LARGE SAMPLE INFERENCE FOR IRREGULARLY SPACED
DEPENDENT OBSERVATIONS BASED ON SUBSAMPLING

BY

DIMITRIS N. POLITIS
EFSTATHIOS PAPARODITIS
JOSEPH P. ROMANO

TECHNICAL REPORT NO. 495
MARCH 1996

PREPARED UNDER THE AUSPICES
OF
NATIONAL SCIENCE FOUNDATION
GRANT DMS 94-03826

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
Large Sample Inference for Irregularly Spaced Dependent Observations Based on Subsampling

By

Dimitris N. Politis
Efstathios Paparoditis
Joseph P. Romano

Technical Report No. 495
March 1996

Prepared Under the Auspices

Of

National Science Foundation Grant DMS 94-03826

Department of Statistics
Stanford University
Stanford, California
Large sample inference for irregularly spaced dependent observations based on subsampling

Dimitris N. Politis                      Efstathios Paparoditis
Purdue University and University of Cyprus University of Cyprus
Joseph P. Romano
Stanford University
February 28, 1996

Abstract

In many contexts, e.g., queueing theory, spatial statistics, etc., the data may consist of measurements of some quantity at irregularly scattered points in time and/or space; in other words, the data might correspond to a realization of a marked point process over a compact subset of the space of points. In this paper, we formulate a modified version of the general subsampling methodology which was originally put forth in Politis and Romano (1994) for data observed over points on a lattice, and show that it leads to valid large sample inferences in a very general estimation set-up involving data from a marked point process.

AMS 1980 subject classifications: primary 62M20; secondary 62G05.

Keywords and phrases: Confidence limits, large sample inference, marked point processes, nonparametric estimation, Poisson processes, random fields, subsampling.

*Research partially supported by NSF grants DMS 94-04329 and DMS 94-03826. Direct correspondence to: Dimitris N. Politis, Department of Mathematics and Statistics, University of Cyprus, P.O. Box 537, Nicosia, Cyprus; Fax: +357-2-339061; e-mail: politis@pythagoras.mas.ucy.ac.cy or politis@stat.purdue.edu
1 Introduction

Suppose \( \{X(t), t \in \mathbb{R}^d\} \) is a continuous parameter random field in \( d \) dimensions, with \( d \in \mathbb{Z}^+ \), i.e., a collection of random variables \( X(t) \) taking values in an arbitrary state space \( S \), and indexed by the variable \( t \in \mathbb{R}^d \). In the important special case where \( d = 1 \), the random field \( \{X(t)\} \) is just a continuous time stochastic process. The probability law of the random field \( \{X(t), t \in \mathbb{R}^d\} \) will be denoted by \( P_X \).

Our objective is statistical inference pertaining to the unknown probability law \( P_X \) on the basis of data. In case the data are of the form \( \{X(t), t \in E\} \), with \( E \) being a finite subset of the rectangular lattice \( \mathbb{Z}^d \), a general methodology for large sample non-parametric inference was developed in Politis and Romano (1994); see also Wu (1990), Sherman (1992), Bickel et al. (1994) and Shao and Tu (1995) for related subsampling ideas in the context of dependent and independent data. However, in many important cases, e.g., queueing theory, spatial statistics, mining and geostatistics, meteorology, etc., the data correspond to observations of \( X(t) \) at non-lattice, irregularly spaced points. For instance, if \( d = 1 \), \( X(t) \) might represent the required service time for a customer arriving at a service station at time \( t \). If \( d = 2 \), \( X(t) \) might represent a measurement of the quality or quantity of the ore found in location \( t \), or a measurement of precipitation at location \( t \) during a fixed time interval, etc. As a matter of fact, in case \( d > 1 \), irregularly spaced data seem to be the rule rather than the exception; cf., for example, Cressie (1993), Karr (1991), Ripley (1981).

A useful and parsimonious way to model the irregularly scattered \( t \)-points is to assume they are generated by a Poisson point process observable on a compact subset \( K \subset \mathbb{R}^d \), and assumed independent of the random field \( \{X(t)\} \). So let \( N \) denote such a Poisson process on \( \mathbb{R}^d \) with mean measure \( \Lambda \), i.e., \( EN(A) = \Lambda(A) \) for any set \( A \subset \mathbb{R}^d \), and independent of \( \{X(t)\} \). The point process \( N \) can be expressed as \( N = \sum; \epsilon_t \), where \( \epsilon_t \) is a point mass at \( t \), i.e., \( \epsilon_t(A) = 1 \) or 0 according to whether \( t \in A \) or not; in other words, \( N \) is a random (counting) measure on \( \mathbb{R}^d \). The expected number of \( t \)-points to be found in \( A \) is \( \Lambda(A) \), whereas the actual number of \( t \)-points found in set \( A \) is given by \( N(A) \). The joint (product) probability law of the random field \( \{X(t)\} \)
and the point process $N$ will be denoted by $P$. The observations then are described via the 'marked point process' $\tilde{N} = \sum_i \epsilon_{\{t_i, X(t_i)\}}$, that is, the point process $N$ where each $t$-point is 'marked' by the value of $X$ at that point.

The Poisson assumption can of course be tested with a particular dataset at hand; e.g., in the precipitation example previously mentioned, the locations of a raingage for a region would usually be found in the nearest small town, and locations of towns can plausibly be thought as conforming to Poisson models (cf. Karr (1986)). In general, there are many reasons to justify the popularity of the Poisson assumption. To name just a few:

- Parsimony: the law of the process is completely specified by prescribing the mean measure $\Lambda$.

- Conditional uniformity or 'complete spatial randomness' of the $t$-points: given the value of $N(K)$, i.e., given the number of the $t$-points - which plays the role of our sample size here- the locations of the $t$-points are i.i.d. from probability measure $\xi$, where $\xi(A) = \Lambda(A \cap K)/\Lambda(K)$; see Cressie (1993) or Karr (1991). In particular, if the Poisson process is homogeneous, i.e., if $\Lambda$ is proportional to Lebesgue measure, then the $t$-points are uniformly distributed over $K$.

- Alias-free measurement or identifiability of $P_X$ (rather: of all finite-dimensional distributions corresponding to $P_X$) on the basis of Poisson samples; see Theorem 10.18 in Karr (1991).

Under the assumption of a homogeneous Poisson process generating the $t$-points, Karr (1986) developed an elegant nonparametric estimation methodology for standard problems in the analysis of time series and random fields such as estimation of the mean $\mu$ and the autocovariance $R(\cdot)$, and proved consistency and asymptotic normality of his estimators; see also Karr (1991, ch.10), Lii and Masry (1994), Lii and Tsou (1995), Masry (1978, 1983, 1988), Krickeberg (1982), and Jolivet (1981) for related developments including estimation of the spectrum and bispectrum of the random field $\{X(t)\}$ based on random sampling.
In this report, we develop a nonparametric methodology for the construction of large sample confidence regions for general parameters of the probability law \( P_X \) based on subsampling under the assumption of a homogeneous Poisson process generating the \( t \)-points. In the next section some definitions and notations are given, as well as a useful lemma on mixing properties of the sampled process. In section 3, the subsampling methodology is described and the large sample validity of the proposed confidence regions is established. Stochastic approximations to the subsampling estimators are introduced in section 4. Section 5 contains some illustrative examples, and Section 6 some concluding remarks; all proofs are deferred to the Appendix.
2 Definitions and some different notions on mixing

The random field \( \{X(t)\} \) is called stationary or homogeneous if for any set \( E \subset \mathbb{R}^d \), and for any point \( h \in \mathbb{R}^d \), the joint distribution of the random variables \( \{X(t), t \in E\} \) is identical to the joint distribution of \( \{X(t), t \in E + h\} \), where \( E + h = \{y : y = t + h, \text{ with } t \in E\} \). The random field \( \{X(t)\} \) is called weakly stationary or \( L_2 \)-stationary if, for any \( t, h \in \mathbb{R}^d \), \( EX(t)^2 < \infty \), \( EX(t) = \mu \), and \( Cov(X(t), X(t + h)) = R(h) \), i.e., \( EX(t) \) and \( Cov(X(t), X(t + h)) \) do not depend on \( t \) at all. In the presence of finite second moments, stationarity implies weak stationarity.

