A Thinned Block Bootstrap Variance Estimation Procedure for Inhomogeneous Spatial Point Patterns

Yongtao GUAN and Ji Meng LOH

When modeling inhomogeneous spatial point patterns, it is of interest to fit a parametric model for the first-order intensity function (FOIF) of the process in terms of some measured covariates. Estimates for the regression coefficients, say $\hat{\beta}$, can be obtained by maximizing a Poisson maximum likelihood criterion. Little work has been done on the asymptotic distribution of $\hat{\beta}$ except in some special cases. In this article we show that $\hat{\beta}$ is asymptotically normal for a general class of mixing processes. To estimate the variance of $\hat{\beta}$, we propose a novel thinned block bootstrap procedure that assumes that the point process is second-order reweighted stationary. To apply this procedure, only the FOIF, and not any high-order terms of the process, needs to be estimated. We establish the consistency of the resulting variance estimator, and demonstrate its efficacy through simulations and an application to a real data example.

KEY WORDS: Block bootstrap; Inhomogeneous spatial point process; Thinning.

1. INTRODUCTION

A main interest when analyzing spatial point pattern data is to model the first-order intensity function (FOIF) of the underlying process that has generated the spatial point pattern. Heuristically, the FOIF is a function that describes the likelihood for an event of the process to occur at a given location (see Sec. 2 for its formal definition). We say a process is homogeneous if its FOIF is a constant and inhomogeneous otherwise. In practice, we often wish to model the FOIF in relation to some measured covariates. For example, for the Beilschmiedia pendula Lauraceae (BPL) data given in Section 5, we want to model the FOIF in terms of two important variables of landscape features, elevation and gradient. Because the FOIF characterizes the probability finding a BPL tree at a given location with the associated elevation and gradient values, the study of this function can yield valuable insights into how these landscape features affect the spatial distribution of BPL trees. If a significant relationship can be established, then this will provide evidence in support of the niche assembly theory, which states that different species benefit from different habitats determined by local environmental features (e.g., Waagepetersen 2007).

To model the FOIF, we assume that it can be expressed as a parametric function of the available covariates. Specifically, let $N$ denote a spatial point process defined over $\mathbb{R}^2$, with the FOIF at $s \in \mathbb{R}^2$ given by $\lambda(s; \beta)$, where $\beta$ is a $p \times 1$ vector of unknown regression coefficients associated with the covariates, and let $D$ be the region in which a realization of $N$ has been observed. Our goal is then to estimate and make inference on the regression coefficients $\beta$.

To estimate $\beta$, we consider the following maximum likelihood criterion:

$$U(\beta) = \frac{1}{|D|} \sum_{x \in D \cap N} \log \lambda(x; \beta) - \frac{1}{|D|} \int_D \lambda(s; \beta) \, ds. \quad (1)$$

The maximizer of (1) is taken as an estimator for $\beta$ (denoted by $\hat{\beta}$ throughout this section). Note that (1) is proportional to the true maximum likelihood if $N$ is an inhomogeneous Poisson process, that is, when the events of the process occurring in disjoint sets are completely independent. For the BPL example, however, the locations of the BPL trees are likely to be clustered, possibly due to seed dispersal and/or correlation among environmental factors that have not been accounted for by the model. Schoenberg (2004) showed that under some mild conditions, $\hat{\beta}$ obtained by maximizing (1) is still consistent for $\beta$ for a class of spatial–temporal point process models, even if the process is not Poisson, however, he did not provide the asymptotic distribution of $\hat{\beta}$. Waagepetersen (2007) significantly extended the scope of the method by deriving asymptotic properties of $\hat{\beta}$, including asymptotic normality for a wide class of spatial cluster processes.

To make inference on $\beta$, we need information on the distributions of $\hat{\beta}$. One standard approach to this is to derive the limiting distribution of $\hat{\beta}$ under an appropriate asymptotic framework and then use it as an approximation for the distribution of $\hat{\beta}$ in a finite-sample setting. We note that currently available asymptotics for inhomogeneous spatial point processes are inadequate because they either assume complete spatial independence, that is, the process is Poisson (e.g., Rathbun and Cressie 1994), or use a parametric model for the dependence of the process (e.g., Waagepetersen 2007). For data arising from biological studies such as the BPL data, however, the underlying biological process generating the spatial point patterns is rather complex and often not well understood. Thus using a specific form for the dependence may be debatable and could lead to incorrect inference on the regression parameters $\beta$ (because the distribution of $\hat{\beta}$ depends on the dependence structure of the process).

In this article we study the distributional properties of $\hat{\beta}$ under an increasing domain setting. To quantify the dependence of the process, we use a more flexible, model-free mixing condition (Sec. 2) but do not assume any specific parametric structure on the dependence. Our main result shows that under some mild conditions, the standardized distribution of $\hat{\beta}$ is asymptotically normal. If the variance of $\hat{\beta}$ is known, the approximate confidence intervals for $\beta$ can be obtained, so that the inference on $\beta$ becomes straightforward. Thus our result further extends the scope of applications of (1) beyond the work of Schoenberg (2004) and Waagepetersen (2007).
One complication in practice is that the variance of \( \hat{\beta} \) is unknown and must be estimated. From Theorem 1 in Section 2, we see that the variance of \( \hat{\beta} \) depends on the second-order cumulant function (SOCF) of the process, a function that is related to the dependence structure of the process. To avoid specifying the SOCF, we develop a nonparametric thinned block bootstrap procedure to estimate the variance of \( \hat{\beta} \) by using a combination of a thinning algorithm and block bootstrap. Specifically, in the thinning step, we retain (i.e., do not thin) an observed event from \( N \) with a probability that is proportional to the inverse of the estimated FOIF at the location of the event. If \( N \) is second-order reweighted stationary (see Sec. 2) and if \( \hat{\beta} \) is close to \( \beta \), then the thinned process should resemble a second-order stationary (SOS) process. In Section 3 we show that the variance of \( \hat{\beta} \) can be written in terms of the variance of a statistic defined on an SOS process, which we can accomplish using block bootstrap. In Section 3 we prove that the resulting variance estimator is \( L_2 \)-consistent for the target variance, and in Section 4 we report a simulation study done to investigate the performance of the proposed procedure. To illustrate its use in a practical setting, we also apply the proposed procedure to the BPL data in Section 5.

