Statistica Sinica **14**(2004), 69-101

BOOTSTRAPPING A SPATIAL POINT PROCESS

J. M. Loh and M. L. Stein

Columbia University and University of Chicago

Abstract: We consider the problem of resampling or bootstrapping a point process to get confidence intervals for the reduced second moment function. We propose a resampling scheme for spatial data, which we call the marked point method. This is a variant of the block of blocks bootstrap first introduced by Künsch (1989). A simulation study with a Poisson, a clustered and a regular point process on the unit square in $\mathbb{R}^2$ shows that the marked point method yields confidence intervals that are closer to the nominal (95%) level than resampling by tiling (block bootstrap) and by using subsets (subsampling). The confidence intervals obtained by the marked point method also tend to be shorter, after accounting for differences in empirical coverage. Finally, the marked point method is very much computationally less intensive so that, even with moderate sample sizes, the marked point method takes considerably less computing time. We also find that the simple method of dividing the sample and treating the subsamples as independent replicates works reasonably well. We apply some of these methods to a set of astronomy data.

Key words and phrases: Marked point bootstrap, reduced second moment function, resampling.

1. Introduction

The reduced second moment function $K(r)$ is a commonly used measure of clustering for point processes. It is defined as the expected number of points within a distance r of a typical point of a point process, divided by the intensity of the process. While various methods to estimate $K(r)$ are available, getting standard errors for the estimates is more difficult.

Resampling of point processes has been studied by a number of researchers (e.g., Hall (1985), Künsch (1989), Liu and Singh (1992), Politis and Romano (1992a) and Lahiri (1992, 1993)), but little has been said about resampling for estimating second-order structure of point processes. When resampling to estimate the second-order structure, the resampling scheme used is particularly important since a poor resampling scheme may produce pairs of points in the new resamples that do not reflect properties of the original process.

In this work, we focus on finding confidence intervals for the isotropic estimator $\hat{K}(r)$ of $K(r)$ (Ripley (1988)). We use a simulation study on the unit square in $\mathbb{R}^2$ with an unclustered, a clustered and a regular process to compare

different methods of finding confidence intervals. We consider an approximation method and three different resampling schemes. The approximation method is based on dividing the observation region into N subregions and assuming independence and normality of the statistic of interest in the different subregions. The variance of $\hat{K}(r)$ is then estimated from the sample variance of the N separate estimates of $K(r)$. The resampling schemes used are the tiling method (or block bootstrap), the subsets method (similar to subsampling) and a new method which is a variant of the block of blocks bootstrap of Künsch (1989), Politis and Romano (1992b) and Bühlmann and Künsch (1995). We call this method the marked point method because of the way it is formulated, using marks assigned to observed points.

We find that the marked point method tends to produce nominal 95% confidence intervals whose empirical coverages are closer to the 95% level. Furthermore, the lengths of the intervals obtained by the marked point method are often shorter than the intervals obtained by the other methods. Finally, resampling of spatial point processes can be computationally demanding. The marked point method is very much computationally less intensive than the other two resampling schemes considered, allowing resampling to be more readily used for point processes.

We describe the reduced second moment function in Section 2. The methods considered in this work for obtaining standard errors are described in Section 3. In particular we introduce the marked point method as a better alternative to the tiling and subsets resampling methods. The results of a simulation study on the unit square in $\mathbb{R}^2$ are given in Section 4. In Section 5 we discuss further issues regarding resampling a point process, including the use of resampling subregions of a different shape from the observation window and the advantage of toroidal wrapping. This work was motivated by our investigations into estimating $K(r)$ for a particular set of astronomy data, known as an absorber catalog. The results of simulations of absorber catalogs as well as confidence intervals for $K(r)$ for the actual data are given in Section 6.

2. The Reduced Second Moment Function

Second-order characteristics of a stationary spatial point process describe the dependence of pairs of points of the process. A commonly used measure of the second-order characteristics of a stationary point process is the reduced second moment function

$$K(r) = \lambda^{-1}\mathrm{E}[N(\mathbf{x}, r)|\text{point at } \mathbf{x}],$$

where λ is the intensity of the process and $N(\mathbf{x}, r)$ is the number of points within distance r of $\mathbf{x}$ but with the point at $\mathbf{x}$ excluded. Thus $\lambda K(r)$ can be thought of as

the expected number of points within distance r of a typical point of the process. Together, $K(r)$ and λ completely determine the first and second moments of a stationary, isotropic point process. See Stoyan, Kendall and Mecke (1995) for more information on the reduced second moment function, and the subject of second order analysis of stationary point processes in general.

For a homogeneous Poisson process in $\mathbb{R}^d$, $K(r) = \mu_d r^d$, where μ_d is the volume of a unit sphere in d dimensions. Values of $K(r)$ for a process are often compared with those for the Poisson process. Values larger or smaller than $\mu_d r^d$ respectively indicate a more clustered or more regular process than the Poisson process.

When estimating $K(r)$ for a point process observed in a bounded window $A \subset \mathbb{R}^d$, the common practice is to first estimate $\lambda^2 v_d(A)K(r)$, where $v_d(.)$ indicates Lebesgue measure in d dimensions, and to divide by an estimator of $\lambda^2 v_d(A)$. To estimate $\lambda^2 v_d(A)K(r)$, we need to count the number of neighbors within distance r of each observed point. A problem when doing this is that for each point within r of the boundary of A, the exact number of points within r of it is not known. If A had no edges, $\sum_{\mathbf{x}} \sum_{\mathbf{y}:\mathbf{y}\neq\mathbf{x}} 1\{|\mathbf{y}-\mathbf{x}| \leq r\}$, where $1\{.\}$ is the indicator function, or more compactly, $\sum_{\mathbf{x}\neq\mathbf{y}} 1\{|\mathbf{y}-\mathbf{x}| \leq r\}$, would be an unbiased estimator of $\lambda^2 v_d(A)K(r)$. When A does have edges, $\sum_{\mathbf{x}\neq\mathbf{y}} 1\{|\mathbf{y}-\mathbf{x}| \leq r\}$ has expected value less than $\lambda^2 v_d(A)K(r)$. The difference can be substantial if much of A is within r of its boundary. The resulting estimator of $K(r)$, which Ripley (1988) calls the naive estimator, has a negative bias.