Denote by \( |A| \) the Lebesgue measure (i.e., the volume) of set \( A \in \mathbb{R}^d \), and by \( dt \) the corresponding Lebesgue measure (volume) element situated at \( t \). Let \( \mathbb{R}_+ \) denote the nonnegative part of the real axis, and let \( K \) be a compact, convex subset of \( \mathbb{R}_+^d \), which will be our observation region, i.e., the region where the point process \( N \) is observed. For two points \( t = (t_1, \ldots, t_d) \) and \( u = (u_1, \ldots, u_d) \) in \( \mathbb{R}^d \), denote the \( l_\infty \) or sup-distance in \( \mathbb{R}^d \) by \( \rho(t, u) = \sup_j |t_j - u_j| \), and for two sets \( E_1, E_2 \) in \( \mathbb{R}^d \), define \( \rho(E_1, E_2) = \inf\{\rho(t, u) : t \in E_1, u \in E_2\} \). Let \( \delta(K) \) denote the supremum of the diameters of all \( l_\infty \) balls (i.e., hypercubes) contained in \( K \), and let \( \Delta(K) \) denote the infimum of the diameters of all \( l_\infty \) balls that contain \( K \).

Observe that, because \( N(K) \) is Poisson distributed with mean \( \Lambda(K) \), if

\[
\Lambda(K) \to \infty
\]

as \( |K| \to \infty \), then this is sufficient to ensure that

\[
N(K)/\Lambda(K) \xrightarrow{a.s.} 1;
\]

in other words, our sample size \( N(K) \) will tend to infinity with probability one as \( |K| \to \infty \) provided \( \Lambda(K) \to \infty \). The Poisson process \( N \) is homogeneous if its mean measure \( \Lambda \) is proportional to Lebesgue measure, i.e., if \( \Lambda(dt) = \lambda dt \), for some constant \( \lambda > 0 \). For a homogeneous Poisson process it follows immediately that equation (1) -and thus equation (2) as well- hold true.

It is easy to construct examples where without some form of weak dependence holding statistical inference may be inaccurate even for large samples. Therefore, the
random field \( \{X(t)\} \) will be assumed to satisfy a certain weak dependence condition. In order to quantify the strength of dependence we will make use of strong mixing coefficients to be defined below.

As in the recent literature on mixing for random fields (see e.g. Doukhan (1994)), we define a collection of strong mixing coefficients by

\[
\alpha_X(k; l_1, l_2) \equiv \sup_{E_1, E_2 \subset \mathbb{R}^d} \{|P(A_1 \cap A_2) - P(A_1)P(A_2)| : A_i \in \mathcal{F}(E_i), |E_i| \leq l_i, i = 1, 2, \rho(E_1, E_2) \geq k\}
\]

where \( \mathcal{F}(E_i) \) is the \( \sigma \)-algebra generated by \( \{X(t), t \in E_i\} \). A weak dependence (i.e., mixing) 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. If we let \( \alpha_X(k) = \alpha_X(k; \infty, \infty) \) be the usual strong mixing coefficient of M. Rosenblatt, 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 \( \{X(t)\} \) is simply said to be strong mixing.

We will now introduce two slightly weaker forms of mixing coefficients for random fields that will be sufficient for our purposes. So we define

\[
\alpha_X(k; l_1) \equiv \sup_{E_1 \subset \mathbb{R}^d} \{|P(A_1 \cap A_2) - P(A_1)P(A_2)| : A_i \in \mathcal{F}(E_i), i = 1, 2, |E_i| \leq l_1, \rho(E_1, E_2) \geq k, E_2 = E_1 + t, t \in \mathbb{R}^d\}
\]

and

\[
\bar{\alpha}_X(k; l_1) \equiv \sup_{E_1, \text{convex} \subset \mathbb{R}^d} \{|P(A_1 \cap A_2) - P(A_1)P(A_2)| : A_i \in \mathcal{F}(E_i), i = 1, 2, |E_i| \leq l_1, \rho(E_1, E_2) \geq k, E_2 = E_1 + t, t \in \mathbb{R}^d\}
\]

Note that, by definition, \( \alpha_X(k; l_1) \) and \( \alpha_X(k; l_1, l_1) \) are different entities; however, the only difference between them is that in the calculation of \( \alpha_X(k; l_1) \) the sets \( E_1, E_2 \) are restricted to have the same shape, i.e., one is just a translate of the other. No such restriction exists for the calculation of \( \alpha_X(k; l_1, l_1) \); consequently, \( \alpha_X(k; l_1) \leq \alpha_X(k; l_1, l_1) \). Thus, a mixing condition based on the coefficients \( \alpha_X(k; l_1) \) is easier to satisfy than a mixing condition based on the coefficients \( \alpha_X(k; l_1, l_2) \). Similarly, \( \bar{\alpha}_X(k; l_1) \leq \alpha_X(k; l_1) \) because of the convexity restriction in calculation of \( \bar{\alpha}_X(k; l_1) \).

The essence of the following lemma is that the strong mixing properties of the random field \( \{X(t)\} \) are passed on to the marked point process \( \tilde{N} \), at least in the sense
that covariances of arbitrary functions can be bounded by the mixing coefficients; the lemma is patterned after Lemma 4.1 of Masry (1988).

**Lemma 1** Let \( N \) be a Poisson process on \( \mathbb{R}^d \) independent of the random field \( \{ X(t), t \in \mathbb{R}^d \} \); neither \( N \) nor \( \{ X(t) \} \) are assumed homogeneous here. Let \( E_1, E_2 \) be two subsets of \( \mathbb{R}^d \) such that \( |E_i| = l_i, i = 1, 2 \), and \( \rho(E_1, E_2) = k > 0 \). Also let \( V_i \) be a random variable measurable with respect to the \( \sigma \)-algebra generated by the part of the marked point process \( \tilde{N} \) whose \( t \)-points happen to fall inside the set \( E_i \); also assume that \( |V_i| \leq 1 \) almost surely for \( i = 1, 2 \). Then

\[
|\text{Cov}(V_1, V_2)| \leq 4\alpha_X(k; l_1, l_2).
\]

As a clarification, we note that \( V_i \) is some function of the \( t \)-points generated by \( N \) that happen to fall inside the set \( E_i \), along with their corresponding \( X(t) \) 'marks'. The ordering of the \( t \)-points is immaterial, and therefore arbitrary, as long as each \( t \)-point is properly paired with its corresponding \( X(t) \) mark. For example, if \( g_{i,k} \) is a real valued, bounded function of \( k \) two-dimensional arguments, then we may say that \( V_i = g_{i,N(E_i)}([t_1^{(i)}, X(t_1^{(i)})], [t_2^{(i)}, X(t_2^{(i)})], \ldots, [t_{N(E_i)}^{(i)}, X(t_{N(E_i)}^{(i)})]) \). Ignoring momentarily the assumed boundedness, \( V_i \) can be thought of as being a general statistic calculated from the data \( \{[t_1^{(i)}, X(t_1^{(i)})], [t_2^{(i)}, X(t_2^{(i)})], \ldots, [t_{N(E_i)}^{(i)}, X(t_{N(E_i)}^{(i)})]\} \).

If the sets \( E_1, E_2 \) are simple translates of one another then the following lemma is also true; its proof is identical to the proof of Lemma 1.

**Lemma 2** Let \( N \) be a Poisson process on \( \mathbb{R}^d \) independent of the random field \( \{ X(t), t \in \mathbb{R}^d \} \); neither \( N \) nor \( \{ X(t) \} \) are assumed homogeneous here. Let \( E_1 \) be a subset of \( \mathbb{R}^d \), and let \( E_2 = E_1 + t \) for some \( t \in \mathbb{R}^d \). Let \( |E_1| = l_1 \), and \( \rho(E_1, E_2) = k > 0 \). Also let \( V_i \) be a random variable measurable with respect to the \( \sigma \)-algebra generated by the part of the marked point process \( \tilde{N} \) whose \( t \)-points happen to fall inside the set \( E_i \); also assume that \( |V_i| \leq 1 \) almost surely for \( i = 1, 2 \). Then

\[
|\text{Cov}(V_1, V_2)| \leq 4\alpha_X(k; l_1).
\]
If the set $E_1$ is convex, then also

$$|\text{Cov}(V_1, V_2)| \leq 4\bar{a}_X(k; l_1).$$
3 Subsampling stationary marked point processes

Now and for the remainder of the paper we will assume that the random field \( \{X(t)\} \) is stationary and independent of the homogeneous Poisson process \( N \) of rate \( \lambda > 0 \). Note that \( \lambda \) is in general unknown but estimable by \( N(K)/|K| \); see (2). Our goal will be to construct a confidence region for a parameter \( \theta = \theta(P_X) \), on the basis of observing \( \{X(t)\} \) for the \( t \)-points generated by the Poisson process \( N \) over the compact, convex set \( K \), i.e., observing the marked point process \( \tilde{N} \) over the corresponding region. The parameter \( \theta \) will be assumed to generally take values in a normed linear space (e.g., a Banach space) \( \Theta \) endowed with norm \( || \cdot || \); the case where \( \Theta = \mathbb{R} \) is an important special case.