Before proceeding to the next section, we note that resampling methods including the block bootstrap have been extensively applied in spatial statistics (e.g., Politis 1999; Lahiri 2003). However, most of these methods were developed for stationary, quantitative spatial processes that are observed on a regularly spaced grid. In the regression setting also for quantitative processes, Cressie (1993, sec. 7.3.2) discussed both semiparametric and parametric bootstrap methods to resample the residuals, whereas Sherman (1997) proposed a subsampling approach. Politis and Sherman (2001) and McElroy and Politis (2007) considered using subsampling for marked point processes. The former assumed the point process to be stationary, whereas the latter assumed it to be Poisson. None of the aforementioned resampling procedures can be used for inhomogeneous spatial point processes due to the unique feature of the data. Note that by nature an inhomogeneous spatial point process is not quantitative, is observed at random locations, is nonstationary, and can be non-Poisson.

2. NOTATION AND PRELIMINARY ASYMPTOTIC RESULTS

Let \( N \) be a two-dimensional spatial point process observed over a domain of interest \( D \). For a Borel set \( B \subset \mathbb{R}^2 \), let \( |B| \) denote the area of \( B \) and let \( N(B) \) denote the number of events from \( N \) that fall in \( B \). We define the \( k \)-th order intensity and cumulant functions of \( N \) as

\[
\lambda_k(s_1, \ldots, s_k) = \lim_{|d|k| \to 0} \frac{E[N(ds_1) \cdots N(ds_k)]}{|ds_1| \cdots |ds_k|},
\]

\( i = 1, \ldots, k, \)

and

\[
Q_k(s_1, \ldots, s_k) = \lim_{|d|k| \to 0} \frac{\text{cum}[N(ds_1), \ldots, N(ds_k)]}{|ds_1| \cdots |ds_k|},
\]

\( i = 1, \ldots, k. \)

Here \( ds \) is an infinitesimal region containing \( s \) and \( \text{cum}(y_1, \ldots, y_k) \) is the coefficient of \( \kappa^{t_1} \cdots t_k \) in the Taylor series expansion of \( \log[E[\exp(\sum_{j=1}^k y_j)]) \) about the origin (e.g., Brillinger 1975). For the intensity function, \( \lambda_k(s_1, \ldots, s_k) |ds_1| \cdots |ds_k| \) is the approximate probability that \( ds_1, \ldots, ds_k \) each contains an event. For the cumulant function, \( Q_k(s_1, \ldots, s_k) \) describes the dependence among sites \( s_1, \ldots, s_k \), where a near-0 value indicates near independence. Specifically, if \( N \) is Poisson, then \( Q_k(s_1, \ldots, s_k) = 0 \) if at least two of \( s_1, \ldots, s_k \) are different.

We study the large-sample behavior of \( \hat{\beta} \) under an increasing-domain setting, where \( \hat{\beta} \) is obtained by maximizing (1). Specifically, consider a sequence of regions, \( D_n \). Let \( \partial D_n \) denote the boundary of \( D_n \), let \( |\partial D_n| \) denote the length of \( \partial D_n \), and let \( \hat{\beta}_n \) denote \( \hat{\beta} \) obtained over \( D_n \). We assume that

\[
C_1n^2 \leq |D_n| \leq C_2n^2, \quad C_1n \leq |\partial D_n| \leq C_2n
\]

for some \( 0 < C_1 \leq C_2 < \infty \). (2)

Condition (2) requires that \( D_n \) must become large in all directions (i.e., the data are truly spatial) and that the boundary must not be too irregular. Many commonly used domain sequences satisfy this condition. To see an example of this, let \( A \subset (0, 1) \times (0, 1) \) be the interior of a simple closed curve with nonempty interior. If we define \( D_n \) as \( A \) inflated by a factor \( n \), then \( D_n \) satisfies condition (2). Note that this formulation incorporates a wide variety of shapes, including rectangular and elliptical shapes.

To formally state the large-sample distributional properties of \( \hat{\beta}_n \), we need to quantify the dependence in \( N \). We do this by using the model-free strong mixing coefficient (Rosenblatt 1956), defined as

\[
\alpha(p; k) \equiv \sup \{|P(A_1 \cap A_2) - P(A_1)P(A_2)| : A_1 \in \mathcal{F}(E_1), A_2 \in \mathcal{F}(E_2), E_2 = E_1 + s, |E_1| = |E_2| \leq p, d(E_1, E_2) \geq k\}.
\]

Here the supremum is taken over all compact and convex subsets \( E_1 \subset \mathbb{R}^2 \) and over all \( s \in \mathbb{R}^2 \) such that \( d(E_1, E_2) \geq k \), where \( d(E_1, E_2) \) is the maximal distance between \( E_1 \) and \( E_2 \) (e.g., Guan, Sherman, and Calvin 2006), and \( \mathcal{F}(E) \) is the \( \sigma \)-algebra generated by the random events of \( N \) that are in \( E \). We assume the following mixing condition on \( N \):

\[
\sup_p \alpha(p; k) \cdot \frac{p}{p} = O(k^{-\epsilon}) \quad \text{for some} \ \epsilon > 2. \quad (3)
\]

Condition (3) states that for any two fixed sets, the dependence between them must decay to 0 at a polynomial rate of the inter-set distance, \( k \). The speed in which the dependence decays to 0 also depends on the size of the sets (i.e., \( p \)). In particular, for a fixed \( k \), condition (3) allows the dependence to increase as \( p \) increases. Any point process with a finite dependence range, such as the Matérn cluster process (Stoyan and Stoyan 1994), satisfies this condition. Furthermore, it is also satisfied by the log-Gaussian Cox process (LGCP), which is very flexible in modeling environmental data (e.g., Möller, Sørbye, and Waagepetersen 1998), if the correlation of the underlying
Gaussian random field decays at a polynomial rate faster than $2 + \epsilon$ and has a spectral density bounded below from 0. This is due to corollary 2 of Doukhan (1994, p. 59).

