RESAMPLING MARKED POINT PROCESSES

BY

DIMITRIS N. POLITIS, EFSTATHIOS PAPARODITIS, AND JOSEPH P. ROMANO

TECHNICAL REPORT NO. 515

May 1997

PREPARED UNDER THE AUSPICES

OF

NATIONAL SCIENCE FOUNDATION
GRANT DMS95-043709

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
Resampling marked point processes

Dimitris N. Politis
Department of Mathematics
University of California, San Diego
La Jolla, CA 92093-0112

Efstathios Paparoditis
Dept. of Mathematics and Statistics
University of Cyprus
P.O. Box 537, Nicosia, Cyprus

Joseph P. Romano
Department of Statistics
Stanford University
Stanford, CA 94305

Abstract

In many contexts, such as queueing theory, spatial statistics, etc., the data consists 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 introduce different resampling techniques for marked point processes, and show that they lead to valid large sample inferences concerning the sampling distribution of the sample mean. The proposed techniques may be viewed as variants of the well-known block-resampling methods originally designed for data observed over points on a rectangular integer lattice.

AMS 1980 subject classifications: primary 62M20; secondary 62G05.
Keywords and phrases: Bootstrap, confidence intervals, large sample inference, nonparametric estimation, point processes, Poisson process, random fields, resampling.
1 Introduction and notation

Suppose \( \{X(t), t \in \mathbb{R}^d\} \) is a homogeneous random field in \( d \) dimensions, with \( d \in \mathbb{Z}^+ \), i.e., a collection of real-valued random variables \( X(t) \) that are indexed by the continuous parameter \( t \in \mathbb{R}^d \). In the important special case where \( d = 1 \), the random field \( \{X(t)\} \) is just a continuous time, stationary stochastic process. The probability law of the random field \( \{X(t), t \in \mathbb{R}^d\} \) will be denoted by \( P_X \). We will generally assume that \( E|X(t)|^2 < \infty \) in which case homogeneity (i.e. strict stationarity) implies weak stationarity, namely that for any \( t, h \in \mathbb{R}^d, E X(t) = \mu \), and \( \text{Cov}(X(t), X(t+h)) = R(h) \), i.e., \( E X(t) \) and \( \text{Cov}(X(t), X(t+h)) \) do not depend on \( t \) at all.

Our objective is statistical inference pertaining to features of the unknown probability law \( P_X \) on the basis of data; in particular, this paper will focus on estimation of the common mean \( \mu \). For the case where the data are of the form \( \{X(t), t \in E\} \), with \( E \) being a finite subset of the rectangular lattice \( \mathbb{Z}^d \), different block-resampling techniques have been developed in the literature; see, for example, Hall (1985), Carlstein (1986), Künsch (1989), Lahiri (1991), Liu and Singh (1992), Politis and Romano (1992 a,b,c, 1993, 1994), Raïs (1992), and Sherman and Carlstein (1994, 1996). 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 homogeneous Poisson point process observable on a compact subset \( K \in \mathbb{R}^d \), and assumed independent of the random field \( \{X(t)\} \); cf. Karr (1986, 1991) for a thorough discussion on the plausibility of the Poisson assumption. So let \( N \) denote such a homogeneous Poisson process on \( \mathbb{R}^d \), independent
of \{X(t)\}, and possessing mean measure \(\Lambda\), i.e., 
\[EN(A) = \Lambda(A)\] for any set \(A \subset \mathbb{R}^d\);
note that homogeneity of the process allows us to write 
\[\Lambda(A) = \lambda|A|,\] where \(\lambda\) is a
positive constant signifying the "rate" of the process, and 
\(|\cdot|\) denotes Lebesgue measure
(volume). The point process \(N\) can then be expressed as
\[N = \sum_i \epsilon_i,\] where \(\epsilon_i\) is a
point mass at \(t\), i.e., \(\epsilon_i(A)\) is 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' \(\tilde{N} = \sum_i \epsilon_{(t_i, X(t_i))}\), i.e., the point process \(N\) where each \(t\)-point
is 'marked' by the value of \(X\) at that point.

Hence, in this paper, our objective will be interval estimation of \(\mu\) on the basis of
measurements of the value of \(X(\cdot)\) at a finite number of generally non-lattice, irregularly
spaced points \(t \in \mathbb{R}^d\). The observed 'marked point process' is then defined as the
collection of pairs \(\{(t_j, X(t_j)), j = 1, ..., N(K)\}\), where \(\{t_j\}\) are the points at which
the \(\{X(t_j)\}\) 'marks' happen to be observed; see Daley and Vera-Jones (1988), Karr
(1991), or Krickeberg (1982) for more details on marked point processes.

The paper is organized as follows: Section 2 contains some useful notions on mixing,
and some necessary background on mean estimation; in Section 3 the marked point
process "circular" bootstrap is introduced and studied, while in Section 4 the marked
point process "block" bootstrap is introduced and studied; some concluding remarks
are presented in Section 5, while all proofs are deferred to Section 6.
2 Some background and a useful lemma on mixing

The continuous parameter random field \( \{X(t), t \in \mathbb{R}^d\} \) will be assumed to satisfy a certain weak dependence condition that will be quantified in terms of mixing coefficients. Let \( \rho(\cdot, \cdot) \) denote sup-distance (i.e., the distance arising from the \( l_\infty \) norm) on \( \mathbb{R}^d \); the strong mixing coefficients of Rosenblatt (1985) are then defined as

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

where \( \mathcal{F}(E_i) \) is the \( \sigma \)-algebra generated by \( \{X(t), t \in E_i\} \). Alternatively, in a random field set-up where \( d > 1 \) it is now customary to consider mixing coefficients that in general also depend on the size (volume) of the sets considered; see e.g., Doukhan (1994). Thus define

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

Note that \( \alpha_X(k; l_1, l_2) \leq \alpha_X(k) \), and that in essence \( \alpha_X(k) = \alpha_X(k; \infty, \infty) \). A random field is said to be strong mixing if \( \lim_{k \to \infty} \alpha_X(k) = 0 \). There are many interesting examples of strong mixing random fields; see Rosenblatt (1985). However, there is a big class of random fields of great interest in spatial statistics, i.e., Gibbs (Markov) random fields in \( d > 1 \) dimensions, that are not strong mixing but instead satisfy a condition on the decay of the \( \alpha_X(k; l_1, l_2) \) coefficients; see Doukhan (1994).

Nevertheless, for our results a yet weaker notion of mixing is required. So we define the coefficients

\[
\bar{\alpha}_X(k; l) \equiv \sup\{|P(A_1 \cap A_2) - P(A_1)P(A_2)| : A_i \in \mathcal{F}(E_i), i = 1, 2, E_2 = E_1 + t, |E_1| = |E_2| \leq l, \rho(E_1, E_2) \geq k\},
\]

where the supremum now is taken over all \( t \in \mathbb{R}^d \), and over all compact and convex sets \( E_1 \subset \mathbb{R}^d \). As before, we may also define \( \bar{\alpha}_X(k) = \bar{\alpha}_X(k; \infty) \).