Let \( T_K = T_K(\tilde{N}) \) be a random variable measurable with respect to the \( \sigma \)-algebra generated by the part of the marked point process \( \tilde{N} \) whose \( t \)-points happen to fall inside the set \( K \). Similarly to the \( V_i \) appearing in Lemma 1, \( T_K \) is a function of the arguments \( [t_1^{(i)}, X(t_1^{(i)})], [t_2^{(i)}, X(t_2^{(i)})], \ldots [t_{N(K)}^{(i)}, X(t_{N(K)}^{(i)})] \), where \( t_1, t_2, \ldots, t_{N(K)} \) are the \( t \)-points generated by \( N \) (and ordered by some fashion) that happen to fall inside the set \( K \); the functional form of this function only depends on \( K \) and on \( N(K) \). Note that, because of homogeneity of the random field \( \{X(t)\} \), and of the Poisson process \( N \), the functional form of \( T_K(\cdot) \) does not depend on the absolute location of the set \( K \) within \( \mathbb{R}^d \); thus, for example, the probability law of \( T_K(\tilde{N}) \) is identical to that of \( T_{K+y}(\tilde{N}) \), for any \( y \in \mathbb{R}^d \).

The statistic \( T_K \) is employed to consistently estimate \( \theta \) at some rate \( \tau_{\lambda(K)} \), where \( \tau. \) will be assumed to be a known, regularly varying function of index \( \zeta \), i.e., for \( u \in \mathbb{R}_+ \)

\[
\tau_u = u^\zeta s(u),
\]

with \( s(\cdot) \) being a known, normalized slowly varying function, such that for any \( l > 0, \lim_{x \to \infty} \frac{s(\sqrt{l}x)}{s(x)} = 1 \) (see Bingham, Goldie, and Teugels (1987)), and \( \zeta > 0 \) a known constant. Nevertheless, the requirement that \( \zeta \) and \( s(\cdot) \) are known can be side-stepped – see Remark 2. Assumption (3) in connection with (2) will be used in our substituting \( \tau_{N(K)} \) in place of \( \tau_{\lambda(K)} \) since then \( \tau_{N(K)}/\tau_{\lambda(K)} \xrightarrow{a.s.} 1 \) as \( |K| \to \infty \). Note that because
of the homogeneity of the point process \( \Lambda(K) = \lambda |K| \), and —in view of (3)— we could drop the constant \( \lambda \) and say that \( T_K \) is consistent for \( \theta \) at rate \( \tau_{|K|} \); nevertheless, it seems more intuitive to have the rate of convergence depend on the (actual or expected) sample size directly.\footnote{Although this is not a crucial point, let us here add that considerations of invariance also lead us to prefer having the rate as a function of \( \Lambda(K) = \lambda |K| \) or \( N(K) \) rather than just \( |K| \); to see this, observe that \( \Lambda(K) \) and \( N(K) \) are dimensionless quantities, i.e., pure numbers, whereas \( |K| \) has the dimensions (and units) of volume. For example, suppose we have that asymptotically \( \text{Var}(\tau_{\Lambda(K)} T_K) \approx \text{Var}(\tau_{N(K)} T_K) \approx \sigma^2 \), and \( \text{Var}(\tilde{\tau}_{|K|} T_K) \approx \tilde{\sigma}^2 \); if the units of measurement of volume are changed (i.e., a scaling is performed), then \( \tau_{\Lambda(K)}, \tau_{N(K)} \) and \( \sigma^2 \) all remain unchanged (invariant), whereas both \( \tau_{|K|} \) and \( \tilde{\sigma}^2 \) have to change (in opposing directions) to accommodate the fact that \( \text{Var}(T_K) \) has to remain unchanged.}

Let \( J_K(P) \) be the sampling law and \( J_K(x, P) \) the sampling (cumulative) distribution function of the real valued random variable \( \tau_{\Lambda(K)}(T_K - \theta(P_X)) \), i.e., \( J_K(x, P) = \text{Prob}_P\{\tau_{\Lambda(K)}(T_K - \theta(P_X)) \leq x\} \). The only assumption that will essentially be needed is existence of a large-sample distribution for this random variable stated as assumption \( A \) below:

**Assumption A.** \( J_K(P) \) converges weakly to a limit law \( J(P) \), with corresponding distribution function \( J(z, P) \), as \( \delta(K) \to \infty \).

Note that one of the implications of Assumption \( A \) is that, as \( \delta(K) \) increases, \( J_K(P) \) progressively loses its dependence on the shape of \( K \).

If \( \Theta = \mathbb{R} \), we can also define \( J^*_K(P) \) to be the sampling law and \( J^*_K(x, P) \) the sampling (cumulative) distribution function of \( \tau_{\Lambda(K)}(T_K - \theta(P_X)) \), i.e., \( J^*_K(x, P) = \text{Prob}_P\{\tau_{\Lambda(K)}(T_K - \theta(P_X)) \leq x\} \); then we can also formulate assumption \( A^* \) stated below:

**Assumption \( A^* \).** \( J^*_K(P) \) converges weakly to a limit law \( J^*(P) \), with corresponding distribution function \( J^*(z, P) \), as \( \delta(K) \to \infty \).
Many practically important statistics have been shown to possess asymptotic distributions under some conditions, usually in connection with homogeneous Poisson sampling. For example, Karr (1986) in the case of homogeneous Poisson sampling showed\(^2\) that the sample mean \(\hat{\mu} = N(K)^{-1} \int_K X(t)N(dt)\) is consistent for \(\mu\), and asymptotically normal at rate \(\sqrt{|K|}\) and asymptotic variance equal to \(\int R(t)dt + \lambda^{-1} \bar{R}(0)\); here, and in what follows, if the region of integration is left unspecified, it is implied that integration takes place over the whole of \(\mathbb{R}^d\). Incidentally, note the obvious typographical mistake in formula (3.10) of Karr (1986) for the asymptotic variance. Similarly, Karr (1986) showed asymptotic normality at some bandwidth-dependent rate for the nonparametric estimate of the autocovariance \(R(t)\). Other examples of statistics of interest possessing well-defined large sample distributions are given by Karr (1991), Krickeberg (1982), Lii and Tsou (1995), Masry (1978, 1983, 1988), and Jolivet (1981).

In any case, assumption \(A\) or \(A^*\) is minimal in the sense that in order to conduct inference with no parametric assumptions on \(P_X\) one is inevitably led to large sample theory, and existence of an asymptotic distribution is an almost \textit{sine qua non} condition. One should also note that we pose no restrictions on the statistic \(T_K\) other -in essence- that it be a proper statistic, i.e., a (measurable) function of our data at hand. Nonetheless, as will be seen shortly, we will be able to derive an estimator of the large-sample distribution of \(T_K\) with no further assumptions other than mixing; this unique generality is afforded to us only via the subsampling methodology.

Recall that our observation region \(K\) is a compact, convex subset of \(\mathbb{R}_+^d\). Let \(c = c(K)\) be a number in \((0, 1)\) depending on \(K\), and define a scaled down replica of \(K\) by \(B = \{ct : t \in K\}\), where \(t = (t_1, \ldots, t_d)\) and \(ct = (ct_1, \ldots, ct_d)\); \(B\) has the same shape as \(K\) but smaller dimensions. Also define the displaced sets \(B + y\) and the set of ‘allowed’ displacements \(K_{1-c} = \{y \in K : B + y \subset K\}\). By temporarily ‘ignoring’ all

\(^2\)Interestingly, Karr (1986) also showed that the estimator \(N(K)^{-1} \int_K X(t)N(dt)\) has smaller finite-sample variance than the estimator \(\Lambda(K)^{-1} \int_K X(t)N(dt)\) even in the case where \(\lambda\), and therefore \(\Lambda(K)\) as well, are known! This observation leads us to conjecture that, even though \(\tau_{N(K)}/\tau_{M(K)} \xrightarrow{a.s.} 1\) as \(|K| \to \infty\), it might be more accurate in finite samples to use \(\tau_{N(K)}\) instead of \(\tau_{M(K)}\); in other words, to have the rate of convergence depend on the actual, rather than the expected, sample size.
data with locations outside of $B + y$ we can compute the statistic $T_{B + y} = T_{B + y}(\tilde{N})$, for $y \in K_{1-c}$, as if the point process $N$ with its attached 'marks' were observed only over $B + y$. So let $S_{K,y} = T_{B + y}(\tilde{N})$ denote this 'subsample' value.