A number of edge-corrected estimators of $K(r)$ have been developed to deal with this problem (see, for example, Stoyan, Kendall and Mecke (1995), Stein (1993) and Ripley (1988)). The key idea is to get unbiased estimates of $\lambda^2 v_d(A)K(r)$ by introducing weights $w_A(\mathbf{x},\mathbf{y})$ for each observed pair $(\mathbf{x},\mathbf{y})$. These weights compensate for the points that we cannot observe due to the boundary of A. Specifically, with proper weights, we have

$$\mathrm{E}\left[\sum_{\mathbf{x}\neq\mathbf{y}} 1\{|\mathbf{y}-\mathbf{x}| \leq r\} w_A(\mathbf{x},\mathbf{y})\right] = \lambda^2 v_d(A)K(r). \tag{1}$$

We obtain an estimator of $K(r)$ by dividing $\sum_{\mathbf{x}\neq\mathbf{y}} 1\{|\mathbf{y}-\mathbf{x}| \leq r\} w_A(\mathbf{x},\mathbf{y})$ by an estimator of $\lambda^2 v_d(A)$. Usual estimators of $\lambda^2 v_d(A)$ are $n^2/v_d(A)$ and $n(n-1)/v_d(A)$. We use $n(n-1)/v_d(A)$ in this work. The estimator $v_d(A)\sum_{\mathbf{x}\neq\mathbf{y}} 1\{|\mathbf{y}-\mathbf{x}| \leq r\} w_A(\mathbf{x},\mathbf{y})/[n(n-1)]$ is not unbiased for $K(r)$. However, if the point process is ergodic, the estimator is consistent for $K(r)$ as the observation window A increases such that the diameter $d(A)$ of its largest inscribed circle tends to

infinity (see Nguyen and Zessin (1979), specifically example 3 in Corollary (4.20), for details). In this work, we use the isotropic correction estimate of $K(r)$.

2.1. The isotropic estimator

The isotropic correction (Ripley (1988)) can be used to correct for edge effects when estimating $K(r)$ of a stationary and isotropic process observed in a bounded window A. Consider such a process observed in $A \subset \mathbb{R}^d$. Suppose we observe a point at $\mathbf{x} \in A$ and another point at $\mathbf{y} \in A$ within distance r of $\mathbf{x}$. For the isotropic correction, the weight $w_A(\mathbf{x}, \mathbf{y})$ is given a value equal to the reciprocal of the fraction of the area of the shell of radius $|\mathbf{y} - \mathbf{x}|$ centered at $\mathbf{x}$ that is contained in A. Specifically,

$$w_A(\mathbf{x}, \mathbf{y}) = \frac{v_{d-1}(\partial \mathcal{B}_0(\mathbf{x}, |\mathbf{y} - \mathbf{x}|))}{v_{d-1}(\partial \mathcal{B}_0(\mathbf{x}, |\mathbf{y} - \mathbf{x}|) \cap A)} = \frac{\eta_d |\mathbf{y} - \mathbf{x}|^{d-1}}{v_{d-1}(\partial \mathcal{B}_0(\mathbf{x}, |\mathbf{y} - \mathbf{x}|) \cap A)},$$

where η_d is the surface area of a unit sphere in d dimensions and $\partial \mathcal{B}_0(\mathbf{x}, |\mathbf{y} - \mathbf{x}|)$ is the shell with center $\mathbf{x}$ and radius $|\mathbf{y} - \mathbf{x}|$. These weights $w_A(\mathbf{x}, \mathbf{y})$ can be large if only a small portion of $\partial \mathcal{B}_0(\mathbf{x}, |\mathbf{y} - \mathbf{x}|)$ is in A. Ripley (1988) shows that, with these weights, (1) holds for r up to the circumradius of A. With n points observed in A, the isotropic estimator of $K(r)$ is then given by

$$\hat{K}_I(r) = \frac{v_d(A)}{n(n-1)} \sum_{\mathbf{x} \neq \mathbf{y}} 1\{|\mathbf{y} - \mathbf{x}| \leq r\} \frac{\eta_d |\mathbf{y} - \mathbf{x}|^{d-1}}{v_{d-1}(\partial \mathcal{B}_0(\mathbf{x}, |\mathbf{y} - \mathbf{x}|) \cap A)} \cdot \frac{v_d(A)}{v_d(A_{|\mathbf{y}-\mathbf{x}|})},$$

where $A_{|\mathbf{y}-\mathbf{x}|}$ is the set of all $\mathbf{z} \in A$ that are a distance $|\mathbf{y} - \mathbf{x}|$ from at least one other $\mathbf{z}_1 \in A$. We have used Ohser's extension (Ohser (1983)) in (2) by including the factor $v_d(A)/v_d(A_{|\mathbf{y}-\mathbf{x}|})$. Including this factor allows the estimate to be valid for distances up to the diameter of A.

3. Confidence Intervals for $K(r)$

Here, we describe a number of methods to obtain confidence intervals for $K(r)$ of an observed stationary, isotropic point process. We do this in the context of a square observation window in $\mathbb{R}^2$. Extension to higher dimensions is conceptually not difficult, although the problem of edge effects is more severe and the computation involved more demanding. We use $\hat{K}(r)$ and $\tilde{K}(r)$ to respectively denote the actual and the bootstrapped estimates of $K(r)$.

Suppose we observe a point process in a square observation window $A \subset \mathbb{R}^2$ with area a. An estimate of $K(r)$ is obtained by dividing an estimate of $\lambda^2 a K(r)$ by an estimate of $\lambda^2 a$. Our estimator of $\lambda^2 a K(r)$ is of the form

$$\sum_{\mathbf{x}} \sum_{\mathbf{y}: \mathbf{y} \neq \mathbf{x}} 1\{|\mathbf{y} - \mathbf{x}| \leq r\} w_A(\mathbf{x}, \mathbf{y}), \tag{2}$$

where $w_A(\mathbf{x}, \mathbf{y})$ is a weight of some form to account for the edge effects.