It is easy to see now that \( \bar{\alpha}_X(k) \leq \alpha_X(k) \), so that if the random field is \( \alpha \)-strong mixing, then it will necessarily be \( \bar{\alpha} \)-strong mixing as well, i.e., \( \lim_{k \to \infty} \bar{\alpha}_X(k) = 0 \).
other words, $\bar{\alpha}$-mixing is easier to satisfy than regular $\alpha$-mixing. Similarly, $\bar{\alpha}_X(k; l_1) \leq \alpha_X(k; l_1, l_1)$, and if the $\alpha_X(k; l_1, l_1)$ coefficients are small, then the same will be true for the $\bar{\alpha}_X(k; l_1)$ coefficients; our bootstrap results will consequently be based on this weaker notion. See Doukhan (1994), Roussas and Ioannides (1987), and Ivanov and Leonenko (1986, p. 34) for discussion and references on strong mixing coefficients.

We now give a useful lemma, its essence being that mixing properties of the continuous parameter random field are inherited by the observed marked point process; the lemma is a generalization of a result of Masry (1988) who considered renewal point processes on the real line (case $d = 1$).

**Lemma 1** Let $N_g$ be a general Poisson process (not necessarily homogeneous) on $\mathbb{R}^d$, possessing mean measure $\Lambda_g$, and assumed independent of the random field $\{X(t), t \in \mathbb{R}^d \}$. Let $E_1, E_2$ be two subsets of $\mathbb{R}^d$ such that $\rho(E_1, E_2) = k > 0$, and define $\bar{Y}_i = N_g(E_i)^{-1} \int_{E_i} X(t)N(dt)$ and $\bar{Y}_i = (\Lambda_g(E_i))^{-1} \int_{E_i} X(t)N(dt)$ for $i = 1, 2$; also assume that $E|X(t)|^p = C_p < \infty$ for some $p > 2$. Then

$$|\text{Cov}(\bar{Y}_1, \bar{Y}_2)| \leq 10C_p^{2/p}(\alpha_X(k; |E_1|, |E_2|))^{1-2/p}$$

and

$$|\text{Cov}(\bar{Y}_1, \bar{Y}_2)| \leq 10C_p^{2/p}(\alpha_X(k; |E_1|, |E_2|))^{1-2/p}.$$ 

If $E_1, E_2$ are compact, convex, and are translates of one another, i.e. if $E_1 = E_2 + t$, then we also have

$$|\text{Cov}(\bar{Y}_1, \bar{Y}_2)| \leq 10C_p^{2/p}(\bar{\alpha}_X(k; |E_1|))^{1-2/p}$$

and

$$|\text{Cov}(\bar{Y}_1, \bar{Y}_2)| \leq 10C_p^{2/p}(\bar{\alpha}_X(k; |E_1|))^{1-2/p}.$$ 

Although we will subsequently use Lemma 1 in the special case of a homogeneous Poisson process, the lemma's generality is noteworthy.
We now consider estimation of the mean of the random field \( \mu = EX(t) \) on the basis of observing \( \{X(t)\} \) for the \( t \)-points generated by the homogeneous Poisson process \( N \) over the compact, convex set \( K \subset \mathbb{R}^d \). It is natural to estimate \( \mu \) by the sample mean which—as hinted at in Lemma 1—can be defined in two asymptotically equivalent ways:

\[
\bar{X}_K = \frac{1}{\lambda |K|} \int_K X(t)N(dt),
\]

and

\[
\tilde{X}_K = \frac{1}{N(K)} \int_K X(t)N(dt);
\]

the difference between the two is division by expected or actual sample size respectively. Obviously, if \( \lambda \) is unknown, then our only practical choice is \( \tilde{X}_K \). Note that Karr (1986) presents some arguments in favor of using \( \bar{X}_K \) even if \( \lambda \) is known; however, in this paper we will study both \( \bar{X}_K \) and \( \tilde{X}_K \) in the interest of completeness.

It is immediate that \( \tilde{X}_K \) is unbiased for \( \mu \); \( \bar{X}_K \) is also unbiased as a conditioning (on \( N \)) argument shows. See, for instance, Karr (1986) where it is also shown that, under some regularity assumptions, as \( |K| \to \infty \), \( \bar{X}_K \) and \( \tilde{X}_K \) are both consistent and asymptotically normal at rate \( \sqrt{|K|} \) with the same asymptotic variance. We explicitly give Karr’s (1986) theorem below as it will be useful for our bootstrap theory.

**Theorem [Karr (1986)]** Let \( R(t) = Cov(X(0), X(t)) \), and assume that

\[
\int R(t)dt < \infty,
\]

where \( \int \) is short-hand for \( \int_{\mathbb{R}^+} \). Also assume that

\[
\frac{1}{\sqrt{|K|}} \int_K (X(t) - \mu) dt \xrightarrow{L} N \left(0, \int R(t)dt\right),
\]

as \( diam(K) \to \infty \), where \( diam(K) \) denotes the supremum of the diameters of all \( L_\infty \) balls contained in \( K \). Then, as \( diam(K) \to \infty \), we have:

\[
\sqrt{|K|} (\tilde{X}_K - \mu) \xrightarrow{D} N \left(0, \sigma^2\right), \quad \sqrt{|K|} (\bar{X}_K - \mu) \xrightarrow{D} N \left(0, \sigma^2\right)
\]

and furthermore

\[
\lim Var \left(\sqrt{|K|} \tilde{X}_K\right) = \lim Var \left(\sqrt{|K|} \bar{X}_K\right) = \sigma^2 = \int R(t)dt + \lambda^{-1} R(0).
\]

6
Note that we have taken the liberty of correcting an obvious typo in the variance formula (3.10) of Karr (1986). To verify the assumptions of Karr's theorem, the following lemma may be used.

**Lemma 2** If $|E|X(t)|^{2+\delta} < \infty$ for some $\delta > 0$, and $\bar{a}_X(k; l_1) \leq \text{const.}(1 + l_1)^{\gamma d - 1} k^{-d - \epsilon}$ for some $\epsilon > 2d/\delta$, and some $\gamma < \frac{\delta(\delta - 2d)}{2d \delta (1 + \delta)}$, then equations (1) and (2) hold true.

Different sufficient conditions for equations (1) and (2) are given in Yadrenko (1983). Nevertheless, to actually use the asymptotic normality of the sample mean to construct confidence intervals for the mean $\mu$, the asymptotic variance must be explicitly estimated. While it is relatively easy to estimate $R(0)$, and $\lambda$ is consistently estimable by $N(K)/|K|$, consistent estimation of $\int R(t) dt$ is not a trivial matter, particularly in the case of irregularly spaced data considered here.