As in Politis and Romano (1994) we define subsampling estimators of $J_k^*(x, P)$ and $J_K(x, P)$ by $L_k^*(x)$ and $L_K(x)$ respectively, where

$$ L_K(x) = |K_{1-c}|^{-1} \int_{K_{1-c}} 1\{{\tau}_N(B+y)||S_{K,y} - T_K|| \leq x\} dy $$

and whereas, if $\theta$ and $T_K$ are real valued,

$$ L_k^*(x) = |K_{1-c}|^{-1} \int_{K_{1-c}} 1\{{\tau}_N(B+y)(S_{K,y} - T_K) \leq x\} dy. $$

The following theorem establishes asymptotic consistency of $L_k^*$, and asymptotic validity of confidence intervals for $\theta$ that are constructed based on quantiles of $L_k^*$.

**Theorem 1** Let $\theta$ be real valued and assume assumption $A^*$ and (3). Let $\delta(K) \to \infty$, and let $c = c(K) \in (0, 1)$ be such that $c \to 0$ but $c\delta(K) \to \infty$. Finally, assume that

$$ |K_{1-c}|^{-1} \int_0^{(1-c)\Delta(K)} u^{d-1} \delta_X(u; c|K|) du \to 0. $$

Then,

(i) $L_k^*(x) \overset{P}{\to} J^*(x, P)$ for every continuity point $x$ of $J^*(x, P)$.

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

$$ \sup_x |L_k^*(x) - J^*(x, P)| \overset{P}{\to} 0. $$

(iii) Let $q_k^*(1 - \alpha) = \inf\{x : L_k^*(x) \geq 1 - \alpha\}$, and let $q^*(1 - \alpha) = \inf\{x : J^*(x, P) \geq 1 - \alpha\}$. If $J^*(x, P)$ is continuous at $x = q^*(1 - \alpha)$, then

$$ \text{Prob}_P\{\tau_{M(K)}(T_K - \theta(P_X)) \leq q_k^*(1 - \alpha)\} \to 1 - \alpha. $$

Thus the asymptotic coverage probability of the interval $(T_K - \tau_{N(K)}^{-1}q_k^*(1 - \alpha), \infty)$ is the nominal level $1 - \alpha$.

The following theorem establishes asymptotic consistency of $L_K$, and asymptotic validity of confidence regions for $\theta$ that are constructed based on quantiles of $L_K$. 

12
Theorem 2 Assume assumption A and (3). Let \( \delta(K) \to \infty \), and let \( c = c(K) \in (0, 1) \) be such that \( c \to 0 \) but \( c\delta(K) \to \infty \). Finally, assume that \( |K_{1-c}|^{-1} \int_0^{(1-c)\Delta(K)} u^{d-1} \alpha_X(u; c^d|K|) du \to 0. \) Then,

(i) \( L_K(x) \xrightarrow{P} J(x, P) \) for every continuity point \( x \) of \( J(x, P) \).

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

\[
\sup_x |L_K(x) - J(x, P)| \xrightarrow{P} 0.
\]

(iii) Let \( q_K(1-\alpha) = \inf\{x : L_K(x) \geq 1-\alpha\} \), and let \( q(1-\alpha) = \inf\{x : J(x, P) \geq 1-\alpha\} \).

If \( J(x, P) \) is continuous at \( x = q(1-\alpha) \), then

\[
\text{Prob}_P\{\tau_{\Lambda(K)}||T_K - \theta(P_X)|| \leq q_K(1-\alpha)\} \to 1 - \alpha.
\]

Thus the asymptotic coverage probability of the confidence region \( \{\theta \in \Theta : \tau_{\Lambda(K)}||T_K - \theta|| \leq q_K(1-\alpha)\} \) is the nominal level \( 1 - \alpha \).

Remark 1. If the limit law \( J^*(x, P) \) happens to be known, then a mixture of the subsampling distribution estimator \( L_K^*(x) \) with \( J^*(x, P) \) may well possess higher order accuracy properties as in Booth and Hall (1993) and Bertail (1994); however, we will not pursue the subject of higher order accuracy in this paper.

Remark 2. If \( \zeta \) and \( s(\cdot) \) appearing in (3) happen to be unknown they can be estimated consistently via a preliminary round of subsampling in the same fashion as in Bertail et al. (1995). Then the estimated rate \( \hat{\tau}_{\Lambda(K)} \), say) can be used in place of the true rate \( \tau_{\Lambda(K)} \) in the construction of our subsampling confidence intervals and regions without destroying their asymptotic validity.

It is of interest to consider what happens if \( K \) does not expand towards \( \mathbb{R}^d \) in all directions as implied by the assumption \( \delta(K) \to \infty \). If \( d = 2 \), for example, one can envision a situation where the practitioner is faced with an observation region \( K \) that is long and thin; notwithstanding the thinness in one direction, if \( |K| \) is large enough, the sample size \( N(K) \) will be large enough so accurate estimation may still be possible.

To avoid cumbersome notations –and because this actually may be a most useful set-up in practice– let us now assume that \( K \) is the rectangle \([0, k_1] \times [0, k_2] \times \cdots \times [0, k_d] \),
and that $B$ is the smaller rectangle $B = [0, b_1] \times [0, b_2] \times \ldots \times [0, b_d]$, where $b_i = ck_i$, for $i = 1, 2, \ldots, d$. Therefore, $K_{1-c} = [0, (1-c)k_1] \times [0, (1-c)k_2] \times \ldots \times [0, (1-c)k_d]$ is a rectangle with the property that for all $y \in K_{1-c}$, $B + y$ is a subset of $K$.

Now assume that

$$0 < k_i \uparrow \infty \text{ for } i = 1, \ldots, d^* \text{ and } 0 < k_i \uparrow K_i < \infty \text{ for } i = d^* + 1, \ldots, d$$

(6)

where $d^*$ is some positive integer less or equal to $d$. Then the following two corollaries to Theorems 1 and 2 respectively are true.

**Corollary 1** Let $\theta$ be real valued and assume (3) and (6). Assume that $J^*_K(P)$ converges weakly to a limit law $J^*(P)$, with corresponding distribution function $J^*(x, P)$, as $|K| \to \infty$ under (6). Let $c = c(K) \in (0, 1)$ be such that $c \to 0$ but $c^d|K| \to \infty$ as $|K| \to \infty$ under (6). Finally, assume that $|K_{1-c}|^{-1} \int_0^{(1-c)\max_k k_i} u^{d^*-1} \alpha_X(u; c^d|K|)du \to 0$. Then, conclusions (i), (ii), and (iii) of Theorem 1 remain true.

**Corollary 2** Assume (3), (6) and that $J_K(P)$ converges weakly to a limit law $J(P)$, with corresponding distribution function $J(x, P)$, as $|K| \to \infty$ under (6). Let $c = c(K) \in (0, 1)$ be such that $c \to 0$ but $c^d|K| \to \infty$ as $|K| \to \infty$ under (6). Finally, assume that $|K_{1-c}|^{-1} \int_0^{(1-c)\max_k k_i} u^{d^*-1} \alpha_X(u; c^d|K|)du \to 0$. Then, conclusions (i), (ii), and (iii) of Theorem 2 remain true.

To appreciate that the mixing condition $|K_{1-c}|^{-1} \int_0^{(1-c)\Delta(K)} u^{d-1} \alpha_X(u; c^d|K|)du \to 0$ appearing in the assumptions of Theorems 1 and 2 (or its equivalent condition $|K_{1-c}|^{-1} \int_0^{(1-c)\max_k k_i} u^{d^*-1} \alpha_X(u; c^d|K|)du \to 0$ appearing in the assumptions of Corollaries 1 and 2) is actually quite weak, let us consider the one-dimensional case; it is quite obvious (see also Politis and Romano (1994) for an analogous phenomenon) that, if $d = 1$, the assumed mixing conditions are indeed satisfied provided the process $\{X(t)\}$ is simply strong mixing, i.e., if $\alpha_X(u) \to 0$ as $u \to \infty$. Therefore, the following corollary is immediate.