A simple way to obtain confidence intervals is to divide the region A into N congruent subregions and compute N separate estimates $\hat{K}_1(r), \ldots, \hat{K}_N(r)$. An approximate $100(1-\alpha)\%$ confidence interval is then given by

$$\hat{K}(r) \pm t_{N-1,\alpha/2} \sqrt{\frac{\widehat{\mathrm{Var}}\{\hat{K}_i(r)\}}{N}}, \tag{3}$$

where $\hat{K}(r)$ is the overall estimate of $K(r)$, $\widehat{\mathrm{Var}}\{\hat{K}_i(r)\}$ is the sample variance of $\hat{K}_1(r), \ldots, \hat{K}_N(r)$ and $t_{N-1,\alpha/2}$ is the $(1-\alpha/2)$th percentile of the t_{N-1} distribution. Here we assume that the estimates from the N subregions are approximately independent and Gaussian, and that the variance of $\hat{K}(r)$ is $1/N$ times that of $\hat{K}_i(r)$. We call this the splitting method.

A simple way to resample a point process is by tiling (e.g., Hall (1985)), which is an extension of block-sampling methods for bootstrapping time series data. Künsch (1989) and Liu and Singh (1992) also discuss the tiling method for resampling observations in one dimension. For square A in two dimensions, this usually involves placing N square subregions of area a/N randomly in A and copying the pattern within each subregion. These subregions, together with the copied patterns, are arranged in some predetermined systematic way to reproduce A (see Figure 1). The points resulting from this arrangement of the copied patterns form a sample of the point process, $\hat{\mathbf{x}}_i$, $i = 1, \ldots, \hat{n}$. With this sample, a new estimate

$$\tilde{K}(r) = \frac{a}{\hat{n}(\hat{n}-1)} \sum_{i=1}^{\hat{n}} \sum_{\substack{j=1 \\ j \neq i}}^{\hat{n}} 1\{|\hat{\mathbf{x}}_j - \hat{\mathbf{x}}_i| \leq r\} w_A(\hat{\mathbf{x}}_i, \hat{\mathbf{x}}_j) \tag{4}$$

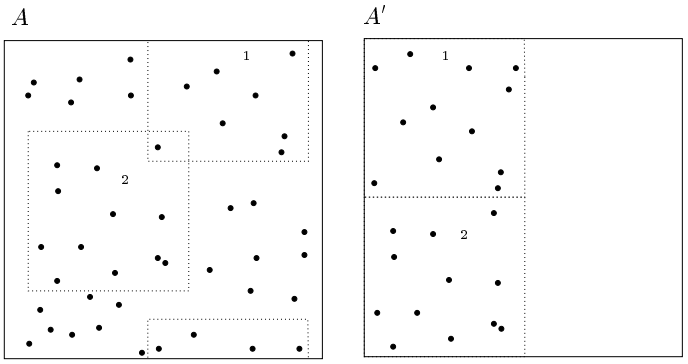

Figure 1. This diagram shows the tiling method in progress. The observed point process is shown on the left. Two tiles have been placed in A, the pattern copied and arranged in A' on the right. The tile that ended up on the upper left corner of A' (tile 1) was wrapped round A.

can be obtained. This is repeated a large number R of times. Assuming that the sampling distribution of $\tilde{K}(r) - \hat{K}(r)$ is similar to that of $\hat{K}(r) - K(r)$, the $(1-\alpha/2)$th and $(\alpha/2)$th quantiles of $\hat{K}(r) - K(r)$ can be estimated by the $(R+1)(1-\alpha/2)$th and $(R+1)(\alpha/2)$th ordered values of $\tilde{K}(r) - \hat{K}(r)$. Then a $100(1-\alpha)\%$ confidence interval for $K(r)$, called the basic bootstrap interval by Davison and Hinkley (1997), is given by

$$\left[2\hat{K}(r) - \tilde{K}_{(R+1)(1-\alpha/2)}(r), \quad 2\hat{K}(r) - \tilde{K}_{(R+1)\alpha/2}(r)\right]. \tag{5}$$

This method of resampling will be referred to as the tiling method. There are various modifications to this resampling scheme. One modification is to partition the observation window A into subregions and obtain samples by independently sampling from these subregions. Another modification is to place the subregions randomly but allow them to fall partly outside A. We treat the window as being wrapped on a torus so that the parts of the subregions that fall outside A copy the pattern on the other side of A (see Figure 1). Politis and Romano (1992a) suggest using toroidal wrapping before resampling. An advantage of wrapping is that all points are chosen with equal probability. Finally, for an isotropic process, the resampling subregions can also be rotated, rather than always be placed in the same orientation. In our simulations we use randomly placed tiles with toroidal wrapping and do not rotate the tiles.

The main problem with the tiling method is that points that are not near each other in the actual sample are put close together in adjacent resampled tiles. Points near the edges of adjacent tiles produce artificial point pairs that may distort the second-order characteristics of the point process and create bias in the estimates. The clearest example occurs in the case of a hard core process, in which no two points can be within a distance r_0 of each other. This necessarily means that $K(r) \equiv 0$ for all $r < r_0$, but this will generally not be true for $\tilde{K}(r)$ estimated from samples obtained by tiling. Indeed, Lahiri (1993) shows that with longer range dependence, the tiling method begins to fail as putting independent blocks together destroys the long-range dependence present in the original observations.

A possible way to get round this problem is not to tile at all. We randomly obtain copies of the point pattern using square subregions of A in the same way as the tiling method. However, instead of tiling the subregions together to recreate A, we treat each subregion separately from the others. Only point pairs that occur in the same subregion contribute to the bootstrap estimate $\tilde{K}(r)$. Suppose we use N square subregions A_i, $i = 1, \ldots, N$, of area a/N to randomly copy the point pattern. We consider the new sample as an 'observation' of the point process over N widely separate regions. Let $\mathbf{x}_{ij}$ and $\mathbf{x}_{ik}$ represent two different points in A_i, with $j, k = 1, \ldots, n_i$. Then the two points $\mathbf{x}_{ij}$ and $\mathbf{x}_{ik}$ together contribute $1\{|\mathbf{x}_{ij} - \mathbf{x}_{ik}| \le r\}[w_{A_i}(\mathbf{x}_{ij}, \mathbf{x}_{ik}) + w_{A_i}(\mathbf{x}_{ik}, \mathbf{x}_{ij})]$ to the estimate of