In addition to conditions (2) and (3), we also need some mild conditions on the intensity and cumulant functions of $N$. In what follows, let $f^{(i)}(\beta)$ denote the $i$th derivative of $f(\beta)$. We assume that

$$
\lambda(s; \beta) \text{ is bounded below from 0,}
$$
(4)

$$
\lambda^{(2)}(s; \beta) \text{ is bounded and continuous with respect to } \beta,
$$
(5)

and

$$
\sup_{s_i} \int \cdots \int |Q_k(s_1, \ldots, s_k)| ds_2 \cdots ds_k < C
$$
for $k = 2, 3, 4$. (6)

Conditions (4) and (5) are straightforward conditions that can be checked directly for a proposed FOIF model. Condition (6) is a fairly weak condition. It also requires the process $N$ to be weakly dependent, but from a different perspective from (3). In the homogeneous case, (6) is implied by Brillinger mixing, which holds for many commonly used point process models, such as the Poisson cluster process, a class of doubly stochastic Poisson process (i.e., Cox process), and certain renewal process (e.g., Heinrich 1985). In the inhomogeneous case, $Q_k(s_1, \ldots, s_k)$ often can be written as $\lambda(s_1) \cdots \lambda(s_k)\psi_k(s_2 - s_1, \ldots, s_k - s_1)$, where $\psi_k(\cdot)$ is a cumulant function for some homogeneous process. Then (6) holds if $\lambda(s)$ is bounded and $\int \cdots \int |\psi_k(u_2, \ldots, u_k)| du_2 \cdots du_k < C$. Processes satisfying this condition include, but are not limited to, the LGCP, the inhomogeneous Neyman–Scott process (INSP; Waagepetersen 2007), and any inhomogeneous process obtained by thinning a homogeneous process satisfying this condition.

**Theorem 1.** Assume that conditions (2)–(6) hold and that $\hat{\beta}_n$ converges to $\beta_0$ in probability, where $\beta_0$ is the true parameter vector. Then

$$
|D_n|^{1/2}(\Sigma_n)^{-1/2}(\hat{\beta}_n - \beta_0) \xrightarrow{d} N(0, I_p),
$$
where $I_p$ is a $p \times p$ identity matrix,

$$
\Sigma_n = |D_n|^{-1}(A_n)^{-1}B_n(A_n)^{-1},
$$

$$
A_n = \int_{D_n} \lambda^{(1)}(s; \beta_0)\lambda^{(1)}(s_1; \beta_0)\, ds,
$$

and

$$
B_n = A_n + \int_{D_n} \lambda^{(1)}(s_1; \beta_0)\lambda^{(2)}(s_2; \beta_0)\, Q_2(s_1, s_2) ds_1 ds_2.
$$

**Proof.** See Appendix A.

We assume in Theorem 1 that $\hat{\beta}_n$ is a consistent estimator for $\beta_0$. Schoenberg (2004) established the consistency of $\hat{\beta}_n$ for a class of spatial–temporal processes under some mild conditions. He assumed that the spatial domain was fixed but that the time domain increased to infinity. His results can be directly extended to our case. Therefore, we simply assume consistency in Theorem 1. In connection with previous results, our asymptotic result coincides with that of Rathbun and Cressie (1994) in the inhomogeneous spatial Poisson process case and Waagepetersen (2007) in the INSP case. Furthermore, note that $\Sigma_n$ depends only on the first- and second-order properties of the process.

### 3. THINNED BLOCK BOOTSTRAP

#### 3.1 The Proposed Method

Theorem 1 states that $|D_n|^{1/2}(\Sigma_n)^{-1/2}(\hat{\beta}_n - \beta_0)$ converges to a standard multivariate normal distribution as $n$ increases. This provides the theoretical foundation for making inference on $\beta$. In practice, however, $\Sigma_n$ is typically unknown and thus must be estimated. From the definition of $\Sigma_n$ in Theorem 1, we see that two quantities (i.e., $A_n$ and $B_n$) need to be estimated to estimate $\Sigma_n$. Note that $A_n$ depends only on the FOIF and thus can be estimated easily once a FOIF model has been fitted to the data, however, the quantity $B_n$ also depends on the SOCF $Q_2(\cdot)$. Unless an explicit parametric model is assumed for $Q_2(\cdot)$, which (as argued in Sec. 1) may be unrealistic, it is difficult to directly estimate $B_n$. In this section we propose a thinned block bootstrap approach to first estimate a term that is related to $B_n$, and then use it to produce an estimate for $B_n$.

From now on, we assume that the point process $N$ is second-order reweighted stationary (SORWS); that is, there exists a function $g(\cdot)$ defined in $\mathbb{R}^2$ such that $\lambda_2(s_1, s_2) = \lambda(s_1)\lambda(s_2)g(s_1 - s_2)$. The function $g(\cdot)$ is often referred to as the pair correlation function (PCF; e.g., Stoyan and Stoyan 1994). Note that our definition of second-order reweighted stationarity is a special case of the original definition given by Baddeley, Møller, and Waagepetersen (2000). Specifically, these two definitions coincide if the PCF exists. Examples of SORWS processes include the INSP, the LGCP, and any point process obtained by thinning a standard multivariate normal distribution as $n$ increases.

The essence of our approach is based on the fact that an SORWS point process can be thinned to be SOS by applying some proper thinning weights to the events (Guan 2007). Specifically, we consider the following thinned process:

$$
\Psi = \left\{ x : x \in N \cap D_n, P(x \text{ is retained}) = \frac{\min_{s \in D_n} \lambda(s)}{\lambda(x)} \right\}.
$$
(7)

Clearly, $\Psi$ is SOS, because its first- and second-order intensity functions can be written as

$$
\lambda_n = \min_{s \in D_n} \lambda(s) \quad \text{and} \quad \lambda_{2,n}(s_1, s_2) = (\lambda_n)^2g(s_1 - s_2),
$$
where $s_1, s_2 \in D_n$. Based on $\Psi$, we then define the following statistic:

$$
S_n = \sum_{x \in \Psi \cap D_n} \lambda^{(1)}(x).
$$
(8)

Note that $Q_2(s_1, s_2) = \lambda(s_1)\lambda(s_2)[g(s_1 - s_2) - 1]$ because $N$ is SORWS. Thus the covariance matrix of $S_n$, where $S_n$ is defined in (8), is given by

$$
\text{cov}(S_n) = \lambda_n \int_{D_n} \lambda^{(1)}(s) \left[ \lambda^{(1)}(s) \right]^T \, ds
$$
where $\mathbb{R}^2 \backslash D_n = \{s: s \in \mathbb{R}^2 \text{ but } s \notin D_n\}$. Condition (12) is only slightly stronger than $\int_{\mathbb{R}^2} |g(s) - 1| \, d\mathbf{s} < C$, which is implied by condition (6). In the case where the $D_n$’s are sufficiently regular (e.g., a sequence of $n \times n$ squares) such that $(|D_n + s|) \cap \mathbb{R}^2 \backslash D_n < C|s||D_n|^{1/2}$, then condition (12) is implied by the condition $\int_{\mathbb{R}^2} ||g(s) - 1|| \, d\mathbf{s} < C$, which holds for any process that has a finite range of dependence. Furthermore, it holds for the LGCP if the correlation function of the underlying Gaussian random field generating the intensities decays to 0 at a rate faster than $|s|^{-3}$.