**Corollary 3** Let $d = 1$, and assume that the stochastic process $\{X(t)\}$ is strong mixing; also assume (3). Let $K = [0, k_1]$, and let $c = c(k_1) \in (0, 1)$ be such that $c \to 0$ but
$ck_1 \to \infty$ as $k_1 \to \infty$.

(a) If assumption $A$ holds, then conclusions (i), (ii), and (iii) of Theorem 2 are true.

(b) If $θ$ happens to be real valued and assumption $A^*$ holds, then conclusions (i), (ii), and (iii) of Theorem 1 are true.

Remark 3. Some discussion on proper choice of $c$ is in order here. Note that the (rather vague) assumption that $c = c(K) \to 0$ but $c\delta(K) \to \infty$ as $\delta(K) \to \infty$ in Theorems 1 and 2 (or its analog $c \to 0$ but $c^d |K| \to \infty$ as $|K| \to \infty$ in our corollaries) suffices for consistency of our subsampling estimators and asymptotic validity of the corresponding confidence regions. Nevertheless, to implement the subsampling method in a practical finite-sample situation, a particular choice of $c$ is required. The situation here is quite analogous to the problem of choosing the block size in subsampling or resampling stationary dependent observations as discussed in Politis and Romano (1992, 1994), and in general requires considerations of second order accuracy. Also, as shown in Hall et al. (1996), the ‘quest’ for the ‘optimal’ block size ultimately depends on the optimality criterion employed; for example, it might be a criterion of accuracy of variance, distribution, or quantile estimation. Even if the vague assumption that $c = c(K) \to 0$ but $c\delta(K) \to \infty$ is ‘narrowed’ to something like $c = A\delta(K)^{-a}$ for some well-defined constants $a \in (0, 1)$ and $A > 0$, the situation typically is that $a$ is known, but $A$ is unknown because it depends on unknown parameters of the underlying probability law. Thus, to effectively use a ‘recipe’ such as $c = A\delta(K)^{-a}$ requires estimation of the underlying unknown parameters that influence $A$, and this typically is a problem harder by an order of magnitude than the original estimation problem at hand. Unfortunately, the problem of choice of $c$ has many common features with the open (or rather, still unconclusively handled) problem of bandwidth choice in nonparametric curve estimation.
4 Stochastic approximations

In the previous section it was shown that the subsampling distributions $L_K$ and $L^*_K$ are consistent estimators of their corresponding large-sample distributions, and can be used to derive asymptotically valid confidence regions for the unknown $\theta$. Nevertheless, equations (4) and (5) involve integrals which—to be of practical use—have to be approximated by finite sums. There are two general ways of performing this approximation, namely:

- Deterministic approximation: For example, the region $K_{1-c}$ can be 'tiled' by a grid consisting of 'small' cells, and then the integral can be approximated by the corresponding Riemann sum.

- Monte-Carlo or stochastic approximation: Points can be dropped 'at random' on $K_{1-c}$, and the averages of $1\{\tau_{N(B+y)}||S_{K,y} - T_K|| \leq x\}$ and $1\{\tau_{N(B+y)}(S_{K,y} - T_K) \leq x\}$ over the dropped $y$-points will be used to approximate $L_K$ and $L^*_K$, respectively.

Both ways are valid provided the number of points over which the approximations are computed is big enough; see also Politis and Romano (1994) for a related discussion. Since the method of deterministic approximation is quite obvious, we will only elaborate on the second method of stochastic approximation.

Let $n$ be a positive integer, and generate $n$ i.i.d. random variables $z_1, z_2, \ldots, z_n$ with distribution that is uniform on $K_{1-c}$; this construction (which is performed independently of the random field $\{X(t)\}$ and of the Poisson process $N$) essentially amounts to generating a realization of the point process $H_n = \sum_{i=1}^n \epsilon_{z_i}$. Stochastic approximations $\tilde{L}_K$ and $\tilde{L}^*_K$ to $L_K$ and $L^*_K$ can now be constructed by

\[
\tilde{L}_K(x) = n^{-1} \int_{K_{1-c}} 1\{\tau_{N(B+y)}||S_{K,y} - T_K|| \leq x\} H_n(dy) \tag{7}
\]

and,

\[
\tilde{L}^*_K(x) = n^{-1} \int_{K_{1-c}} 1\{\tau_{N(B+y)}(S_{K,y} - T_K) \leq x\} H_n(dy), \tag{8}
\]
the latter being defined only if $\theta$ and $T_K$ are real valued. Define also the corresponding quantiles $\tilde{q}_K(1-\alpha) = \inf\{x : \tilde{L}_K(x) \geq 1-\alpha\}$, and $\tilde{q}_K^+(1-\alpha) = \inf\{x : \tilde{L}_K(x) \geq 1-\alpha\}$.

The following theorem shows that the proposed stochastic approximations $\tilde{L}_K^+(x)$ and $\tilde{L}_K(x)$ may be used in place of $L_K^+(x)$ and $L_K(x)$ for constructing asymptotically valid confidence regions for $\theta$.

**Theorem 3** Let $n \rightarrow \infty$ as $d(K) \rightarrow \infty$.

(a) Under the assumptions of Theorem 1, and if $J^*(x, P)$ is continuous at $x = q^*(1-\alpha)$, then the asymptotic coverage probability of the interval $(T_K - \tau_{N(K)}^{-1}\tilde{q}_K^+(1-\alpha), \infty)$ is the nominal level $1 - \alpha$.

(b) Under the assumptions of Theorem 2, and if $J(x, P)$ is continuous at $x = q(1-\alpha)$, then the asymptotic coverage probability of the confidence region $\{\theta \in \Theta : \tau_{N(K)} ||T_K - \theta|| \leq \tilde{q}_K(1-\alpha)\}$ is the nominal level $1 - \alpha$.

It is perhaps intuitive to have $n$ depend on the original sample size $N(K)$, say, for instance, to take $n = N(K)$. The construction of the new point process $H_n$ is then modified to read: conditionally on the value of $n$, but independently of the random field $\{X(t)\}$ and of the locations of the points generated by the original Poisson process $N$, generate $n$ i.i.d. random variables $z_1, z_2, \ldots, z_n$ with uniform distribution on $K_{1-c}$; the following corollary to Theorem 3 is now easily available.

**Corollary 4** Let $n = h(N(K))$, where $h(\cdot)$ is a monotone, strictly increasing function.

(a) Under the assumptions of Theorem 1, and if $J^*(x, P)$ is continuous at $x = q^*(1-\alpha)$, then the asymptotic coverage probability of the interval $(T_K - \tau_{N(K)}^{-1}\tilde{q}_K^+(1-\alpha), \infty)$ is the nominal level $1 - \alpha$.

(b) Under the assumptions of Theorem 2, and if $J(x, P)$ is continuous at $x = q(1-\alpha)$, then the asymptotic coverage probability of the confidence region $\{\theta \in \Theta : \tau_{N(K)} ||T_K - \theta|| \leq \tilde{q}_K(1-\alpha)\}$ is the nominal level $1 - \alpha$.

**Remark 4.** Note that $H_n$ is in some sense 'mimicking' the original point process $N$; this is especially apparent if we let $n = N(K)$ as in Corollary 4 because in that case $H_n$ becomes itself a Poisson process, identically distributed as $N$. Therefore,
we might further consider what happens if we actually put $N$ in place of $H_n$ (in addition to $N(K)$ in place of $n$) in equations (7) and (8). We conjecture that the resulting approximators $\hat{L}_K$ and $\tilde{L}_K$ would still be usable in the effort to construct asymptotically valid confidence regions for $\theta$ in a manner analogous to Corollary 4. Nevertheless, this is harder to prove, and there does not seem to yield any practical advantage other than saving us from the trouble of generating $n$ i.i.d. uniform random variables.
5 Some illustrative examples

We conclude this work by presenting some illustrative examples where the subsampling methodology may be of practical interest; for the examples we assume that $X(t)$ is real valued.