The resampling methodology that is introduced in this paper is able to yield confidence intervals for the mean without explicit estimation of the asymptotic variance; alternatively, the resampling method may provide an estimate of the asymptotic variance to be used in connection with the asymptotic normality result of Karr (1986). The "circular" resampling methodology of the next section uses a blocking argument similar to the circular bootstrap of Politis and Romano (1992c, 1993), while the "block" resampling methodology of our section 4 employs arguments similar to the block bootstrap of Künsch (1989) and Liu and Singh (1992).
3 "Circular" bootstrap for marked point processes

Our goal will be to construct bootstrap confidence intervals for $\mu$ on the basis of observing $\{X(t)\}$ for the $t$-points generated by the Poisson process $N$ over the compact, convex set $K \subset \mathbb{R}^d$. In this section we will further assume that $K$ is a "rectangle", i.e. that $K = \{t = (t_1, \ldots, t_d) : 0 \leq t_i \leq K_i, i = 1, \ldots, d\};$ in the next section the general case of $K$ being simply convex will be addressed. Thus, the "circular" bootstrap for marked point processes is described as follows.

- Begin by imagining that $K$ is "wrapped-around" on a compact torus; in other words, we interpret the index $t$ as being modulo $K$. If $t \not\in K$, we will redefine $t = t(\text{mod } K)$, where the $i$th coordinate of vector $t(\text{mod } K)$ is $t_i(\text{mod } K_i)$. With this redefinition, we have data $X(t)$ even if $t \not\in K$.

- 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 let $l = \lfloor 1/e^d \rfloor$, where $\lfloor \cdot \rfloor$ denotes the integer part.

- Generate random points $Y_1, Y_2, \ldots, Y_l$ independent and identically distributed from a uniform distribution on $K$, and define

$$\hat{X}^* \equiv l^{-1} \left( \sum_{i=1}^l \frac{1}{\lambda |B|} \int_{B + Y_i} X(t) N(dt) \right),$$

and

$$\hat{X}^* \equiv l^{-1} \left( \sum_{i=1}^l \frac{1}{N(B + Y_i)} \int_{B + Y_i} X(t) N(dt) \right).$$

- This generation of the points $Y_1, Y_2, \ldots, Y_l$ and subsequently of $\hat{X}^*$ and $\hat{X}^*$ is governed by a probability mechanism which we will denote by $P^*$; note that the generation is performed conditionally on the marked point process data that were actually observed.

- Let $P(\sqrt{|K|}(\hat{X}_K - \mu) \leq z)$ denote the distribution function of the sample mean (centered, and normalized), and let $P^*(\sqrt{|K|}(\hat{X}^* - E^*\hat{X}^*) \leq z)$ and
$P^*(\sqrt{K}((X^* - E^*X^*) \leq x) \leq x)$ denote the conditional (given the marked point process data) distribution function of its bootstrap counterparts; $E^*$ and $\text{Var}^*$ denote expected value and variance respectively under the probability mechanism $P^*$.

An intuitive way of visualizing the construction of $\tilde{X}^*$ and $\hat{X}^*$ is to imagine a "re-tiling" of an area comparable to the rectangle $K$ by putting side-by-side the small rectangles $B + Y_i$, carrying along at the same time the t-points and their corresponding $X$-marks that the marked point process $\tilde{N}$ originally generated in $B + Y_i$; as a final step recalculate the sample mean of the "re-tiled" process to get $\tilde{X}^*$ and $\hat{X}^*$.

We are now ready to state one auxiliary and two main results.

**Theorem 1** Assume equations (1) and (2), let $E|X(t)|^{6+\delta} < \infty$, where $\delta > 0$, and that $R(0) = \text{Var}(X(t)) > 0$. Also assume that $\tilde{\alpha}_X(k; l_1) \leq \text{const.}(1 + l_1)^{\tilde{\gamma}}k^{-\tilde{\beta}}$ for some $\tilde{\beta} > 3d$, and $0 \leq \tilde{\gamma} \leq \tilde{\beta}/d$. Let $\min_i K_i \to \infty$, $\max_i K_i = O(\min_i K_i)$, and let $c = c(K) \to 0$, but in such a way that $c^d|K| \to \infty$. Then,

$$E^*\hat{X}^* = \tilde{X}_K, \quad \frac{\text{Var}^*\hat{X}^*}{\text{Var}\tilde{X}_K} \xrightarrow{p} 1,$$

and

$$\sup_x |P^*(\sqrt{K}((X^* - \tilde{X}_K) \leq x) - P(\sqrt{K}((\tilde{X}_K - \mu) \leq x)| \xrightarrow{p} 0.$$

**Remark 1** The result of Theorem 1 could be compactly expressed as

$$d_2 \left( P^*(\sqrt{K}((X^* - \tilde{X}_K) \leq x), P(\sqrt{K}((\tilde{X}_K - \mu) \leq x) \right) \xrightarrow{p} 0,$$

where $d_2 (\cdot, \cdot)$ is Mallows metric between distributions – cf. Bickel and Freedman (1981) or Shao and Tu (1995, p. 73). Convergence in $d_2$ is stronger than weak convergence, as it also implies convergence of the first two moments. Nevertheless, this compact expression would obscure the fact that we have exact equality of the first moments here, as opposed to only approximate equality as in Theorem 3 of our next section.
Remark 2 Also recall that in Lemma 2 we gave a sufficient mixing condition that 
together with our moment condition implies equations (1) and (2); however, since 
in general there exist different sufficient conditions, we followed Karr (1986) in placing 
equations (1) and (2) in the assumptions of the theorems.

Theorem 2 Under the conditions of Theorem 1 we have

\[ E^* \bar{X}^* = \bar{X}_K, \quad \frac{\text{Var}^* \bar{X}^*}{\text{Var} \bar{X}_K} \xrightarrow{P} 1, \]

and

\[ \sup_x |P^*(\sqrt|K|(\bar{X}^* - \bar{X}_K) \leq x) - P(\sqrt|K|(\bar{X}_K - \mu) \leq x)| \xrightarrow{P} 0. \]

Remark 3 Although we have stated the results only for the sample mean (in its two 
forms), an application of the \( \delta \)-method (cf. Bickel and Freedman (1981)) immediately 
shows that our resampling methodology is also valid for smooth functions of the sample 
mean. The "circular" bootstrap (and the "block" bootstrap discussed in section 4) 
can be extended to other mean-like statistical functionals as well, e.g. appropriately 
differentiable statistics (see e.g. Künsch (1989) and Liu and Singh (1992)), or general 
linear statistics (see e.g. Politis and Romano (1993)).

So far we have motivated the introduction of resampling for marked point processes 
in terms of by-passing the difficult problem of estimating the asymptotic variance \( \sigma^2 \) 
which is required in order to use the asymptotic normality of the sample mean for 
confidence intervals. Nevertheless, the bootstrap typically has a further advantage 
as compared to the asymptotic normal distribution, namely that it yields a more 
accurate distribution approximation; in other words, the bootstrap is typically "higher 
order accurate"—see e.g. Efron and Tibshirani (1993) or Shao and Tu (1995) and the 
references therein.

Nevertheless, the property of higher order accuracy will apply only to the bootstrap 
distribution of the standardized or studentized sample mean. For example, the block 
bootstrap in the case of stationary sequences in discrete time was recently shown to be
higher order accurate; see Lahiri (1991) for the standardized sample mean, and Götze and Künsch (1996) for the studentized sample mean. The same higher order accuracy property characterizes the circular bootstrap; see Politis and Romano (1992c).