**Theorem 2.** Assume that conditions (2), (4), (5), and (6) hold. Then

$$\frac{1}{|D_n|} \text{cov}(\hat{\theta}_m) \leq \frac{1}{|D_n|} \text{cov}(\theta_m).$$

If we further assume (12), then $\frac{1}{|D_n|} E[|\text{cov}(\hat{\theta}_m) - \text{cov}(\theta_m)|^2] \leq C[1/k_n + 1/|D_i|]$ for some $C < \infty$.

**Proof.** See Appendix B.

Theorem 2 establishes the $L_2$ consistency of the proposed covariance matrix estimator for a wide range of values for the number of thinned processes and the subblock size. But to apply the proposed method in practice, these values must be determined. For the former, the proposed thinning algorithm can be applied as many times as necessary until a stable estimator can be obtained. For the latter, Theorem 2 suggests that the “optimal” rate for the subblock size is $|D_n|^{1/2}$, where the word “optimal” is in the sense of minimizing the mean squared error. This result agrees with the findings for the optimal subblock size for other resampling variance estimators obtained from quantititative processes (e.g., Sherman 1996). If we further assume that the best rate can be approximated by $c_n = c|D_n|^{1/2}$, where $c$ is an unknown constant, then we can adapt the algorithm of Hall and Jing (1996) to estimate $c_n$ and thus determine the “optimal” subblock size.

Specifically, let $D_m^i, i = 1, \ldots, k_n$, be a set of subblocks contained in $D_n$ that have the same size and shape, let $S_{m,j}$ be $S_n$ in (8) obtained from the $j$th thinned process on $D_m^i$, and let $S_{m,j}$ be the mean of $S_{m,j}$. We then define the sample covariance matrix of $S_{m,j}$ averaged over the $K$ replicate thinned point processes,

$$\hat{\theta}_m = \frac{1}{K} \sum_{i=1}^{k_n} \sum_{j=1}^{k_n} (S_{m,j} - \bar{S}_{m,j})(S_{m,j} - \bar{S}_{m,j})'.$$

If $D_m^i$ is relatively small compared with $D_n$, then the foregoing should be a good estimator for the true covariance matrix of $S_{m,j}$, where $S_{m,j}$ is $S_n$ in (8) defined on $D_n$. For each $D_m^i$, we then apply (11) to estimate $\text{cov}(\hat{\theta}_m)$ over a fine set of candidate subblock sizes, say $c_m = c|D_m^i|^{1/2}$. Let $\hat{\theta}_m^j(j,k)$ and $[\hat{\theta}_m(j,k)]'$ be the $(j,k)$th element of $\hat{\theta}_m$ and the resulting estimator for using $c_m$. We define the best $c_m$ as the one that has the smallest value for the following criterion:

$$M(c_m) = \frac{1}{k_n} \sum_{i=1}^{k_n} \sum_{j=1}^{k_n} \sum_{k=1}^{k_n} [\hat{\theta}_m^j(j,k) - \hat{\theta}_m(j,k)]'^2.$$

The “optimal” subblock size for $D_n$, $c_n$, then can be estimated easily using the relationship $c_n = c_m(|D_n|/|D_m|)^{1/2}$.
4. SIMULATIONS

4.1 Log-Gaussian Cox Process

We performed a simulation study to test the results of our theoretical findings. On an observation region \( D \) of size \( L \times L \), where \( L = 1, 2, \) and \( 4 \), we simulated 1,000 realizations of an inhomogeneous Poisson process with intensity function \( \lambda(s) \) given by

\[
\log \lambda(s) = \alpha + \beta X(s) + G(s), \tag{13}
\]

where \( G \) is a mean-0 Gaussian random field. This is the LGCP model, with inhomogeneity due to the covariate \( X \).

For values of \( X \), we used a single realization from a different mean-0 Gaussian random field. We kept this realization fixed throughout the whole simulation study; that is, we used the same \( X \) for each of the 1,000 point process realizations. Figure 1 shows a surface plot of \( X \) for \( L = 4 \). The values of \( X \) for \( L = 1 \) and 2 are the lower-left corner of those shown in the figure. For \( G \), we used the exponential model \( C(h) = \sigma^2 \exp(-h/\rho) \) for its covariance function, where \( \rho = .1 \) and .2 and \( \sigma = .1 \). We set \( \alpha \) and \( \beta \) to be 7.02 and 2. We generated the point process patterns using the rpoisson function in the spatstat R package (Baddeley and Turner 2005). The average number of points simulated were about 1200, 4,500, and 16,700 for \( L = 1, 2, \) and 4.

For each of the 1,000 point patterns, we obtained the estimates \( \hat{\alpha} \) and \( \hat{\beta} \) by maximizing the Poisson likelihood (using the ppm function in spatstat). Next, we thinned the point pattern according to the criterion described in Section 3, resulting in 20 independent thinned processes for each realization. The average number of points in the thinned patterns was roughly 490, 1,590, and 5,400 for \( L = 1, 2, \) and 4, so about 30–40% of the points were retained. We then applied the bootstrap procedure to each of the thinned processes to get the variance estimates for \( \hat{\alpha} \) and \( \hat{\beta} \). From (13), we found that the quantities to be computed with each bootstrap sample were \( \sum \hat{\lambda}(s^*) \) for \( \alpha \) and \( \sum X(s^*) \hat{\lambda}(s^*) \) for \( \beta \), where \( s^* \) represents the locations of the points in a bootstrap sample. We used nonoverlapping square blocks in the bootstrap procedure. For \( L = 1 \), we used blocks with side lengths \( 1/4, 1/3, \) and \( 1/2 \), whereas for \( L = 2, 4 \), we used blocks with side lengths \( 1/4, 1/3, 1/2, \) and 1. The number of bootstrap samples was 499.

We computed the variances of \( \hat{\alpha} \) and \( \hat{\beta} \) from the bootstrap variances using Theorem 1, (9), and the relationship of \( \text{cov}(S_n) \) to \( B_n \). We averaged the variance estimates across the independent thinnings. Finally, we constructed nominal 95% confidence intervals, yielding a total of 1,000 confidence intervals for \( \alpha \) and \( \beta \).