Example 1: The mean $\mu$. Let $\hat{\mu} = N(K)^{-1} \int_K X(t)N(dt)$ be the sample mean. As mentioned in Section 3, Karr (1986) has shown, under some regularity assumptions, that $\hat{\mu}$ is consistent for $\mu$, and asymptotically normal at rate $\sqrt{|K|}$, with asymptotic variance equal to $\int R(t)dt + \lambda^{-1}R(0)$. Nevertheless, to actually use this asymptotic normality to construct confidence intervals for the mean the asymptotic variance must be estimated. While it is relatively easy to estimate $R(0)$—see our Example 2—, and $\lambda$ is a.s. consistently estimable by $N(K)/|K|$, consistent estimation of $\int R(t)dt$ is not a trivial matter. To see this, note that $\int R(t)dt = f(0)$, where $f(w) \equiv \int R(t)e^{-i(t-w)}dt$ is the spectral density function defined for $w \in \mathbb{R}^d$, and $(t \cdot w)$ denotes the inner product.

Since an asymptotic distribution exists, Assumption $A^*$ is satisfied and the subsampling methodology is immediately seen to be applicable under a mixing condition. Obtaining confidence intervals for $\mu$ using the subsampling distribution $L^*_K(x)$, rather than the asymptotic normal distribution, side-steps the thorny issue of having to perform nonparametric estimation of the spectral density function at 0. Nevertheless, as discussed in Politis and Romano (1994), the variance corresponding to the subsampling distribution $L^*_K(x)$ may be—under additional moment and mixing conditions—a consistent estimator of the asymptotic variance of $T_K$. Since the focus of the present paper is obtaining confidence intervals under the weakest assumptions possible, we will not diverge towards the issue of variance estimation here; see also Carlstein (1986), Possolo (1991), Sherman (1992), and Sherman and Carlstein (1994) for related results on variance estimation based on subsampling dependent data.

Example 2: The autocovariance $R(t)$. The usual estimator of $R(t)$ at some
point \( t \) is given as

\[
\hat{R}(t) = \frac{1}{\lambda^2|K|} \int_K \int_K W_K(t - t_1 + t_2) X(t_1) X(t_2) N(dt_1)(N - \epsilon t_1)(dt_2),
\]

where \( W_K(t) = a_K^{-d} W(t/a_K) \), and the kernel \( W \) is a positive, bounded, isotropic probability function on \( \mathbb{R}^d \); if \( \lambda \) is unknown, its estimator \( N(K)/|K| \) may be used instead. Under regularity conditions, and if the bandwidth \( a_K \to 0 \) but with \( a_K^d|K| \to \infty \) as \( \delta(K) \to \infty \), Karr (1986) showed asymptotic normality of \( \hat{R}(t) \) at rate \( \sqrt{a_K^d|K|} \). Karr (1986) also calculated the asymptotic variance which is found to depend on the 4th order cumulant function of the random field \( X(t) \). Thus, unless the 4th order cumulant function is known to vanish, e.g. if the random field \( X(t) \) is Gaussian, to obtain confidence intervals for \( R(t) \) using the asymptotic distribution it is necessary to estimate the 4th order cumulant function; this is a task quite more formidable than estimation of \( R(t) \). Subsampling again side-steps this difficulty, and immediately yields valid confidence intervals for \( R(t) \) since again Assumption \( A^* \) holds.

**Example 3: Estimating the integral \( \int R(t)g(t)dt \).** Let \( g \) be a continuous function with compact support satisfying \( g(0) = 0 \), and suppose it is required to estimate the integral \( \int R(t)g(t)dt \). Karr (1986, 1991) proposed the estimator

\[
\frac{1}{\lambda^2|K|} \int_K M(dt_1) \int g(t_2 - t_1) M(dt_2),
\]

where \( M(dt) = Y(t)N(dt) \), and showed its asymptotic normality at rate \( \sqrt{|K|} \). In this case, however, the asymptotic variance is too difficult to calculate, let alone estimate. Again, subsampling comes to our rescue, yielding valid confidence intervals for \( \int_A R(t)dt \) since Assumption \( A^* \) holds.

**Example 4: Probability and spectral density estimation.** Let \( X(t) \) have a distribution that admits a probability density function \( p(x) \), and let \( f(w) \) denote the spectral density function as defined in Example 1 – and assuming that the defining integral exists. Different smoothing methods have been developed for consistent estimation of \( p(x) \) and \( f(w) \) (at some given points \( x \in \mathbb{R} \) and \( w \in \mathbb{R}^d \)) and asymp-
Asymptotic normality of the proposed estimates at well-defined rates has been established; see, e.g., Lii and Masry (1994), and Masry (1978, 1983, 1988). Nevertheless, besides the issue of complicated asymptotic variances there is a more serious issue here if our objective is to use the asymptotic normal distribution to obtain confidence intervals for \( p(x) \) and \( f(w) \) on the basis of their estimators denoted by \( \hat{p}(x) \) and \( \hat{f}(w) \); this is the issue of bias. It is well-known that the asymptotic distribution of an optimally smoothed probability or spectral density estimator has a remaining bias term; the term "optimally" here implies a bandwidth choice that "balances" the bias and the standard deviation of the nonparametric estimator at hand, therefore minimizing the mean squared error of estimation. Consequently, the use of the asymptotic normal distribution of an optimally smoothed \( \hat{p}(x) \) or \( \hat{f}(w) \) will only yield confidence intervals for \( E\hat{p}(x) \) and \( E\hat{f}(w) \) respectively, rather than for \( p(x) \) and \( f(w) \) as desired; see also Politis and Romano (1994) for a more detailed discussion on the bias issue. Nevertheless, the asymptotic distribution \( J^*(x, P) \) of our Assumption \( A^* \) does not have to have zero mean for our subsampling methodology to apply. Valid confidence intervals for \( p(x) \) and \( f(w) \) can immediately be obtained from conclusion (iii) of our Theorem 1 here; an intuitive explanation of this phenomenon is that subsampling provides an automatic built-in estimation of the bias, with subsequently bias-corrected confidence intervals.

**Example 5: The bispectrum and a test for normality.** As the spectral density function \( f(w) \) is simply the Fourier transform of the 2nd order cumulant function of the random field \( \{X(t)\} \) (i.e., the autocovariance), the bispectrum \( f(w, l) \) is the Fourier transform of the 3rd order cumulant function (which—by stationarity of \( \{X(t)\} \)—is a function of only two—rather than three—arguments). For the time series case \( d = 1 \), Lii and Tsou (1995) developed nonparametric, kernel-smoothed, consistent estimators \( \hat{f}(w, l) \) of the bispectrum \( f(w, l) \). Under some regularity conditions, Lii and Tsou (1995) also showed asymptotic multivariate normality at some well-defined bandwidth-dependent rate which we will denote by \( \tau_{[K]} \). To elaborate, if \( A = \{(w_j, l_j) \),
for \( j = 1, \ldots, J \) is a discrete set of points, we can define our (multivariate) statistic \( T_K \) to have as its \( j \)th coordinate the value \( \hat{f}(w_j, l_j) \), and our (multivariate) parameter \( \theta \) to have as its \( j \)th coordinate the value \( f(w_j, l_j) \). Note now that asymptotic (multivariate) normality of \( T_K \) implies (by the continuous mapping theorem) existence of an asymptotic distribution for \( \|T_K - \theta\|_\infty \equiv \sup_j |\hat{f}(w_j, l_j) - f(w_j, l_j)| \); in other words, letting \( \Theta = \mathbb{R}^J \) equipped with the \( l_\infty \) or sup-norm \( \| \cdot \|_\infty \), the results of Lii and Tsou (1995) imply that our Assumption \( A \) is satisfied, and thus subsampling may be used to construct an approximate \((1 - \alpha)100\% \) confidence region for \( \theta \) of the following type: 
\[
\tau_K \sup_j |\hat{f}(w_j, l_j) - f(w_j, l_j)| \leq q_K(1 - \alpha).
\]