Although the necessary tools (e.g. Edgeworth expansions for the distribution of the sample mean) are not yet available to prove higher order accuracy of the bootstrap in a marked point process setting, we conjecture that this higher order accuracy of the standardized or studentized sample mean indeed obtains. With this in mind we offer the following easy corollary of Theorems 1 and 2, that has to do with the bootstrap distribution of the standardized sample mean. A similar result would be true for the studentized sample mean, in which case a double (or iterated) bootstrap (cf. Hall (1992)) would be required if we intend to studentize $\bar{X}_K$ and $\bar{X}_K$ using the bootstrap variance estimator.

**Corollary 1** Under the conditions of Theorem 1 we also have

$$\sup_x \left| P^* \left( \frac{\bar{X}^* - \bar{X}_K}{\sqrt{\text{Var}^* \bar{X}^*}} \leq x \right) - P \left( \frac{\bar{X}_K - \mu}{\sqrt{\text{Var} \bar{X}_K}} \leq x \right) \right| \xrightarrow{P} 0,$$

(3)

and

$$\sup_x \left| P^* \left( \frac{\bar{X}^* - \bar{X}_K}{\sqrt{\text{Var}^* \bar{X}^*}} \leq x \right) - P \left( \frac{\bar{X}_K - \mu}{\sqrt{\text{Var} \bar{X}_K}} \leq x \right) \right| \xrightarrow{P} 0.$$

(4)
4 "Block" bootstrap for marked point processes

Now our observation region $K$ can be any compact, convex subset of $\mathbb{R}^d$; for our asymptotic results, $K$ will be assumed to expand uniformly in all directions, i.e. we will assume that $diam(K) \to \infty$. As before $diam(K)$ is the supremum of the diameters of all $l_\infty$ balls contained in $K$, and we also define $Diam(K)$ to be the infimum of the diameters of all $l_\infty$ balls that contain $K$. The "block" bootstrap for marked point processes is described as follows.

- As in the previous section, 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 $ct = (ct_1, \ldots, ct_d)$. However, since the "wrap-around" will not be used here, we define the set of 'allowed' displacements $K_{1-c} = \{y \in \mathbb{R}^d : B + y \subset K\}$, and as before we let $l = [1/c^d]$.

- Generate random points $Y_1, Y_2, \ldots, Y_l$ independent and identically distributed from a uniform distribution on $K_{1-c}$, and as before let

$$
\tilde{X}^* \equiv l^{-1} \sum_{i=1}^l \frac{1}{|B|} \int_{B+Y_i} X(t)N(dt),
$$

and

$$
\check{X}^* \equiv l^{-1} \sum_{i=1}^l \frac{1}{\lambda(B + Y_i)} \int_{B+Y_i} X(t)N(dt).
$$

This generation of the points $Y_1, Y_2, \ldots, Y_l$ and subsequently of $\tilde{X}^*$ and $\check{X}^*$ is governed by a probability mechanism which we will denote by $P^*$, with moments denoted by $E^*$, $Var^*$, etc. Note again that this generation is done conditionally on the marked point process data observed, thus $P^*$ is really a conditional probability.

- Let $P^*(\sqrt{|K|}(\tilde{X}^* - E^*\tilde{X}^*) \leq x)$ and $P^*(\sqrt{|K|}(\check{X}^* - E^*\check{X}^*) \leq x)$ denote the conditional (given the marked point process data) distribution functions of the bootstrap sample means.
We are now ready to state another set of consistency results.

**Theorem 3** Assume equations (1) and (2), let $E|X(t)|^{\delta+\delta} < \infty$, where $\delta > 0$, and that $R(0) = \text{Var}(X(t)) > 0$. Also assume that $\tilde{a}(k; l_1) \leq \text{const.}(1 + l_1)^{\tilde{q}} k^{-\tilde{\beta}}$ for some $\tilde{\beta} > 3d$, and $0 \leq \tilde{q} \leq 3d$. Let $\text{diam}(K) \to \infty$, $\text{Diam}(K) = O(\text{diam}(K))$, and let $c = c(K) \to 0$, but in such a way that $c^d|K| \to \infty$. Then,

$$\frac{\text{Var}^* \hat{X}^*}{\text{Var} \hat{X}_K} \xrightarrow{P} 1,$$

and

$$\sup_x |P^*(\sqrt{|K|} (\hat{X}^* - E^* \hat{X}^*) \leq x) - P(\sqrt{|K|} (\hat{X}_K - \mu) \leq x)| \xrightarrow{P} 0.$$

**Theorem 4** Under the conditions of Theorem 3 we have

$$\frac{\text{Var}^* \hat{X}^*}{\text{Var} \hat{X}_K} \xrightarrow{P} 1,$$

and

$$\sup_x |P^*(\sqrt{|K|} (\hat{X}^* - E^* \hat{X}^*) \leq x) - P(\sqrt{|K|} (\hat{X}_K - \mu) \leq x)| \xrightarrow{P} 0.$$

**Remark 4** Note that $E^* \hat{X}^* \neq \hat{X}_K$; instead, $E^* \hat{X}^* = \hat{X}_K + O_P(c)$ as the proof of Theorem 3 shows. Therefore in order not to introduce bias in the bootstrap distribution, it is required to center the bootstrap distribution around its bootstrap mean (thus forcing it to have mean zero as in the true distribution of the sample mean); see Lahiri (1991) and Politis and Romano (1992c, 1993) for a similar discussion in the discrete-time case.

**Remark 5** In comparing the "circular" bootstrap of section 3 to the "block" bootstrap of this section we note that the "block" bootstrap is valid even for nonrectangular, convex observation regions $K$, at the expense of having to explicitly center the bootstrap distribution—see previous Remark—which the "circular" bootstrap does automatically.
Similarly to Corollary 1, we now offer a result on the "block" bootstrap distribution of the standardized sample mean in anticipation of its possible higher order accuracy properties.

**Corollary 2** Under the conditions of Theorem 3 we also have

\[
\sup_x \left| P^\ast\left( \frac{\bar{X}^* - \bar{X}_K}{\sqrt{\text{Var}^* X^*}} \leq x \right) - P\left( \frac{\bar{X}_K - \mu}{\sqrt{\text{Var} X_K}} \leq x \right) \right| \xrightarrow{P} 0,
\]

and

\[
\sup_x \left| P^\ast\left( \frac{\bar{X}^* - \bar{X}_K}{\sqrt{\text{Var}^* X^*}} \leq x \right) - P\left( \frac{\bar{X}_K - \mu}{\sqrt{\text{Var} X_K}} \leq x \right) \right| \xrightarrow{P} 0.
\]
5 Concluding remarks

In this paper, we have introduced two different (but closely related) resampling techniques for marked point processes, and have showed that they both lead to consistent estimation of the sampling distribution of the sample mean; thus, the bootstrap estimate of sampling distribution can be effectively used for the construction of confidence intervals and hypothesis testing regarding the unknown true mean $\mu$.