Table 1 gives the empirical coverage of the confidence intervals for \( \beta \). The coverage for \( \alpha \) (not shown) is similar but slightly lower. For both values of \( \rho \), we found that the empirical coverage increased toward the nominal level as the observation region size increased. This increase in coverage agrees with what we would expect from the theory. The empirical coverage for \( \rho = .1 \) is typically closer to the nominal level than for \( \rho = .2 \). Note that \( \rho = .1 \) corresponds to a weaker dependence. Thus we would expect our method to work better if the dependence in the data were relatively weaker. It is interesting to note that for \( L = 2 \) and 4, the best coverage for \( \rho = .2 \) was achieved by using a block larger than its counterpart for \( \rho = .1 \). Thus for errors with longer-range dependence, it appears that the block bootstrap procedure requires larger blocks to work well.

Table 2 gives the bootstrap estimates of the standard errors of \( \hat{\beta} \) for \( \rho = .2 \), averaged over the 1,000 realizations. The standard deviations of these estimates over the 1,000 realizations are included in brackets. The “true” standard errors computed using the estimated regression parameters from the independent realizations are also given. The table shows that the bootstrap estimates have a negative bias, becoming proportionally closer to the “true” standard errors as the region size \( L \) increases.

In summary, we found that our method worked reasonably well when the regression parameters were estimated. We found a slight undercoverage in each instance, but noted that this undercoverage becomes smaller as the sample size increases. We also estimated the standard errors by using the true regression parameters in our thinning step, and found that the coverage increased only slightly.

Table 1. Empirical coverage of 95% nominal confidence intervals of \( \beta \) in the LGCP case

<table>
<thead>
<tr>
<th>Region size</th>
<th>1/4</th>
<th>1/3</th>
<th>1/2</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \rho = .1 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>.85</td>
<td>.77</td>
<td>.49</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>.91</td>
<td>.90</td>
<td>.88</td>
<td>.49</td>
</tr>
<tr>
<td>4</td>
<td>.93</td>
<td>.93</td>
<td>.93</td>
<td>.90</td>
</tr>
<tr>
<td>( \rho = .2 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>.84</td>
<td>.76</td>
<td>.50</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>.87</td>
<td>.88</td>
<td>.87</td>
<td>.50</td>
</tr>
<tr>
<td>4</td>
<td>.88</td>
<td>.89</td>
<td>.90</td>
<td>.89</td>
</tr>
</tbody>
</table>

Figure 1. Image of the covariate \( X \) used in the simulation. The range of \( X \) is between \(-.6066 \) and \(.5303 \), where darker colors represent larger values of \( X \).
4.2 Inhomogeneous Neyman–Scott Process

Theorem 1 gives the variance in terms of the FOIF and the SOCF. Another approach for variance estimation is to use the plug-in method. As point out by one referee, we may simply estimate the PCF or the $K$-function using the original data (Baddeley et al. 2000). This will yield an estimate of the SOCF that in turn can be plugged into the expression for $B_0$ in Theorem 1 to obtain an estimate for the asymptotic variance. Waagepetersen (2007) studied the performance of this method in the INSP case.

To compare the performance of our method with the plug-in method, we also simulated data from the INSP model given by Waagepetersen (2007). To do so, we first simulated a homogeneous Poisson process with intensity $\kappa = 50$ as the parent process. For each parent, we then generated a Poisson number of offspring. We defined the position of each offspring relative to its parent by a radially symmetric Gaussian random variable (e.g., Diggle 2003). Let $\omega$ denote the standard deviation of the variable. We used $\omega = .02, .04$, representing relatively strong and weak clustering. Finally, we thinned the offspring as done by Waagepetersen (2007) by setting the probability of retaining an offspring equal to the intensity at the offspring location divided by the maximum intensity in the study region. For the intensity, we used $\log \lambda(s) = \alpha + \beta X(s)$, where $\alpha$, $\beta$, and $X(s)$ were as defined in the LGCP case.

We simulated 1,000 realizations of the process on both a $1 \times 1$ square and a $2 \times 2$ square. For each realization, we applied the thinned block bootstrap and the plug-in method of Waagepetersen (2007) to estimate the second-order parameters, $\kappa$ and $\omega$, by a minimum contrast estimation procedure. Specifically, we obtained these estimates by minimizing

$$\int_0^d [\hat{K}(t) - K(t; \kappa, \omega)]^2 dt, \tag{14}$$

with respect to $(\kappa, \omega)$ for a specified $a$ and $c$, where $\hat{K}(t)$ and $K(t; \kappa, \omega)$ are the empirical and theoretical $K$ functions. We used $a = 4\omega$ and $c = .25$, then used the estimated values of $\kappa$ and $\omega$ to estimate the standard errors of $\hat{\alpha}$ and $\hat{\beta}$. To do this, we used the inhom.thomas.asympcov function in the InhomCluster R package, which is available on Waagepetersen’s website at http://www.math.aau.dk/~rw/spccode/.

\begin{table}[h]
\centering
\caption{Empirical coverage of 95\% nominal confidence intervals of $\beta$ obtained by the thinned block bootstrap method and the plug-in method in the INSP case}
\begin{tabular}{|c|c|c|c|c|}
\hline
Region size & Plug-in & 1/4 & 1/3 & 1/2 \\
\hline
$\omega = .02$ & & & & \\
1 & .96 & .91 & .89 & .81 \\
2 & .96 & .94 & .94 & .94 \\
$\omega = .04$ & & & & \\
1 & .93 & .87 & .87 & .77 \\
2 & .95 & .90 & .90 & .91 \\
\hline
\end{tabular}
\end{table}
We reanalyzed the data, because the INSP model was only a “crude” model for the data, as pointed out by Waagepetersen (2007). The INSP model attributed the possible clustering among the BLP trees to a one-round seed-dispersal process. In reality, however, the clustering might be due to many different factors (e.g., important environmental factors other than elevation and gradient) that were not included in the model. Even if the clustering was due to seed dispersal alone, it probably was due to seed dispersal that had occurred in multiple generations. As a result, the validity of the model, and the subsequent estimated standard errors and inference on the regression parameters, required further investigation. Our proposed thinned blocking bootstrap method, on the other hand, requires much weaker conditions and thus might yield more reasonable estimates for the standard errors and more reliable inference on the regression parameters.

To estimate the standard errors for $\hat{\beta}_1$ and $\hat{\beta}_2$, we applied (12) with $K = 100$ and $B = 999$ and used $200 \times 100$ m sub-blocks. We chose $K = 100$ because the trace plots for the estimated standard errors became fairly stable if 100 or more thinned realizations were used. We selected the subblock size from using the data-driven procedure discussed at the end of Section 3. Table 4 gives the estimated standard errors for $\hat{\beta}_1$ and $\hat{\beta}_2$, using the thinned block bootstrap and the plug-in method of Waagepetersen (2007), along with the respective 95% confidence intervals based on these estimated standard errors. Note that the intervals based on the thinned block bootstrap are slightly shorter than those from the plug-in method, but nevertheless lead to the same conclusion as the plug-in method. Specifically, they suggest that $\beta_2$ is significant but $\beta_1$ is not.