$\lambda^2 aK(r)$, where $w_{A_i}(\mathbf{x}_{ij}, \mathbf{x}_{ik})$ and $w_{A_i}(\mathbf{x}_{ik}, \mathbf{x}_{ij})$ are weights based on the smaller regions A_i. For example, if the isotropic correction is used, $w_{A_i}(\mathbf{x}_{ij}, \mathbf{x}_{ik}) = \eta_d |\mathbf{x}_{ik} - \mathbf{x}_{ij}|^{d-1} / v_{d-1}(\partial \mathcal{B}_0(\mathbf{x}_{ij}, |\mathbf{x}_{ik} - \mathbf{x}_{ij}|) \cap A_i)$. Thus the estimate of $K(r)$ is given by

$$\tilde{K}(r) = \frac{a}{\sum n_i (\sum n_i - 1)} \sum_{i=1}^{N} \sum_{j \neq k}^{n_i} 1\{|\mathbf{x}_{ij} - \mathbf{x}_{ik}| \leq r\} w_{A_i}(\mathbf{x}_{ij}, \mathbf{x}_{ik}), \tag{6}$$

where we use $\sum n_i (\sum n_i - 1)/a$ as our estimate for $\lambda^2 a$. We refer to this method as the subsets method. The subsets method is similar to the subsampling method introduced in Politis and Romano (1994).

Notice that (6) involves a sum over fewer pairs of points than in the tiling method at (4), because points in different subregions are not considered. The fewer number of pairs in (6) is compensated for by the weights, which are larger being based on smaller subregions. Limiting pairs of points to those occurring in the same subregion avoids the problem of artificially produced pairs, but results in more pronounced edge effects. This may outweigh the advantage gained by not tiling. Some artificially produced pairs will be present if toroidal wrapping is used, but they will be considerably fewer than the number that would have been obtained by tiling. To avoid such pairs altogether, we can treat the different parts of a wrapped subregion separately. In our simulations we use randomly placed subregions with toroidal wrapping and do not split up a wrapped subregion.

Künsch (1989) introduced a procedure to reduce the effect of putting independent blocks together when resampling equally spaced observations in one dimension. By considering statistics T_N that depend on an m-dimensional marginal with m fixed, Künsch (1989) first defined blocks of m consecutive observations. Specifically, with observations $X_t, t = 1, \ldots, N$, define $Y_{t,m} = (X_t, \ldots, X_{t+m-1})$, $t = 1, \ldots, N-m+1$. Tiling is applied to Y_t rather than to X_t directly. For example, the sample autocovariance at lag s is given by $\hat{T}_N = \sum_{i=1}^{N-s} X_i X_{i+s}/(N-s)$. Take $m = s+1$ and set $\phi(Y_{t,m}) = \phi(X_t, \ldots, X_{t+m-1}) = X_t X_{t+m-1}$. Then with bootstrapped observations $\tilde{Y}_{t,m}$, the bootstrap estimate of T_N is given by $\tilde{T}_N = \sum_{i=1}^{N-m+1} \phi(\tilde{Y}_{t,m})/(N-m+1)$. Politis and Romano (1992b) and Bühlmann and Künsch (1995) further developed this block of blocks bootstrap method when resampling time series, by allowing m to increase slowly.

Braun and Kulperger (1998) suggest a method to estimate the second-order intensity $\rho(\tau)$ of a point process in $\mathbb{R}$ by bootstrapping the process observed in an interval $(0, T]$. Using what they call the marked point process bootstrap, they give each observed point $x \in (0, T]$ a mark equal to the number of points in the interval $(x + \tau, x + \tau + h]$ for fixed small h and resample using sub-intervals without wrapping. We introduce a number of adaptations of their approach to make it more suitable for the spatial setting and propose a resampling scheme to

obtain confidence intervals for $K(r)$. To each point $\mathbf{x}$, we give a mark $m_{\mathbf{x}}$ that is the sum of all the weights $w_A(\mathbf{x}, \mathbf{y})$ for points $\mathbf{y}$ within r of $\mathbf{x}$. Thus $m_{\mathbf{x}}$ is the total contribution to the estimate of $\lambda^2 aK(r)$ by all the points within r of $\mathbf{x}$. Specifically,

$$m_{\mathbf{x}} = \sum_{\mathbf{y}:\mathbf{y}\neq\mathbf{x}} 1\{|\mathbf{y}-\mathbf{x}| \le r\} w_A(\mathbf{x}, \mathbf{y}). \tag{7}$$

Note that this expression for $m_{\mathbf{x}}$ is equal to the inner sum at (2). The sum of all the marks of the observed points is given by (2), the estimate of $\lambda^2 aK(r)$.

The points, together with the marks, are resampled using some resampling scheme (randomly placed or fixed subregions, with or without wrapping). For the example shown in Figure 2, point $\mathbf{y}$ is resampled but not point $\mathbf{z}$. The mark given to $\mathbf{x}$ is $m_{\mathbf{x}} = w_A(\mathbf{x}, \mathbf{y}) + w_A(\mathbf{x}, \mathbf{z})$. Thus the mark given to $\mathbf{x}$ contains information about the presence of $\mathbf{z}$. Furthermore, although point $\mathbf{v}$ is within r of $\mathbf{x}$ in the resampled process, this is not recorded in the mark given to $\mathbf{x}$ and vice versa. This avoids the problem of artificially produced pairs. The estimate of $\lambda^2 aK(r)$ for the new resample is obtained by adding up the marks of the resampled points. Specifically, suppose we use subregions $A_i, i = 1, \ldots, N$, to resample the point process, resulting in A_i containing points $\mathbf{x}_{ij}$, $j = 1, \ldots, n_i$. Each $\mathbf{x}_{ij}$ has a mark of the form of (7): $m_{ij} = \sum_{\mathbf{y}:\mathbf{y}\neq\mathbf{x}_{ij}} 1\{\mathbf{y} \in A : |\mathbf{y} - \mathbf{x}_{ij}| \le r\} w_A(\mathbf{x}_{ij}, \mathbf{y})$. Then, if we use $\sum n_i(\sum n_i - 1)/a$ to estimate $\lambda^2 a$, we have

$$\tilde{K}(r) = \frac{a}{\sum n_i(\sum n_i - 1)} \sum_{i=1}^{N} \sum_{j=1}^{n_i} m_{ij}.$$

We refer to this method as the marked point method. The resampling scheme we use with this method is randomly placed subregions with toroidal wrapping.