Now since the parameter \( \theta \) is really a function evaluated at different points, and because we are using the sup-distance (uniform distance) coordinate-wise on \( \theta \), the confidence region for \( \theta \) has the interpretation of a \((1 - \alpha)100\% \) uniform confidence band for the unknown function \( f(w, l) \) of the type: 
\[
f(w, l) = \hat{f}(w, l) \pm q_K(1 - \alpha)/\tau_K,
\]
for all \((w, l) \in A\). Recall that if the random field \( \{X(t)\} \) is Gaussian, then the 3rd order cumulant function (and therefore the bispectrum as well) vanishes. Thus, the just obtained \((1 - \alpha)100\% \) uniform confidence band for the bispectrum affords us the ability to perform a test of the hypothesis \( H_0: "the \ random \ field \ \{X(t)\} \ is \ Gaussian." \). The test rejects the normal hypothesis \( H_0 \) at approximate level \( \alpha \) if the uniform band 
\[
\hat{f}(w, l) \pm q_K(1 - \alpha)/\tau_K
\]
does not fully (i.e., for all \((w, l) \in A\)) 'contain' the graph of the identically zero function; in other words, the test rejects the normal hypothesis if 
\[
0 \notin [\hat{f}(w, l) - q_K(1 - \alpha)/\tau_K, \hat{f}(w, l) + q_K(1 - \alpha)/\tau_K]
\]
for at least one point \((w, l) \in A\). Equivalently, if one constructs the band \(0 \pm q_K(1 - \alpha)/\tau_K \) around the identically zero function, the test rejects the normal hypothesis if the graph of the estimator \( \hat{f}(w, l) \) is not fully contained in this band for all \((w, l) \in A\), i.e., the test rejects \( H_0 \) if 
\[
|\hat{f}(w, l)| > q_K(1 - \alpha)/\tau_K \text{ for at least one point } (w, l) \in A.
\]
6 Conclusion

In this paper, we have introduced a modified version of the general subsampling methodology of Politis and Romano (1994), and we have demonstrated that it leads to valid large sample inferences in a very general estimation set-up involving data from a marked point process. To appreciate the generality recall that:

- The state space $S$ of the random field $\{X(t)\}$ is completely arbitrary; for example, the 'marks' $\{X(t)\}$ may very well be multivariate with some coordinates being real valued, and others being categorical variables.

- The parameter space $\Theta$ is completely arbitrary (other than the fact that it must be a linear space endowed with norm $|| \cdot ||$); for example, $\Theta$ may be a space of continuous functions over an interval, and $|| \cdot ||$ may be the appropriate $L_{\infty}$-norm in which case Theorem 2 may be used to construct uniform confidence bands for the unknown, estimated function.

- No moment conditions have been imposed, and the assumed mixing conditions are quite weak.

- No parametric assumptions on $P_X$, the law of the random field $\{X(t)\}$, have been imposed, and the assumed existence of a well-defined large-sample distribution (of possibly unknown form) for $T_K$ can be seen to be a minimal assumption since asymptotic inference is our objective.

Future research will focus on the aspects touched upon in our Remarks 1–3, of which the issue of proper choice of $c$ is perhaps the most pressing. Work is also in progress on the issue of variance and bias estimation in the context of marked point process that was briefly touched upon in our Examples 1 and 4 of last section.
7 Appendix: Technical proofs

Proof of Lemma 1. Consider first the identity

$$E^{N(E_1), N(E_2)}V_1V_2 - E^{N(E_1), N(E_2)}V_1E^{N(E_1), N(E_2)}V_2 = E^{N(E_1), N(E_2)}\{E(V_1V_2 | N) - E(V_1 | N)E(V_2 | N)\}$$

$$+ E^{N(E_1), N(E_2)}\{E(V_1 | N)E(V_2 | N)\} - E^{N(E_1), N(E_2)}V_1E^{N(E_1), N(E_2)}V_2,$$

where $E^{N(E_1), N(E_2)}V$ is a short-hand for the conditional expectation $E(V | N(E_1), N(E_2))$. Note that $E(V_i | N)$ is a function of only the $t^{(i)}_1, t^{(i)}_2, \ldots, t^{(i)}_{N(E_i)}$ points. Also note that $t^{(1)}_1, t^{(1)}_2, \ldots, t^{(1)}_{N(E_1)}$ and $t^{(2)}_1, t^{(2)}_2, \ldots, t^{(2)}_{N(E_2)}$ are two different collections of t-points, i.e., there are no common t-points to both collections. Finally, by the Poisson assumption, the generated t-points are conditionally (given $N(E_1), N(E_2)$) i.i.d. Thus, $E(V_1 | N)$ and $E(V_2 | N)$ are conditionally (given $N(E_1), N(E_2)$) independent as functions of the distinct, independent collections of t-points $t^{(1)}_1, t^{(1)}_2, \ldots, t^{(1)}_{N(E_1)}$ and $t^{(2)}_1, t^{(2)}_2, \ldots, t^{(2)}_{N(E_2)}$ respectively. Hence,

$$E^{N(E_1), N(E_2)}\{E(V_1 | N)E(V_2 | N)\} = E^{N(E_1), N(E_2)}\{E(V_1 | N)\}E^{N(E_1), N(E_2)}\{E(V_2 | N)\}$$

$$= E^{N(E_1), N(E_2)}V_1E^{N(E_1), N(E_2)}V_2,$$

where it was used that $E^{N(E_1), N(E_2)}\{E(V | N)\} = E^{N(E_1), N(E_2)}V$, since the $\sigma$-algebra generated by the random variables $N(E_1), N(E_2)$ is coarser than that generated by the whole process $N$.

Thus we have shown that

$$E^{N(E_1), N(E_2)}V_1V_2 - E^{N(E_1), N(E_2)}V_1E^{N(E_1), N(E_2)}V_2 = E^{N(E_1), N(E_2)}\{E(V_1V_2 | N) - E(V_1 | N)E(V_2 | N)\}.$$  (9)

Now note that $E^{N(E_1), N(E_2)}V_i = E^{N(E_i)}V_i$, and that $E^{N(E_i)}V_i$ is a function of the random variable $N(E_i)$ only. Also note that because $E_1, E_2$ are assumed disjoint (since $\rho(E_1, E_2) > 0$), the Poisson process properties imply that the random variables $N(E_1), N(E_2)$ are independent. Hence,

$$E\{E^{N(E_1), N(E_2)}V_1E^{N(E_1), N(E_2)}V_2\} = E\{E^{N(E_1)}V_1E^{N(E_2)}V_2\}$$

24
Therefore, taking expectations on both sides of (9) we finally arrive at the relation
\[ EV_1 V_2 - EV_1 E V_2 = E \{ E(V_1 V_2 | N) - E(V_1 | N) E(V_2 | N) \}. \]  \hspace{1cm} (10)

But by a well known mixing inequality (see e.g. Roussas and Ioannidis (1987)) we have that \( |E(V_1 V_2 | N) - E(V_1 | N) E(V_2 | N)| \leq 4\alpha_X(k; l_1, l_2) \) since, conditionally on the realization of the point process \( N \), events generated by the random variable \( V_i \) are included in the \( \sigma \)-algebra \( \mathcal{F}(E_i) \); thus the assertion of the lemma follows immediately. \( \Box \)

**Proof of Theorem 1.** First note that since \( c\delta(K) \to \infty \), it follows that \( |B + y| \to \infty \) for any \( y \). Because of \( c \to 0 \) however, it follows that \( |B + y|/|K| \to 0 \). Now by assumption (3) for \( \zeta > 0 \) we have that \( \tau_{A(B)}/\tau_{A(K)} \to 0 \), and that \( \tau_{A(B)}/\tau_{N(B+y)} \overset{a.s.}{\to} 1 \).

Let \( x \) be a continuity point of \( J^*(x, P) \). Now by an argument similar to the one used in the proof of Theorem 2.1 in Politis and Romano (1994), and using the fact that by assumption (3), \( \tau_{N(A)}/\tau_{A(A)} \overset{a.s.}{\to} 1 \) for any set \( A \) such that \( |A| \to \infty \), it follows that to prove (i) it is sufficient to show that \( U_K^*(x) \overset{P}{\to} J^*(x, P) \), where
\[
U_K^*(x) = |K_{1-c}|^{-1} \int_{K_{1-c}} 1\{\tau_{A(B+y)}(S_{K,y} - \theta) \leq x\}dy.
\]

Note that
\[
EU_K^*(x) = |K_{1-c}|^{-1} \int_{K_{1-c}} Prob_P\{\tau_{A(B+y)}(S_{K,y} - \theta) \leq x\}dy.
\]

But
\[
Prob_P\{\tau_{A(B+y)}(S_{K,y} - \theta) \leq x\} = J_{B+y}^*(x, P) = J_B^*(x, P),
\]
the latter equality holding being due to homogeneity (translation invariance of the probability laws). So the integrand does not depend on \( y \), and it follows immediately that
\[
EU_K^*(x) = J_B^*(x, P) \to J^*(x, P),
\]
the latter convergence being due to Assumption \( A^* \) in connection with \( \delta(B) = c\delta(K) \to \infty \).
Now look at

\[ \text{Var}(U^*_K(x)) = |K_{1-c}|^{-2} \int_{K_{1-c}} \int_{K_{1-c}} \text{Cov}(I(x), I(y)) \, dx \, dy \]

where \( I(x) \equiv 1\{\tau_{\Lambda(B+X)}(S_{K,X - \theta}) \leq x \} \) and \( I(y) \equiv 1\{\tau_{\Lambda(B+Y)}(S_{K,Y - \theta}) \leq y \} \).