The proposed techniques may be viewed as variants of the well-known "circular" and the "block" bootstrap methods originally designed for data observed over points on a rectangular integer lattice. Some further comments regarding the implementation of the methods are the following:

- For practical implementation, note that since it is possible for us to generate as many pseudo-replicates of $\bar{\hat{X}}^*$, $\bar{\bar{X}}^*$, $\bar{X}^*$, and $\bar{X}^*$ as we wish, the aforementioned "circular" and "block" bootstrap distributions $P^*$ and $P^*$, (as well as their moments $E^*$, $E^*$, $Var^*$, and $Var^*$) can be evaluated approximately by a Monte Carlo procedure simply by looking at the empirical distribution (and moments) of the corresponding generated pseudo-replicates.

- It is interesting to note that it is not necessary to take the uniform distribution on $K$ and $K_{1-c}$ respectively as the distribution used to generate the i.i.d. points $Y_1, Y_2, ..., Y_t$ that are central to the "circular" and "block" bootstrap procedures. For example, the (discrete) uniform distribution on $K \cap (hZ)^d$ and $K_{1-c} \cap (hZ)^d$ can instead be used respectively for the construction of the "circular" and "block" bootstrap for marked point processes without affecting the validity of our asymptotic results; here $hZ = \{hk; k \in Z\} = \{\ldots, -2h, -h, 0, h, 2h, 3h, \ldots\}$, and $h$ is a positive real number that is either a constant, or in general may depend on $K$ but in such a way that guarantees that the cardinality of the set $K \cap (hZ)^d$ and the cardinality of $K_{1-c} \cap (hZ)^d$ both tend to infinity as $diam(K) \to \infty$.

- Our bootstrap results are in principle generalizable to the case where the $t$-points are generated by a Poisson process $N_t$ that is not necessarily homogeneous; it is
for this reason that our Lemma 1 is stated in such general form. Nevertheless, since in effect the bootstrap for the sample mean will not work unless the sample mean is known to be asymptotically normal, a new version of Karr's Theorem would also be required that will not rely on homogeneity of the Poisson process involved; such a central limit theorem is expected to have rate $\sqrt{\Lambda_g(K)}$ as opposed to the $\sqrt{|K|}$ rate of Karr's Theorem. Also required would be appropriate restrictions on the mean measure $\Lambda_g$ of the general Poisson process $N_g$; for example, a natural requirement would be to assume that $\Lambda_g$ is absolutely continuous with respect to Lebesgue measure on $\mathbb{R}^d$, with Radon-Nikodym derivative $\lambda_g(t)$ satisfying $\lambda_g(t) > \epsilon$ for all $t$, where $\epsilon$ is a positive constant.

- In general, it is not even necessary to take $l = [1/c^d]$ for the methods to work; the simple choice $l = [1/c^d]$ is nonetheless recommended in anticipation of higher-order accuracy results – see the discussion before Corollary 1. To be more specific, our Corollaries 1 and 2 remain true as they are stated even if $l$ is taken different from $[1/c^d]$, as long as $l \to \infty$ as $\text{diam}(K) \to \infty$. However, we conjecture that – under some extra conditions – the simple choice $l = [1/c^d]$ will be required in order to have the right-hand-side of equations (3), (4), (5), (6) be of order $o_p(1/|K|)$, implying higher-order accuracy of the "circular" and "block" bootstrap.

- Last but not least is the issue of "optimally" choosing the design parameter $c$ in practice, again with the point of view of improving estimation accuracy. Although our first-order consistency results (e.g. Corollaries 1 and 2) remain true for any choice of $c = c(K)$ satisfying $c \to 0$ but $c^d|K| \to \infty$ as $\text{diam}(K) \to \infty$, it is highly plausible that there is an "optimal" choice of $c$ as a function of $K$ that will minimize the right-hand-side of equations (3), (4), (5), (6). Thus, the issues of higher-order accuracy and of "optimally" choosing $c$ are intertwined, and will be the subject of further research. Note, however, that the quest for the "optimal" $c$ ultimately depends on the optimality criterion employed, e.g., improving the accuracy of distribution estimation will generally entail a different "optimal" choice of $c$ as compared to the one required for improving the accuracy.
of variance estimation, i.e., the approximation $\text{Var}^* \hat{X}^* \simeq \text{Var} \hat{X}_K$; cf. Hall et al. (1996) for a thorough discussion on the analogous problem of "optimal" block size in the time series case.
6 Technical proofs

Proof of Lemma 1. Consider first the identity

\[ E^{N_g(E_1), N_g(E_2)} Y_1 Y_2 - E^{N_g(E_1), N_g(E_2)} \tilde{Y}_1 E^{N_g(E_1), N_g(E_2)} \tilde{Y}_2 = \]

\[ E^{N_g(E_1), N_g(E_2)} \{ E(\tilde{Y}_1 | N_g) - E(\tilde{Y}_1 | N_g) E(\tilde{Y}_2 | N_g) \} + \]

\[ + E^{N_g(E_1), N_g(E_2)} \{ E(\tilde{Y}_1 | N_g) E(\tilde{Y}_2 | N_g) \} - E^{N_g(E_1), N_g(E_2)} \tilde{Y}_1 E^{N_g(E_1), N_g(E_2)} \tilde{Y}_2, \]

where \( E^{N_g(E_1), N_g(E_2)} \tilde{Y} \) is an alternative short-hand notation for the conditional expectation \( E(\tilde{Y} | N_g(E_1), N_g(E_2)) \). Note that \( E(\tilde{Y}_1 | N_g) \) is a function of just the \( t_{1(i)}, t_{2(i)}, \ldots, t_{N_g(E_1)} \) points that were generated by the point process \( N_g \) in the set \( E_i \), for \( i = 1, 2 \).

Also note that \( t_{1(1)}, t_{2(1)}, \ldots, t_{N_g(E_1)} \) and \( t_{1(2)}, t_{2(2)}, \ldots, t_{N_g(E_2)} \) are two different collections of \( t \)-points, i.e., there are no common \( t \)-points to both collections. Thus, by the Poisson assumption, \( E(\tilde{Y}_1 | N_g) \) and \( E(\tilde{Y}_2 | N_g) \) are conditionally (given \( N_g(E_1), N_g(E_2) \)) independent (being functions of the distinct, independent sets of \( t \)-points \( t_{1(1)}, t_{2(1)}, \ldots, t_{N_g(E_1)} \) and \( t_{1(2)}, t_{2(2)}, \ldots, t_{N_g(E_2)} \) respectively). Hence,

\[ E^{N_g(E_1), N_g(E_2)} \{ E(\tilde{Y}_1 | N_g) E(\tilde{Y}_2 | N_g) \} = \]

\[ = E^{N_g(E_1), N_g(E_2)} \{ E(\tilde{Y}_1 | N_g) \} E^{N_g(E_1), N_g(E_2)} \{ E(\tilde{Y}_2 | N_g) \} \]

\[ = E^{N_g(E_1), N_g(E_2)} \tilde{Y}_1 E^{N_g(E_1), N_g(E_2)} \tilde{Y}_2, \]

where it was used that \( E^{N_g(E_1), N_g(E_2)} \{ E(\tilde{Y}_1 | N_g) \} = E^{N_g(E_1), N_g(E_2)} \tilde{Y}_1 \), since the \( \sigma \)-algebra generated by the random variables \( N_g(E_1), N_g(E_2) \) is coarser than that generated by the whole process \( N_g \).