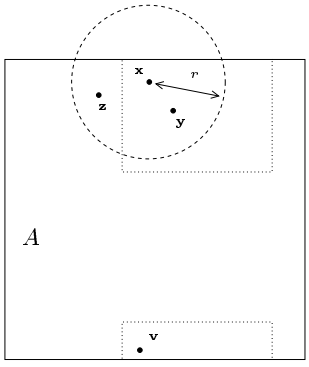

Figure 2. In this diagram only points $\mathbf{y}$ and $\mathbf{z}$ are within distance r of point $\mathbf{x}$ (the dashed circle). A resampling subregion has been wrapped round A. Points that are resampled by this subregion include $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{v}$, but not $\mathbf{z}$. With the marked point method, $\mathbf{x}$ is allocated a mark given by (7).

Note that with the marked point method, the mark given to each point $\mathbf{x}$ consists of the sum of contributions to the estimate of $\lambda^2 aK(r)$ by points near $\mathbf{x}$ in the observed process (see (7)). These marks are resampled together with the points. To compute the bootstrapped estimate, the marks of the resampled points are added together. No new marks are calculated, so that the actual positions of points in the resampled process are not used at all. Thus using toroidal wrapping with the marked point method is merely a simple method to resample points with equal probability and does not affect the bootstrapped estimate.

The marked point method we introduce here is a variant of the block of blocks bootstrap of Künsch (1989), generalizing it to allow resampling of general point processes in $\mathbb{R}^d$ where the observed points are not equally spaced. More importantly, the contribution of a point at $\mathbf{y}$ to $m_{\mathbf{x}}$ includes information about the boundary of the observation window, through the weights used to account for edge effects. When the edge effects are large, as in the case of the absorber catalog (see Section 6), including the weights in the resampling scheme is crucial.

Intuitively, the marked point process method appears to be better than the tiling or subsets methods. There are no artificially produced point pairs. Marks given to the points are contributions to $K(r)$ by other points that were actually found nearby. The main weakness of the subsets method is the large influence of edge effects. With the subsets method, a very much larger number of points are near an edge, resulting in larger weights being used. The larger weights compensate for the fewer number of point pairs but may cause the estimate to have more uncertainty. In contrast, for the marked point process bootstrap, the weights used are those based on the whole region A. For both the tiling and subsets methods, we consider only those points resampled by subregions. Information about points outside the subregions is lost. With the marked point method, however, some information about points outside the subregions is retained in the resamples.

Finally, the marked point method has a computational advantage over the tiling and subsets methods. The splitting method requires the least computation. With N subregions, only $N+1$ estimates of $K(r)$ are needed, the actual estimate and an estimate for each subregion. Among the bootstrap methods, however, the marked point method is least computationally intensive while the tiling method is by far the most computationally intensive. For the marked point bootstrap, the marks given to the observed points are calculated once, when the estimate of $K(r)$ is computed. The resampling process merely involves keeping track of the resampled points and summing up the relevant marks. For the tiling method, each new resample requires a new computation of all the weights. To get a confidence interval using R resamples, $R+1$ estimates of $K(r)$ are needed. The subsets method also requires new computation of weights for each resample, but

is slightly less computationally intensive than the tiling method because there are fewer point pairs and therefore fewer weights to compute.

4. Simulation Studies on the Unit Square

We performed a number of simulation studies to compare the methods described in Section 3. We use the unit square in $\mathbb{R}^2$ as the observation window and consider three different processes: the Poisson, the Neymann-Scott and the soft core processes. The second-order statistic considered is the reduced second moment function $K(r)$ for $0.01 \leq r \leq 0.14$ at intervals of 0.01.

For the Poisson process we used an intensity of 250. For the Neymann-Scott process, the parent intensity was set at $\lambda_p = 25$. Each parent had a Poisson number of daughters with mean 10, uniformly distributed on a disc of radius 0.1 centered on the parent. Stoyan, Kendall and Mecke (1995) call this the Matérn cluster field. The soft core process is obtained as follows: a Poisson process of intensity 500 is simulated on the observation window as well as on a border region around it. Each point is given a random radius ρ with probability density function $f_p(\rho) = 800\rho$ for $0 < \rho < 0.05$ and a random mark m uniform on $[0, 1]$. All points $\mathbf{x}'$ with radius ρ' and mark m' are deleted if there is at least one other point in the set of all simulated points that is less than ρ' away with a smaller mark. The remaining points that lie inside the observation window form a realization of the soft core process. This soft core process has an intensity of about 250. Figure 3 shows a simulated realization of each of these three processes.

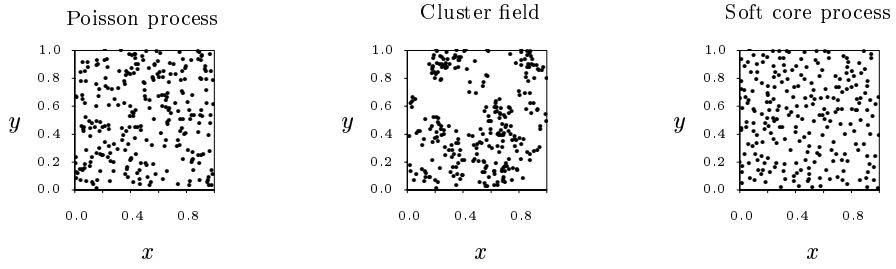

Figure 3. Sample realizations of the three point processes used in the simulation study, with approximately the same intensity of 250.

4.1. Coverage

We simulated 1,000 realizations of each of the three point processes. For each realization we used the splitting, tiling, subsets and marked point methods together with (3) or (5) to compute pointwise nominal 95% confidence intervals. Randomly placed subregions with toroidal wrapping were used for the resampling methods. With each method, we used square subregions with sides of length

0.125, 0.25 and 0.5. Thus for each combination of point process, method and subregion size we have 1,000 confidence intervals for $K(r)$, $0.01 \leq r \leq 0.14$ at 0.01 intervals. The true value of $K(r)$ can be computed directly for the Poisson process and the Matérn cluster field (for the latter, see Stoyan and Stoyan (1994, p.312)). For the soft-core process, we used the mean of the estimates of $\lambda^2 aK(r)$ obtained from $10,000$ simulated realizations of the process, divided by an estimate of $\lambda^2 a$, to be the true value of $K(r)$. We then find the empirical coverage of the different confidence intervals. Figure 4 shows plots of these coverages.