Again by the homogeneity concerning translations of the domain of \( x \), the random field \( \{I(x), x \in \mathbb{R}^d\} \) is stationary as well, and we have \( \text{Cov}(I(x), I(y)) = C(x - y) \) for some \( C(\cdot, \cdot) \).

A straightforward calculation now gives

\[
\text{Var}(U^*_K(x)) = |K_{1-c}|^{-1} \int_{K_{1-c}^+} C(y) \frac{|K_{1-c} \cap K_{1-c} - y|}{|K_{1-c}|} \, dy
\]

\[
\leq |K_{1-c}|^{-1} \int_{K_{1-c}^+} |C(y)| \, dy = |K_{1-c}|^{-1} \int_{B^+} |C(y)| \, dy + |K_{1-c}|^{-1} \int_{K_{1-c}^+ - B^+} |C(y)| \, dy
\]

where \( K_{1-c}^\pm = \{t : t = x - y, \text{ with } x, y \in K_{1-c}\} \) and \( B^\pm = \{t : t = x - y, \text{ with } x, y \in B\} \). Note that \( |C(y)| \leq 1 \) and thus

\[
|K_{1-c}|^{-1} \int_{B^\pm} |C(y)| \, dy = O(\frac{|B^\pm|}{|K_{1-c}|}) = O(c^d) = o(1).
\]

Finally, for \( y \in K_{1-c}^+ - B^\pm \), we have by Lemma 2 that \( |C(y)| \leq 4 \tilde{\alpha}_X(\max_i |y_i - b_i|; |B|) \).

Therefore,

\[
|K_{1-c}|^{-1} \int_{K_{1-c}^+ - B^\pm} |C(y)| \, dy = O \left( |K_{1-c}|^{-1} \int_{K_{1-c}^+ - B^\pm} \tilde{\alpha}_X(\max_i |y_i - b_i|; |B|) \, dy \right)
\]

\[ = O \left( |K_{1-c}|^{-1} \int_0^{(1-c)\Delta(K)} u^{d-1} \tilde{\alpha}_X(u; |B|) \, du \right)
\]

which is \( o(1) \) by assumption. Thus \( \text{Var}(U^*_K(x)) \to 0 \) and (i) is proved.

The proof of (ii) and (iii) now proceeds exactly as in the proof of Theorem 2.1 in Politis and Romano (1994). Note that substituting \( \tau_{N(K)} \) in place of \( \tau_{\Lambda(K)} \) in the confidence limit is permissible asymptotically because assumption (3) implies \( \tau_{N(K)}/\tau_{\Lambda(K)} \xrightarrow{a.s.} 1 \) as \( |K| \to \infty \). □

**Proof of Theorem 2.** The proof of Theorem 2 is similar to the proof of Theorem 1 taking into account the arguments presented in Politis, Romano, and You (1993). □
Proof of Theorem 3. We will show (a) only, and in particular we will just show that \( \tilde{L}_K^*(x) \to J^*(x, P) \) in probability; showing (b) is similar.

Let \( z \) be a continuity point of \( J^*(x, P) \). Let \( \tilde{P} \) denote the joint probability law of \( H_n, \{X(t)\} \) and \( N \), and let \( \tilde{E}, \tilde{Var} \) denote the corresponding expectation and variance; also let \( P^c, E^c, Var^c \) denote conditional probability, expectation and variance respectively, where we condition on the realization of the random field \( \{X(t)\} \) and that of the Poisson process \( N \). As before, \( P \) denotes the joint probability law of \( \{X(t)\} \) and \( N \), and \( E, Var \) denote the corresponding expectation and variance.

Now let \( I(y) = 1\{\tau_{\Lambda(B+y)}(S_{K,y} - \theta) \leq z\} \) which is a new homogeneous random field for \( y \in \mathbb{R}^d \). Note that to show \( \tilde{U}_K^*(x) = J^*(x, P) + o_P(1) \), where

\[
\tilde{U}_K^*(x) = n^{-1} \int_{K_{1-c}} I(y)H_n(dy).
\]

But it is obvious that \( E^c \tilde{U}_K^*(x) = U_K^*(x) \), where \( U_K^*(x) \) was defined in the proof of Theorem 1; thus, \( \tilde{E} \tilde{U}_K^*(x) = E \{E^c \tilde{U}_K^*(x)\} = EU_K^*(x) = J_B^*(x, P) \to J^*(x, P) \). So we look at \( \tilde{V}ar(\tilde{U}_K^*(x)) \).

Observe that

\[
\tilde{E}\{(\tilde{U}_K^*(x))^2|H_n\} = \tilde{E}\{n^{-2} \int_{K_{1-c}} \int_{K_{1-c}} I(y)I(x)H_n(dy)H_n(dx)|H_n\}
\]

\[
= n^{-2} \int_{K_{1-c}} \int_{K_{1-c}} \tilde{E}\{I(y)I(x)|H_n\}H_n(dy)H_n(dx).
\]

But \( \tilde{E}\{I(y)I(x)|H_n\} = EI(y)I(x) = C(y - x) + (J_B^*(x, P))^2 \), where \( C(y - x) = \text{Cov}(I(y), I(x)) \) as defined in the proof of Theorem 1. So

\[
\tilde{E}\{(\tilde{U}_K^*(x))^2|H_n\} = n^{-2} \int_{K_{1-c}} \int_{K_{1-c}} C(y - x)H_n(dy)H_n(dx) + (J_B^*(x, P))^2.
\]

Therefore

\[
\tilde{E}(\tilde{U}_K^*(x))^2 = n^{-2} \int_{K_{1-c}} \int_{K_{1-c}} C(y - x) \frac{n^2}{|K_{1-c}|^2} dy dx + n^{-1} C(0) + (J_B^*(x, P))^2,
\]

27
where it was used that

\[ \mathcal{E} H_n(dx) = \frac{n}{|K_{1-c}|} dx, \]

and

\[ \mathcal{E} H_n(dy) H_n(dx) = \frac{n^2}{|K_{1-c}|^2} dy dx + \frac{n}{|K_{1-c}|} dx \epsilon(x)(dy). \]

Observe now that in the course of the proof of Theorem 1 we had shown that

\[ \frac{1}{|K_{1-c}|^2} \int_{K_{1-c}} \int_{K_{1-c}} C(y-x) dy dx = Var(U_K^*) = o(1). \]

Hence \( \mathcal{V}ar(\bar{U}_K^*(x)) = \mathcal{E}(\bar{U}_K^*(x))^2 - (J_L^*(x, P))^2 = o(1) \), and (i) of part (a) is proved.

\( \square \)

**Acknowledgement.** We are grateful to Elias Masry (University of California at San Diego), George Roussas (University of California at Davis), and Dimitris Gatzouras (Cambridge University) for many helpful discussions and suggestions.

---

\(^3\) It is easy to see that for any set \( A \subset K_{1-c} \) we have \( \text{Prob}_P\{H_n(A) = 1|H_n(K_{1-c}) = n\} = \frac{n!}{n!(n-l)!} p^l (1-p)^{n-l} \), where \( p = |A|/|K_{1-c}| \), from which it follows that \( \mathcal{E} H_n(A) = np, \mathcal{V}ar(H_n(A)) = np(1-p) \), and that \( \mathcal{C}ov(H_n(A), H_n(A')) = 0 \) if \( A \) and \( A' \) are disjoint; the stated facts regarding \( \mathcal{E} H_n(dx) \) and \( \mathcal{E} H_n(dy) H_n(dx) \) are then a consequence of the Poisson approximation to this binomial.

28
References