Thus we have shown that

\[ E^{N_g(E_1), N_g(E_2)} Y_1 Y_2 - E^{N_g(E_1), N_g(E_2)} \tilde{Y}_1 E^{N_g(E_1), N_g(E_2)} \tilde{Y}_2 = \]

\[ E^{N_g(E_1), N_g(E_2)} \{ E(\tilde{Y}_1 \tilde{Y}_2 | N_g) - E(\tilde{Y}_1 | N_g) E(\tilde{Y}_2 | N_g) \}. \quad (7) \]

Now note that \( E^{N_g(E_1), N_g(E_2)} \tilde{Y}_i = E^{N_g(E_i)} \tilde{Y}_i \), and that \( E^{N_g(E_i)} \tilde{Y}_i \) is a function of the random variable \( N_g(E_i) \) 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 $N_g(E_1), N_g(E_2)$ are independent. Hence,

$$E\left\{ E^{N_g(E_1)} E^{N_g(E_2)} \mid Y_1, Y_2 \right\} = E\left\{ E^{N_g(E_1)} Y_1 E^{N_g(E_2)} Y_2 \right\} = E\left\{ E^{N_g(E_1)} Y_1 \right\} E\left\{ E^{N_g(E_2)} Y_2 \right\} = EY_1 EY_2.$$  

Therefore, taking expectations on both sides of (7) we finally arrive at the relation

$$EY_1 Y_2 - EY_1 EY_2 = E\{E(Y_1 Y_2|N_g) - E(Y_1|N_g)E(Y_2|N_g)\}.$$  

(8)

Conditionally on $N_g$, $Y_i$ is just a sum of the $N_g(E_i)$ random variables $X(t)$ with indices that happen to be in $E_i$. But, for any fixed (conditionally) set of points $\{t_k, k = 1, 2, \ldots\}$, Minkowski’s inequality coupled with the assumed homogeneity of the $X$-process yields

$$E^{N_g|N_g(E_i)|^{-1}} \sum_{k=1}^{N_g(E_i)} X(t_k) = E^{N_g|X(t)|^p} = E\{X(t)\}^p,$$  

(9)

by the independence of the $X$-process to $N_g$. Now by a well known mixing inequality (see e.g. Roussas and Ioannidis (1987)) we have that

$$|E(Y_1 Y_2|N_g) - E(Y_1|N_g)E(Y_2|N_g)| \leq 10C^{2/p}_p(\alpha_X(k; E_1, E_2))^{1-2/p}.$$  

Taking expectations (with respect to $N_g$) on the above completes the first assertion of the lemma.

To prove the bound concerning the covariance of $Y_1$ and $Y_2$, the exact same arguments apply only that equation (9) is replaced by the following:

$$E^{N_g|\Lambda_g(E_i)|^{-1}} \sum_{k=1}^{N_g(E_i)} X(t_k) \leq \left( \frac{N_g(E_i)}{\Lambda_g(E_i)} \right)^p E\{X(t)\}^p.$$  

Therefore, using the above bound we obtain

$$|E(Y_1 Y_2|N_g) - E(Y_1|N_g)E(Y_2|N_g)| \leq \frac{N_g(E_1)N_g(E_2)}{\Lambda_g(E_1)\Lambda_g(E_2)} 10C^{2/p}_p(\alpha_X(k; E_1, E_2))^{1-2/p}.$$  

Taking expectations (with respect to $N_g$), and using the independence of $N_g(E_1)$ and $N_g(E_2)$ and that $EN_g(E_i) = \Lambda_g(E_i)$ completes the second assertion of the lemma.
Finally, the last two assertions of the lemma involving $\tilde{\alpha}$-mixing follow similarly to the ones we have proved involving $\alpha$-mixing.

**Proof of Lemma 2.** Due to Theorem 1.7.1 of Ivanov and Leonenko (1986) that actually employs a yet weaker notion than our $\tilde{\alpha}$-mixing, we just need to verify the finiteness of $\int R(t)dt$ which by necessity will then be equal to the limiting variance of $\frac{1}{\sqrt{|K|}} \int_K X(t)dt$; since a covariance kernel is nonnegative definite, the improper integral $\int R(t)dt$ exists, it is nonnegative, but it may be infinite. Note also that, if we are not going to divide by $\int R(t)dt$ to produce a standardized (or uniform) central limit theorem, it is not necessary to prove or assume that $\int R(t)dt \neq 0$.

Note that by the well known mixing inequality (see e.g. Roussas and Ioannidis (1987)) we have that $|\text{Cov}(X(0), X(t))| \leq \text{const.} \tilde{\alpha}_X(\max_i |t_i|; 0)^{1-2/(2+\delta)}$. Thus we have

$$\int |R(t)|dt = O \left( \int \tilde{\alpha}_X(\max_i |t_i|; 0)^{1-2/(2+\delta)} dt \right)$$

$$= O \left( \int_0^\infty y^{d-1} \left( \frac{1}{y^{d+\epsilon}} \right)^{1-2/(2+\delta)} dy \right) = O \left( \int_0^\infty \frac{1}{y^{1-2d(2+\delta)+1\epsilon+1-2/(2+\delta)}} dy \right) < \infty,$$

where the assumed bound on the $\tilde{\alpha}_X$-coefficients was used, together with the assumption $\epsilon > 2d/\delta$.

**Proof of Theorem 1.** Note first that $\bar{X}$ is an average of i.i.d. random variables, each one being distributed as $\frac{1}{|B|} \int_{B+Y} X(t)N(dt)$, where $Y$ has the uniform distribution on $K$.

Therefore,

$$E^* \bar{X}^* = E^* \frac{1}{|B|} \int_{B+Y} X(t)N(dt) = \int_K \int_{B+Y} \frac{1}{|K| |B|} X(t)N(dt) dY = \bar{X}_K.$$

Similarly,

$$\text{Var}^* \bar{X}^* = l^{-1} \text{Var}^* \left( \frac{1}{|B|} \int_{B+Y} X(t)N(dt) \right)$$
and

\[ Var^* \left( \frac{1}{|B|} \int_{B+Y} X(t)N(dt) \right) = \frac{1}{|K|} \int_K \left( \frac{1}{|B|} \int_{B+Y} X(t)N(dt) - \bar{X}_K \right)^2 dY \]

\[ = \frac{1}{|K|} \int_K \left( \frac{1}{|B|} \int_{B+Y} X(t)N(dt) - \mu - (\bar{X}_K - \mu) \right)^2 dY = A_1 + A_2 + A_3, \]

where

\[ A_1 = \frac{1}{|K|} \int_K \left( \frac{1}{|B|} \int_{B+Y} X(t)N(dt) - \mu \right)^2 dY \]

\[ A_2 = \frac{2}{|K|} \int_K \left( \frac{1}{|B|} \int_{B+Y} X(t)N(dt) - \mu \right) (\bar{X}_K - \mu) dY \]

and

\[ A_3 = \frac{1}{|K|} \int_K (\bar{X}_K - \mu)^2 dY. \]

Now note that under the assumed conditions, Karr's (1986) theorem implies that \( \sqrt{|K|}(\bar{X}_K - \mu) \xrightarrow{d} N(0, \sigma^2) \), where \( \sigma^2 = \int R(t)dt + R(0)/\lambda \). Observe also that our assumption \( R(0) = Var(X(t)) > 0 \) implies that \( \sigma^2 > 0 \) since \( \int R(t)dt \geq 0 \) by nonnegative definiteness of \( \{R(t)\} \).