The splitting method is the simplest method and requires the least computation. As such it serves as a benchmark to which the other methods are to be compared. From Figure 4 it appears that the splitting method performs rather well, with coverage close to 95% for both the Poisson and soft core processes. For the clustered process, it also does rather well with subregions of size 0.5. However, coverage is very poor for the smaller subregion sizes.

In all cases, resampling by tiling yields nominal 95% confidence intervals with empirical coverage that is much lower than 95%. The subregion size of 0.125 is clearly not suitable for the clustered process that we used. The cluster size has radius of 0.1. A subregion of size 0.125 will not be able to fully capture even one whole cluster. Since for the Poisson process the distribution of points around a particular position is independent of whether there is an observed point at that position or not, we would expect tiling not to have adverse effects on the second-order properties of the resampled process. Thus it is disappointing that even in this case the coverage is still quite a bit smaller than the 95% nominal level.

For the subsets method, the coverage obtained depends on the size of the subregions used. For all three processes, large subregions produce greater undercoverage. A subregion of size 0.5 is a quarter of the size of A. There is too little variability in the resamples and the spread of $\tilde{K}(r) - \hat{K}(r)$ does not adequately reflect that of $\hat{K}(r) - K(r)$. The subsets method performs well for the Poisson and soft core processes and noticeably worse for the Matérn cluster field.

Using the marked point method, we find smaller differences in coverage between the various subregion sizes. This similarity in coverage of confidence intervals obtained with different subregion sizes is an important and useful property. Often it is difficult to decide what subregion size to use. A method that gives similar results with different subregion sizes will be preferred over another method that gives widely varying confidence intervals. Using subregions of size 0.5 produces slightly less coverage due to the lack of variability in the resamples. Again it is the confidence intervals for the clustered process that do the worst.

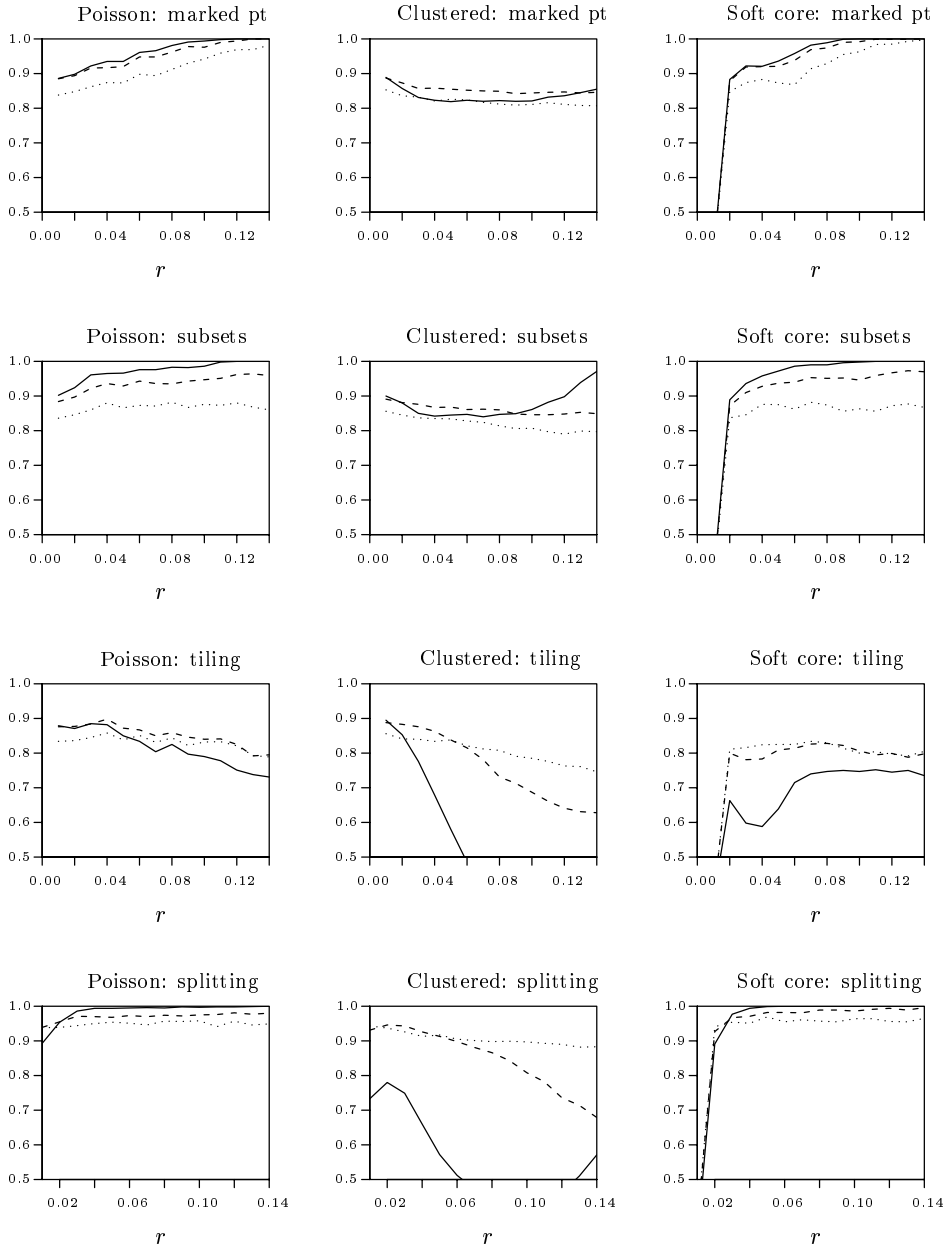

Figure 4. Plots of the empirical coverage of nominal 95% confidence intervals obtained by various methods for the Poisson, Matérn cluster and soft core point processes, using subregion sizes of 0.5 (*dotted line*), 0.25 (*dashed line*) and 0.125 (*solid line*). The plots in each column correspond to the same point process while those in each row correspond to a particular method of obtaining confidence intervals.