From a biological standpoint, this means that the BPL trees prefer to live on slopes but do not really favor either high or low altitudes. Thus our results formally confirm the findings of Waagepetersen (2007) by using less restrictive assumptions.

**Appendix A: Proof of Theorem 1**

Let $\beta_0$ represent the true regression parameter. For ease of presentation but without loss of generality, we assume that $\hat{\beta}$ and $\beta_0$ are both scalars, that is, $p = 1$. In the case of $p > 1$, the proof can be easily generalized by applying the Cramer–Wold device.

**Lemma A.1.** Assume that conditions (4)–(6) hold. Then $|D_n|^2 \times E[(U_n^{(1)}(\beta_0))^2] < C$ for some $C < \infty$.

**Proof.** By repeatedly using Campbell’s theorem (e.g., Stoyan and Stoyan 1994) and the relationship between moments and cumulants, and using the fact that $\lambda^{(1)}(s; \beta_0)/\lambda(s; \beta_0)$ is bounded due to conditions (4) and (5), we can derive that $|D_n|^2 E[(U_n^{(1)}(\beta_0))^2]$ is bounded by the following (ignoring some multiplicative constants):

$$\int \int \int \int |Q_4(s_1, s_2, s_3, s_4)| \, ds_1 \, ds_2 \, ds_3 \, ds_4$$
$$+ \left[ \int \int |Q_2(s_1, s_2)| \, ds_1 \, ds_2 \right]^2$$
$$+ \int \int \int |Q_3(s_1, s_2, s_3)| \, ds_1 \, ds_2 \, ds_3$$
$$+ |D_n| \int |Q_2(s_1, s_2)| \, ds_1 \, ds_2$$
$$+ \int |Q_2(s_1, s_2)| \, ds_1 \, ds_2,$$

where all of the integrals are over $D_n$. It then follows from condition (6) that the foregoing sum is of order $|D_n|^2$. Thus the lemma is proven.

**Proof of Theorem 1**

First, note that

$$U_n^{(1)}(\beta) = \frac{1}{|D_n|} \left[ \sum_{x \in N \cap D_n} \frac{\lambda^{(1)}(x; \beta)}{\lambda(x; \beta)} - \int \lambda^{(1)}(s; \beta) \, ds \right]$$

and

$$U_n^{(2)}(\beta) = \frac{1}{|D_n|} \left[ \sum_{x \in N \cap D_n} \frac{\lambda^{(2)}(x; \beta)}{\lambda(x; \beta)} - \int \lambda^{(2)}(s; \beta) \, ds \right]$$

$$- \frac{1}{|D_n|} \sum_{x \in N \cap D_n} \left[ \frac{\lambda^{(1)}(x; \beta)}{\lambda(x; \beta)} \right]^2$$

$$:= U_n^{(2)}(\beta) - U_n^{(2)}(\beta_0).$$

Using the Taylor expansion, we can obtain

$$|D_n|^{1/2} (\hat{\beta}_n - \beta_0) = -\left[U_n^{(2)}(\beta_0)\right]^{-1} |D_n|^{1/2} U_n^{(1)}(\beta_0),$$

where $\beta_n$ is between $\hat{\beta}_n$ and $\beta_0$. We need to show that

$$U_n^{(2)}(\beta_0) \xrightarrow{p} 0$$

and

$$U_n^{(2)}(\beta_0) \xrightarrow{p} \frac{1}{|D_n|} \int \frac{[\lambda^{(1)}(s; \beta_0)]^2}{\lambda(s; \beta_0)} \, ds,$$
To show the first expression, note that \( E[U_n^{(2)}(\beta_0)] = \theta \). Thus we only need look at the variance term,

\[
\text{var}
\left[
U_n^{(2)}(\beta_0)
\right]
= \frac{1}{|D_n|} \iint \frac{\lambda^{(2)}(s_1; \beta_0)\lambda^{(2)}(s_2; \beta_0)}{\lambda(s_1; \beta_0)\lambda(s_2; \beta_0)} Q_2(s_1, s_2) \, ds_1 \, ds_2
+ \frac{1}{|D_n|} \int \frac{\lambda^{(2)}(s; \beta_0)^2}{\lambda(s; \beta_0)} \, ds,
\]

which converges to 0 due to conditions (4)–(6). Thus \( U_n^{(2)}(\beta_0) \xrightarrow{p} 0 \).

To show the second expression, note that \( E[U_n^{(2)}(\beta_0)] = \frac{1}{|D_n|} \int \frac{\lambda^{(1)}(s; \beta_0)^2}{\lambda(s; \beta_0)} \, ds \). Thus again we need consider only the variance term,

\[
\text{var}
\left[
U_n^{(2)}(\beta_0)
\right]
= \frac{1}{|D_n|} \iint \frac{\lambda^{(1)}(s_1; \beta_0)\lambda^{(1)}(s_2; \beta_0)^2}{\lambda(s_1; \beta_0)\lambda(s_2; \beta_0)} Q_2(s_1, s_2) \, ds_1 \, ds_2
+ \frac{1}{|D_n|} \int \frac{\lambda^{(1)}(s; \beta_0)^2}{\lambda(s; \beta_0)} \, ds,
\]

which converges to 0 due to conditions (4)–(6). Thus \( U_n^{(2)}(\beta_0) \xrightarrow{p} \frac{1}{|D_n|} \int \frac{\lambda^{(1)}(s; \beta_0)^2}{\lambda(s; \beta_0)} \, ds \).