Hence, it follows that \( \bar{X}_K - \mu = O_P(1/\sqrt{|K|}) \), and therefore \( A_3 = O_P(1/|K|) \).

Similarly, because (also by Karr's (1986) theorem) we have \( \frac{1}{|B|} \int_{B+Y} X(t)N(dt) - \mu \) \( = O_P(1/\sqrt{|B||K|}) \), it follows that \( A_2 = O_P(1/\sqrt{|B||K|}) \). Since \( A_2 \) and \( A_3 \) are asymptotically negligible, we now focus on \( A_1 \).

Let \( G(Y) \equiv \left( \frac{1}{\lambda \sqrt{|B|}} \int_{B+Y} X(t)N(dt) - \mu \sqrt{|B|} \right)^2 \), and note that due to homogeneity of the \( X \)-process and of the point process \( N \), \( G(Y) \) is itself a homogeneous random field with index \( Y \in \mathbb{R}^d \). Also note that, since the \( X \)-process has more than four finite moments, \( EG(Y)^2 < \infty \), and homogeneity yields \( Cov(G(Y_1), G(Y_2)) = C(Y_1 - Y_2) \) for some covariance function \( C(\cdot) \).

So we have:

\[ A_1 = \frac{1}{|K||B|} \int_K G(Y) dY, \]

and

\[ EA_1 = \frac{1}{|K||B|} \int_K EG(Y) dY = \frac{1}{|B|} EG(0) \]

21
by homogeneity of the $G$-field. But since $|B| \to \infty$, Karr's (1986) theorem implies that $EG(0) \to \sigma^2$, and thus $EA_1 = \frac{\sigma^2}{|B|} + o(1/|B|)$.

Now look at

$$VarA_1 = \frac{1}{|K||B|^2} \int K \int K C(Y_1 - Y_2)dY_1dY_2$$

$$= \frac{1}{|K||B|^2} \int_{\mathbb{R}^d} C(t) \frac{|K \cap (K - t)|}{|K|} dt$$

where again the homogeneity of $G$ was used.

Recall that $K$ is the rectangle $\{t = (t_1, ..., t_d) : 0 \leq t_i \leq K_i, i = 1, ..., d\}$ and $B$ is the rectangle $\{t = (t_1, ..., t_d) : 0 \leq t_i \leq cK_i, i = 1, ..., d\}$. Define "symmetrized" versions of $K$ and $B$ by

$$K^* = \{t = (t_1, ..., t_d) : |t_i| \leq K_i, \ i = 1, ..., d\}$$

and

$$B^* = \{t = (t_1, ..., t_d) : |t_i| \leq cK_i, \ i = 1, ..., d\}.$$

Now note that

$$VarA_1 = \frac{1}{|K||B|^2} \int K^* C(t) \frac{|K \cap (K - t)|}{|K|} dt$$

$$\leq \frac{1}{|K||B|^2} \int B^* |C(t)| dt + \frac{1}{|K||B|^2} \int_{K^*-B^*} |C(t)| dt.$$

But $\frac{1}{|K||B|^2} \int B^* |C(t)| dt = O(\frac{1}{|K||B|^2})$ since $|C(t)|$ is bounded by $|C(0)|$. Finally note that for $t \in K^*-B^*$, our Lemma 1 affords us the possibility of using mixing bounds for the covariance $C(t)$. Letting $p = 3$ in Lemma 1, we obtain

$$|C(t)| \leq 10(E|G(0)|^3)^{2/3}(\tilde{\alpha}_X(\rho(t, B^*); |B|))^{1/3}.$$

But from Lemma 1.8.1 of Ivanov and Leonenko (1986) it follows that $E|G(0)|^3 \leq \text{const.} \ (E|X(t) - \mu|^{6+\delta})^{6/(6+\delta)}$ which is assumed finite. Putting this all together, and noting that

$$|K|^{-1} \int_{K^*-B^*} (\tilde{\alpha}_X(\rho(t, B^*); |B|))^{1/3} dt \leq |K|^{-1} \int_{K^*} (\tilde{\alpha}_X(\rho(t, 0); |B|))^{1/3} dt$$

22
\[
\leq |K|^{-1} \int_0^{\max_i K_i} y^{d-1} (\bar{\alpha}_X(y; |B|))^{1/3} dy = O\left(\frac{|B|^{\delta/3}}{|K|} \int_0^{\max_i K_i} y^{d-1} y^{-\beta/3} dy\right) = \\
O\left(\frac{|B|^{\delta/3}}{|K|} (\max_i K_i)^{d-\beta/3}\right) = O\left(c^{\delta/3} |K|^{(\tilde{\gamma}-\beta d-1)/3}\right) = o(1),
\]
because \( c \to 0 \) and it was assumed that \( \tilde{\beta} \geq d\tilde{\gamma} \). In the above string of inequalities, the assumed bound on the \( \bar{\alpha}_X \) coefficients was used, as well as the fact that \( \max_i K_i = O(|K|^{1/4}) \) by assumption. Therefore, it follows that
\[
Var A_1 = O\left(\frac{1}{|K||B|}\right) + o\left(\frac{1}{|B|^2}\right) = o\left(\frac{1}{|B|^2}\right).
\]
To recapitulate, we have shown that
\[
IVar^* \bar{X}_* = EA_1 + Op\left(\frac{1}{|B|}\right) = \frac{\sigma^2}{|B|} + Op\left(\frac{1}{|B|}\right).
\]
Since \( |B|/|K| \to 1 \) and \( |K|Var \bar{X}_K \to \sigma^2 \), it follows that \( \frac{Var^* \bar{X}_*}{Var \bar{X}_K} \xrightarrow{P} 1 \) as claimed.

So far we have shown convergence of the first two bootstrap moments of \( \bar{X}_* \) to the corresponding moments of \( \bar{X}_K \). Since \( \bar{X}_K \) is asymptotically normal, to complete the proof of Theorem 1 we need to show that \( \bar{X}_* \) is asymptotically normal too, in \( P^* \) probability. However, recall that \( \bar{X}_* \) is an average of \( l \) i.i.d. random variables, with \( l \to \infty \). Therefore, to show that \( \bar{X}_* \) is asymptotically normal in \( P^* \) (conditional) probability, it suffices (by the Lyapunov central limit theorem) to show that \( E^* |G(Y)|^{3/2} = O_P(1) \). But \( E^* |G(Y)|^{3/2} = \frac{1}{|K|} \int_K |G(Y)|^{3/2} dY \equiv \xi \). Due to homogeneity and \( \bar{\alpha} \)-mixing, a calculation similar to the calculation of \( Var A_1 \) above, yields \( Var \xi \to 0 \). Hence, a weak law of large numbers obtains, and \( \xi \xrightarrow{P} E|G(Y)|^{3/2} \) which is finite by Lemma 1.8.1 of Ivanov and Leonenko (1986) and the assumed finiteness of \( 6 + \delta \) moments of the \( X \)-process. Therefore, \( E^* |G(Y)|^{3/2} = O_P(1) \) and the proof is completed.