Based on this comparison on empirical coverage, the tiling method performs the worst. Empirical coverage is consistently below the nominal 95% level. For large resampling subregions, this may be due to a lack of variability in the resamples. For the smaller subregion sizes, however, it is clear that the underperformance is due to the artificial point pairs produced by putting subregions together. So the most computationally demanding method (see Section 3) is also the weakest method. It is more difficult to decide between the splitting, subsets and marked point methods. The marked point method seems to be more robust with regards to subregion size, with coverage for the subregion of size 0.5 being only slightly lower. The difference across subregion sizes is more noticeable for the subsets method. Choice of subregion size appears to be crucial for the splitting method when applied to a clustered process.

4.2. Widths of the nominal 95% confidence intervals

In this section we compare the widths of nominal 95% confidence intervals obtained by the four methods. The widths of the intervals for the Matérn cluster field behave differently from those for the Poisson and soft core processes, and will be described separately.

For the Poisson and soft core processes, the widths of the intervals obtained by the resampling methods increase with decreasing subregion size. On the other hand, with the splitting method, using the subregion size of 0.25 yielded shorter intervals than using the larger or smaller subregions. At small r, the width of intervals obtained by the resampling methods are comparable and generally shorter than the intervals obtained by the splitting method. At larger r, the tiling method had the smallest intervals. Note, however, that these intervals also have much lower coverage. The main differences in widths between the marked point and subsets methods occur with the smallest subregions where the intervals obtained by the subsets method were consistently wider.

For each r and subregion size, the confidence intervals obtained by the splitting method were the widest. These intervals are sometimes larger by quite a large amount. The only exceptions are $r = 0.14$ with subregion sizes 0.25 and 0.125, where the intervals are slightly narrower. Thus, although the splitting method is a much simpler procedure, the choice of subregion size can be critical and an inappropriate choice may result in very large intervals. This is a major weakness of the splitting method, since it is often unclear what an appropriate subregion size is.

With the clustered process, it is difficult to compare the intervals widths obtained by the tiling method meaningfully since the empirical coverages differ so drastically across subregion size. This is also the case with the splitting method. For the subsets and marked point methods, however, using subregion size of 0.25 produced the widest intervals. With subregion size 0.125, the confidence

intervals obtained by the subsets method become very wide at large r. This may account for the increase in the empirical coverage for $r > 0.10$ (see Figure 4). The coverages obtained by all methods with the subregion size of 0.5 are similar. Here, the splitting method had the widest intervals. The marked point method also has consistently wider intervals than the tiling or subsets method with this subregion size.

4.3. Normalizing coverage

For confidence intervals with similar coverage, we would prefer intervals that are smaller. We compared the widths of the nominal 95% confidence intervals obtained by the different methods in the previous section. However, with their empirical coverages so different, it is difficult to meaningfully compare interval widths.

To better compare the different methods using interval widths as a criterion, we computed confidence intervals so that all of them have empirical coverages of around 85%. We chose 85% because it was a convenient intermediate value to which most of the confidence intervals could be normalized (see below). This normalization is done separately for each r, since the empirical coverages vary with r. For the resampling methods, we used the basic bootstrap interval given in (5), choosing for each r the appropriate α to give empirical coverages of 85%. For the splitting method, we obtained intervals with empirical coverage of 85% by using the interval given at (3), replacing $t_{N-1,\alpha/2}$ with a suitable number. Essentially, for each r, the nominal level of the confidence intervals for $K(r)$ is increased or decreased so that the proportion of intervals that contain $K(r)$ based on simulations under the truth is close to 0.85.

Note that, in practice, this calibration procedure cannot be performed since there will generally be only a single realization of an unknown point process. If calibration is desired, some other method to make a bias correction to estimates, such as the double bootstrap Hinkley and Shi (1989) or the BC_a method Götze and Künsch (1996) must be used. We tried using the double bootstrap to obtain confidence intervals with better empirical coverages, but without success.

Two points regarding our normalization procedure should be mentioned. First, even at the 60% nominal level there were still a few cases with empirical coverage above 90% for $r > 0.10$. Rather than further reduce the nominal coverage, we decided to keep the lowest nominal coverage at 60%. The cases where this occurred were the Poisson and soft core processes using the subsets method and the soft core process using the marked point method, all with subregion sizes of 0.125. Second, with the tiling method using subregions of size 0.125, empirical coverages for the clustered process were lower than 70% for $r \geq 0.10$ even though the nominal levels were 99.8%. Results for these cases are reported together with the others, but this fact should be borne in mind when comparing with other methods.

Tables 1−3, corresponding to the Poisson, Matérn cluster and soft core processes respectively, summarize these results. They show the mean, median, standard deviation and inter-quartile range of the widths of normalized confidence intervals for $K(r)$ for $r = 0.02$, 0.06, 0.10 and 0.14. For the Poisson process, in

Table 1. Descriptive statistics of the widths of normalized confidence intervals for $K(r)$ of the Poisson process, for $r = 0.02, 0.06, 0.10$ and 0.14. The values have been multiplied by 10^5.