Now we want to show that \( |D_n|^{-1/2} U_n^{(1)}(\beta_0) \) converges to a normal distribution. First, we derive the mean and variance of \( |D_n|^{-1/2} U_n^{(1)}(\beta_0) \). Clearly, \( E[|D_n|^{-1/2} U_n^{(1)}(\beta_0)] = 0 \). For the variance,

\[
\text{var}
\left[
|D_n|^{-1/2} U_n^{(1)}(\beta_0)
\right]
= \frac{1}{|D_n|} \iint \frac{\lambda^{(1)}(s_1; \beta_0)\lambda^{(1)}(s_2; \beta_0)^2}{\lambda(s_1; \beta_0)\lambda(s_2; \beta_0)} Q_2(s_1, s_2) \, ds_1 \, ds_2
+ \frac{1}{|D_n|} \int \frac{\lambda^{(1)}(s; \beta_0)^2}{\lambda(s; \beta_0)} \, ds.
\]

Now consider a new sequence of regions \( D_{n}^\ast \), where \( D_{n}^\ast \subset D_n \) and \( |D_{n}^\ast|/|D_n| \to 1 \) as \( n \to \infty \). Let \( U_n^{(1)}(\beta_0) \) be the estimating function for the realization of the point process on \( D_{n}^\ast \). Based on the expression of the variance, we can deduce that

\[
\text{var}
\left[
|D_n|^{-1/2} U_n^{(1)}(\beta_0) - |D_{n}^\ast|^{-1/2} U_n^{(1)}(\beta_0)
\right]
\to 0 \quad \text{as} \quad n \to \infty.
\]

Thus

\[
|D_n|^{-1/2} U_n^{(1)}(\beta_0) \sim |D_{n}^\ast|^{-1/2} U_n^{(1)}(\beta_0),
\]

where the notation \( a_n \sim b_n \) means that \( a_n \) and \( b_n \) have the same limiting distribution. Thus we need only show that \( |D_{n}^\ast|^{-1/2} U_n^{(1)}(\beta_0) \) converges to a normal distribution for some properly defined \( D_{n}^\ast \).

To obtain a sequence of \( D_{n}^\ast \), we apply the blocking technique used by Guan et al. (2004). Let \( l(n) = n^a \), \( m(n) = n^a - n^\epsilon \) for \( 4/(2 + \epsilon) < \eta < \alpha < 1 \), where \( \epsilon \) is defined as in condition (3). We first divide the original domain \( D_n \) into some nonoverlapping \( l(n) \times l(n) \) subareas, \( D_{l(n)}^i \) \( i = 1, \ldots, k_n \). Within each subarea, we further obtain \( D_{m(n)}^i \), an \( m(n) \times m(n) \) square sharing the same center with \( D_{l(n)}^i \). Note that

\[
d(D_{m(n)}^i, D_{m(n)}^j) \geq n^\eta \frac{\epsilon}{i \neq j}.
\]

Condition (2) implies that \( |D_{m(n)}^i|/|D_n| \to 1 \). Thus we need only show that \( |D_{m(n)}^i|^{-1/2} U_n^{(1)}(\beta_0) \) converges to a normal distribution. This is true due to the mixing condition, the result in Lemma A.1, and the application of the Lyapunov’s central limit theorem. The proof is similar to that of Guan et al. (2004); thus we omit the details here.

**APPENDIX B: PROOF OF THEOREM 2**

To simplify notation, let \( g(\mathbf{u}) = g(\mathbf{u}) - 1 \), \( \phi_k(s_1, \ldots, s_k) = \phi_k(s_1, \ldots, s_l)/[(\lambda(s_1) \cdots \lambda(s_l))] \), \( \gamma_k(s_1, \ldots, s_k) = \gamma_k(s_1, \ldots, s_l)/[(\lambda(s_1) \cdots \lambda(s_l))] \) for \( k = 3 \) and \( 4 \) and \( Z(s) = \lambda^{(1)}(s) \). As in the proof of Theorem 1, we assume that \( \hat{\beta} \) and \( \hat{\rho} \) are both scalars, that is, \( p = 1 \). As a result, \( Z(s) \) is a scalar, so that \( \text{cov}(S_n) = var(S_n) \). Furthermore, we consider only the case where \( K = 1 \). Thus the covariance estimator defined in (11) becomes

\[
\hat{\text{var}}(S_n) = \sum_{b=1}^{B} \frac{(s_b - \bar{s}_n)^2}{B - 1}.
\]

As \( B \) goes to infinity, we see that the foregoing converges to

\[
\text{var}(S_n^b | \Psi \cap D_n) \to \frac{1}{|D_n|} \int_{D_n} \int_{D_n} Z(s_1)Z(s_2)\psi(s_1 - s_2) \, ds_1 \, ds_2
\]

where the notation \( \sum_{D} \) is short for \( \sum_{\Psi \cap D} \). We wish to show that \( \text{var}(S_n^b | \Psi \cap D_n) \) converges in \( L_2 \) to

\[
\frac{\lambda_n}{|D_n|} \int_{D_n} [Z(s)]^2 \, ds.
\]

To show this, we need to show that

\[
\frac{1}{|D_n|} E[\text{var}(S_n^b | \Psi \cap D_n)] \to \frac{1}{|D_n|} \text{var}(S_n) \quad \text{(B.1)}
\]

and

\[
\frac{1}{|D_n|} \text{var}[\text{var}(S_n^b | \Psi \cap D_n)] \to 0 \quad \text{(B.2)}
\]

To show (B.1), note that

\[
E[\text{var}(S_n^b | \Psi \cap D_n)]
= \frac{(\lambda_n)^2}{|D_n|} \sum_{i=1}^{k_n} \int_{D_{l(n)}^i} Z(s_1)Z(s_2) g(s_1 - s_2) \, ds_1 \, ds_2
+ \frac{\lambda_n}{|D_n|} \int_{D_n} [Z(s)]^2 \, ds
- \frac{\lambda_n^2}{|D_n|} \sum_{i=1}^{k_n} \sum_{j=1}^{k_n} \int_{D_{l(n)}^i} \int_{D_{l(n)}^j} Z(s_1)Z(s_2) g(s_1 - s_2) \, ds_1 \, ds_2
\]

Thus

\[
\frac{1}{|D_n|} \left[ E[\text{var}(S_n^b | \Psi \cap D_n)] - \text{var}(S_n) \right]
= \frac{\lambda_n}{|D_n|} \int_{D_n} [Z(s)]^2 \, ds
- \frac{(\lambda_n)^2}{|D_n|} \sum_{i \neq j} \int_{D_{l(n)}^i} \int_{D_{l(n)}^j} Z(s_1)Z(s_2) g(s_1 - s_2) \, ds_1 \, ds_2.
\]
elementary derivations yield

\[ T_n = \frac{n}{k_n} \sum_{i=1}^{k_n} \sum_{j=1}^{k_n} \sum_{l=1}^{k_n} \int D_{D_n} |Z(s_1) - Z(s_2)| \, ds \]