**Proof of Theorem 2.** Note again that \( E^* \bar{X}_* \) is an average of \( l \) i.i.d. random variables, each one being distributed as \( \frac{1}{N(B+Y)} \int_{B+Y} X(t) N(dt) \), where \( Y \) has the uniform distribution on \( K \).
Therefore,
\[
E^* \bar{X}^* = E^* \frac{1}{N(B + Y)} \int_{B + Y} X(t)N(dt) = \\
\int_K \int_{B + Y} \frac{1}{|K|N(B + Y)} X(t)N(dt)dY = \bar{X}_K
\]
by a conditioning (on \(N\)) argument similar to the proof of the unbiasedness of \(\bar{X}_K\) as an estimator of \(\mu\) — see Karr (1986, 1991).

Furthermore, recall that under the assumed conditions, Karr’s theorem (1986) implies that \(\sqrt{|K|}(\bar{X}_K - \mu) \Rightarrow N(0, \sigma^2)\), where \(\sigma^2 = \int R(t)dt + R(0)/\lambda\).

Now due to the Poisson assumption, we have \(N(K) \to \lambda|K|\) almost surely. As a matter of fact, \(EN(K) = \text{Var}N(K) = \lambda|K|\); therefore, \(\frac{N(K)}{\lambda|K|} = 1 + O_P\left(\frac{1}{\sqrt{|K|}}\right)\). By the foregoing discussion we have that
\[
\sqrt{|K|}(\bar{X}_K - \bar{X}_K) = \sqrt{|K|}(\bar{X}_K - \mu + \mu - \bar{X}_K) = \frac{N(K) - \lambda|K|}{N(K)\lambda\sqrt{|K|}} \int_K (X(t) - \mu)N(dt) = \\
O_P(N(K)^{-1} \int_K (X(t) - \mu)N(dt)) = O_P\left(\frac{1}{\sqrt{|K|}}\right) = o_P(1)
\]
since \(\bar{X}_K - \mu = O_P\left(\frac{1}{\sqrt{|K|}}\right)\) by Karr’s (1986) result.

Consequently, \(\sqrt{|K|}\bar{X}_K\) and \(\sqrt{|K|}\bar{X}_K\) have the same asymptotic distribution, as well as the same asymptotic variance (by Karr’s (1986) theorem); thus Theorem 2 follows from Theorem 1.

**Proof of Theorem 3.** The proof is quite similar to the proof of Theorem 1; below we point out only the differences and new elements.

Note again that \(\bar{X}^*\) is an average of \(t\) i.i.d. random variables, each one being distributed as \(\frac{1}{\lambda|B|} \int_{B + Y} X(t)N(dt)\), where \(Y\) has the uniform distribution on \(K_{1-c}\).

Therefore,
\[
E^* \bar{X}^* = E^* \frac{1}{\lambda|B|} \int_{B + Y} X(t)N(dt) = \int_{K_{1-c}} \int_{B + Y} \frac{1}{\lambda|B||K_{1-c}|} X(t)N(dt)dY.
\]
Recall that
\[
E^* \bar{X}^* = \int_K \int_{B + Y} \frac{1}{\lambda|B||K|} X(t)N(dt)dY = \bar{X}_K.
\]
Hence,

\[ |K|(\bar{X}_K - \mu) - |K_{1-c}|(E^* \bar{X}^* - \mu) = \int_{K_{1-c}} \int_{B+Y} \frac{1}{\lambda |B|} (X(t) - \mu) N(dt) dY = \]

\[ = \int_{K_{1-c}} O_P(1/\sqrt{|B|}) dY = O_P(|K - K_{1-c}|/\sqrt{|B|}). \]

But \(|K_{1-c}| = (1-c)^d|K| \), and \(|K - K_{1-c}| = O(c|K|)\). Therefore,

\[ E^* \bar{X}^* = \frac{\bar{X}_K - \mu}{(1-c)^d + \mu} + O_P(c/\sqrt{|B|}) = \mu + O_P(1/\sqrt{|K|}) + O_P(c/\sqrt{|B|}) = \mu + o_P(1/\sqrt{|B|}). \]

Similarly,

\[ \text{IVar}^* \bar{X}^* = \frac{1}{|K_{1-c}|} \int_{K_{1-c}} \left( \frac{1}{\lambda |B|} \int_{B+Y} X(t) N(dt) - E^* \bar{X}^* \right)^2 dY \]

\[ = \frac{1}{|K_{1-c}|} \int_{K_{1-c}} \left( \frac{1}{\lambda |B|} \int_{B+Y} X(t) N(dt) - \mu - (E^* \bar{X}^* - \mu) \right)^2 dY = D_1 + D_2 + D_3, \]

where

\[ D_1 = \frac{1}{|K_{1-c}|} \int_{K_{1-c}} \left( \frac{1}{\lambda |B|} \int_{B+Y} X(t) N(dt) - \mu \right)^2 dY \]

\[ D_2 = \frac{2}{|K_{1-c}|} \int_{K_{1-c}} \left( \frac{1}{\lambda |B|} \int_{B+Y} X(t) N(dt) - \mu \right) \left( E^* \bar{X}^* - \mu \right) dY \]

and

\[ D_3 = \frac{1}{|K_{1-c}|} \int_{K_{1-c}} \left( E^* \bar{X}^* - \mu \right)^2 dY. \]

But necessarily \(D_3 = o_P(1/|B|)\), since we have shown that \(E^* \bar{X}^* = \mu + o_P(1/\sqrt{|B|})\).

Similarly, \(D_2 = o_P(1/|B|)\), since \(\frac{1}{|B|} \int_{B+Y} X(t) N(dt) = \mu = O_P(1/\sqrt{|B|})\).

Note now that \(D_1\) is of the exact same form as the quantity \(A_1\) in the proof of Theorem 1; the only differences are: (a) integrating/averaging takes place over \(K_{1-c}\) as opposed to \(K\), (b) \(K\) is not a rectangle anymore, but rather a general compact, convex set, and (c) we generally define "symmetrized" versions of the non-rectangular \(K_{1-c}\) and \(B\) by \(K^*_{1-c} = \text{convex hull of } \{t : |t| \in K_{1-c}\}\), and \(B^* = \text{convex hull of } \{t : |t| \in B\}\), where \(|t| \equiv (|t_1|, ..., |t_d|)\). But the same arguments used in the proof of Theorem 1 to show that \(A_1 = \frac{\sigma^2}{|B|} + o_P(1/|B|)\) can be used to show \(|B| D_1 \xrightarrow{P} \sigma^2\), which in turns implies that \(\frac{\text{Var}^* \bar{X}^*}{\text{Var} \bar{X}_K} \xrightarrow{P} 1\). Finally, the proof of asymptotic normality follows \textit{verbatim} the proof of Theorem 1.
Proof of Theorem 4. The proof follows from Theorem 3 using the same arguments used in the proof of Theorem 2.

References