- \varphi(s_1 - s_2 + c_{j1} + c_{i1})

\[ := -T_n^1 - T_n^2 - T_n^3. \]

\[ T_n^1 \] goes to 0 due to conditions (2) and (5). For \[ T_n^2 \] and \[ T_n^3 \], lengthy yet elementary derivations yield

\[ |T_n^2| \leq \frac{C(n_2)}{|k_n|} \sum_{i=1}^{k_n} \int D_{D_n} |Z(s_1 - c_j_1 + c_{i1})| \, ds \]

\[ \times \int D_{D_n} (D_n - s) \, |\varphi(s)| \, ds \]

\[ - \frac{1}{|D_{D_n}|} \int D_{D_n} \int D_{D_n} |D_{D_n}(s_1 - c_{j1} + c_{i1})| \, ds \]

and

\[ T_n^3 \leq \frac{C(n_3)}{|k_n|} \int D_{D_n} |\varphi(s_1 - s_2)| \, ds. \]

Both terms converge to 0 due to conditions (2), (4), (5), and (6). Thus (B.1) is proven. Furthermore, we can see that both \[ T_n^1 \] and \[ T_n^2 \] are of order \( 1/k_n \), whereas \[ T_n^3 \] is of order \( 1/D_{D_n}/2 \) due to condition (12).

To prove (B.2), we first note that \[ \var{\var{S_n^2}} \] is equal to the sum of the following three terms:

\[ \frac{1}{k_n^2} \sum_{i=1}^{k_n} \sum_{j=1}^{k_n} \sum_{l=1}^{k_n} \left[ \int D_{D_n} \int D_{D_n} |Z(x_1 - c_{j1} + c_{i1})| \, ds \right] \]

\[ \times \int D_{D_n} |Z(x_2 - c_{j2} + c_{i2})| \int D_{D_n} |Z(x_3 - c_{j3} + c_{i3})| \int D_{D_n} |Z(x_4 - c_{j4} + c_{i4})| \]

\[ \left[ \int D_{D_n} \int D_{D_n} |Z(x_1 - c_{j1})| \int D_{D_n} |Z(x_2 - c_{j2})| \int D_{D_n} |Z(x_3 - c_{j3})| \int D_{D_n} |Z(x_4 - c_{j4})| \right]. \]

The first integral in the foregoing is of order \( 1/k_n \) due to conditions (4), (5), and (6). The second integral is bounded by

\[ \frac{2C(n_4)}{|k_n|} \sum_{i=1}^{k_n} \int D_{D_n} \int D_{D_n} \int D_{D_n} \int D_{D_n} \int D_{D_n} |F_1(x_1, s_2) | \]

\( s_3, s_4 \) \) \, ds_1 \, ds_2 \, ds_3 \, ds_4.

The foregoing is of order \( 1/k_n + \alpha_n \) due to conditions (4), (5), (6), and (12), where the order of \( \alpha_n \) is no higher than \( 1/D_{D_n} \).

**APPENDIX C: JUSTIFICATION FOR USING \( \hat{\beta}_n \) IN THE THINNING STEP**

Here we assume \( \var{D_{D_n}} = o(D_{D_n}/2) \). Let \( p_n(x; \theta) = \min \{ D, \lambda(s; \theta) / \lambda(x; \theta) \} \) and let \( r(x) \) be a uniform random variable in \([0, 1] \). If \( r(x) \leq p_n(x; \theta) \), where \( x \in (N \cap D_n) \), then \( x \) will be retained in the thinned process, say \( \Psi(\theta) \). Based on the same set \( \{ r(x) : x \in (N \cap D_n) \} \).
\(D_n(\theta)\)}, we can determine \(\Psi(\theta_0)\) and \(\Psi(\hat{\theta}_n)\). Note that \(\Psi(\theta_0)\) is \(\Psi\) as defined in (7) using the true FOIF. Define \(\Psi_a = \{x \in \Theta(\theta_0), x \neq \Psi(\theta_0)\}\), and \(\Psi_b = \{x \neq \Psi(\theta_0), x \in \Psi(\hat{\theta}_n)\}\). Note that \(P(x \in \Psi_a \cup \Psi_b | x \in N) \leq p_n(x; \theta_0) + p_n(x; \hat{\theta}_n)\). We need to show that
\[
\frac{1}{|D_n|} \left[ \text{var}\left[ S_n^b(x | \Theta(\theta_0)) \cap D_n \right] - \text{var}\left[ S_n^b(x | \Theta(\hat{\theta}_n) \cap D_n \right] \right] \overset{P}{\rightarrow} 0,
\]
where \(\text{var}\left[ S_n^b(x | \Theta(\theta_0)) \cap D_n \right] \) and \(\text{var}\left[ S_n^b(x | \Theta(\hat{\theta}_n) \cap D_n \right] \) are defined in Appendix B. Note that \(\text{var}\left[ S_n^b(x | \Theta(\theta_0)) \cap D_n \right] \) is defined analogously.

Let \(\sum D\) and \(\sum D_n\) denote
\[\sum_{D \cap (\Psi_a \cup \Psi_b)} \quad \text{and} \quad \sum_{\sum_{D_n} \neq 0} \sum_{x_1 \in D \cap N, x_2 \in D \cap (\Psi_a \cup \Psi_b), x_1 \neq x_2} \).

Note that
\[
\frac{1}{|D_n|} \left[ \text{var}\left[ S_n^b(x | \Theta(\theta_0)) \cap D_n \right] - \text{var}\left[ S_n^b(x | \Theta(\hat{\theta}_n) \cap D_n \right] \right] \leq C \frac{1}{|D_n|^2} \sum_{i=1}^{k_n} \sum_{j=1}^{k_n} \sum_{D_{i,j} \cap N} \left( D_{i,j} \cap D_{i,j} \right) \sum_{j=1}^{k_n} \sum_{j=2}^{k_n} \sum_{D_{i,j} \cap D_{i,j}^2} \left( D_{i,j} \cap D_{i,j} \right) \sum_{D_n} \neq 0 \sum_{D_n}^{1} \sum_{D_n}^{1} \left( x_1 \in D_n \cap N \right) \left. \right| \sum_{D_n}^{1} \left( x_1 \in D_n \cap N \right)
\]

Note that \(P(x \in (\Psi_a \cup \Psi_b) | x \in N) \leq \frac{C}{|D_n|^{a/2}}\) with probability 1 for \(0 < a < 1\) due to Theorem 1. Thus all three terms in the foregoing converge to 0 in probability. This in turn leads to the consistency of the variance estimator.

[Received December 2006. Revised May 2007.]

REFERENCES


Politis, D. N. (1999), Subsampling, New York: Springer-Verlag.