Poisson process			Subregion size		
			0.5	0.25	0.125
$r = 0.02$	Splitting	mean (SD)	77 (35)	66 (20)	72 (27)
		median (IQR)	74 (46)	62 (21)	69 (38)
	Tiling	mean (SD)	71 (17)	71 (13)	73 (11)
		median (IQR)	69 (22)	69 (15)	71 (14)
	Subsets	mean (SD)	68 (17)	68 (13)	66 (11)
		median (IQR)	67 (20)	67 (16)	64 (13)
	Marked Pt	mean (SD)	66 (17)	64 (13)	63 (11)
		median (IQR)	64 (22)	63 (15)	62 (13)
$r = 0.06$	Splitting	mean (SD)	225 (100)	192 (55)	175 (18)
		median (IQR)	213 (133)	185 (63)	175 (23)
	Tiling	mean (SD)	213 (45)	228 (35)	257 (31)
		median (IQR)	207 (58)	223 (45)	254 (41)
	Subsets	mean (SD)	205 (44)	199 (32)	202 (26)
		median (IQR)	199 (55)	195 (40)	199 (35)
	Marked Pt	mean (SD)	184 (43)	183 (30)	182 (23)
		median (IQR)	176 (53)	178 (40)	179 (29)
$r = 0.10$	Splitting	mean (SD)	388 (177)	329 (84)	302 (90)
		median (IQR)	360 (228)	320 (98)	286 (50)
	Tiling	mean (SD)	389 (80)	407 (61)	504 (57)
		median (IQR)	378 (106)	398 (76)	498 (72)
	Subsets	mean (SD)	368 (76)	360 (56)	365 (43)
		median (IQR)	359 (97)	351 (70)	359 (56)
	Marked Pt	mean (SD)	311 (59)	303 (41)	302 (29)
		median (IQR)	303 (79)	298 (54)	299 (36)
$r = 0.14$	Splitting	mean (SD)	558 (263)	491 (107)	643 (1223)
		median (IQR)	522 (332)	483 (133)	481 (261)
	Tiling	mean (SD)	655 (134)	714 (111)	832 (95)
		median (IQR)	632 (167)	698 (142)	828 (124)
	Subsets	mean (SD)	594 (119)	581 (87)	1236 (96)
		median (IQR)	576 (149)	567 (107)	1234 (126)
	Marked Pt	mean (SD)	461 (74)	473 (55)	582 (48)
		median (IQR)	451 (90)	466 (72)	578 (63)

almost all cases, the marked point method produced the shortest intervals. These interval widths also tend to have a smaller spread, especially for larger r. The widths of intervals obtained by tiling are the largest in most cases, while those obtained by the splitting method had the largest spreads.

Table 2. Descriptive statistics of the widths of normalized confidence intervals for $K(r)$ of the Matérn cluster field. The values have been multiplied by 10^5.

Matérn cluster			Subregion size		
			0.5	0.25	0.125
$r = 0.02$	Splitting	mean (SD)	228 (322)	267 (298)	171 (65)
		median (IQR)	177 (148)	168 (161)	167 (93)
	Tiling	mean (SD)	175 (74)	153 (53)	154 (49)
		median (IQR)	161 (85)	142 (62)	144 (57)
	Subsets	mean (SD)	169 (72)	156 (54)	152 (49)
		median (IQR)	154 (82)	144 (62)	142 (57)
	Marked Pt	mean (SD)	178 (79)	165 (59)	155 (51)
		median (IQR)	162 (89)	154 (68)	144 (59)
$r = 0.06$	Splitting	mean (SD)	1476 (1164)	1860 (612)	943 (92)
		median (IQR)	1192 (1036)	1000 (804)	943 (126)
	Tiling	mean (SD)	1015 (435)	980 (353)	1217 (429)
		median (IQR)	936 (506)	923 (417)	1130 (514)
	Subsets	mean (SD)	1127 (495)	1012 (369)	959 (320)
		median (IQR)	1037 (583)	947 (434)	902 (379)
	Marked Pt	mean (SD)	1186 (531)	1118 (417)	1054 (351)
		median (IQR)	1090 (616)	1043 (485)	997 (399)
$r = 0.10$	Splitting	mean (std dev)	3209 (2362)	2453 (823)	2145 (304)
		median (IQR)	2671 (2314)	2363 (1113)	2097 (188)
	Tiling	mean (SD)	2494 (1097)	2331 (854)	2324 (785)
		median (IQR)	2295 (1338)	2187 (1024)	2191 (925)
	Subsets	mean (SD)	2460 (1100)	2343 (870)	2224 (736)
		median (IQR)	2267 (1352)	2185 (1050)	2104 (881)
	Marked Pt	mean (SD)	2727 (1229)	2764 (1015)	2547 (804)
		median (IQR)	2500 (1472)	2571 (1238)	2399 (899)
$r = 0.14$	Splitting	mean (std dev)	5217 (3679)	3947 (1219)	4554 (3306)
		median (IQR)	4320 (3541)	3919 (1716)	3746 (1612)
	Tiling	mean (SD)	3859 (1685)	4190 (1582)	3450 (1125)
		median (IQR)	3528 (2138)	3852 (1892)	3256 (1321)
	Subsets	mean (SD)	4158 (1862)	3880 (1451)	3691 (947)
		median (IQR)	3780 (2289)	3609 (1786)	3528 (1173)
	Marked Pt	mean (SD)	4874 (2222)	4482 (1591)	4098 (1162)
		median (IQR)	4318 (2731)	4154 (1904)	3884 (1331)

Table 3. Descriptive statistics of the widths of normalized confidence intervals for $K(r)$ of the soft core process. The values have been multiplied by 10^5.

Soft core process			Subregion size		
			0.5	0.25	0.125
$r = 0.02$	Splitting	mean (SD)	32 (14)	27 (8)	33 (20)
		median (IQR)	31 (18)	26 (10)	29 (18)
	Tiling	mean (SD)	34 (9)	44 (8)	62 (9)
		median (IQR)	33 (10)	43 (10)	61 (11)
	Subsets	mean (SD)	31 (9)	32 (7)	32 (7)
		median (IQR)	31 (11)	31 (9)	32 (8)
	Marked Pt	mean (SD)	28 (8)	28 (7)	27 (6)
		median (IQR)	27 (10)	27 (8)	27 (8)
$r = 0.06$	Splitting	mean (SD)	129 (55)	109 (22)	108 (17)
		median (IQR)	122 (76)	108 (28)	107 (23)
	Tiling	mean (SD)	150 (27)	168 (20)	233 (21)
		median (IQR)	145 (34)	166 (26)	232 (27)
	Subsets	mean (SD)	144 (26)	142 (17)	139 (13)
		median (IQR)	139 (34)	139 (23)	138 (16)
	Marked Pt	mean (SD)	107 (24)	106 (16)	102 (11)
		median (IQR)	103 (33)	104 (21)	101 (14)
$r = 0.10$	Splitting	mean (SD)	211 (93)	184 (42)	207 (185)
		median (IQR)	200 (120)	181 (56)	189 (39)
	Tiling	mean (SD)	250 (41)	280 (33)	326 (28)
		median (IQR)	244 (52)	277 (42)	325 (36)
	Subsets	mean (SD)	238 (39)	228 (28)	231 (19)
		median (IQR)	231 (48)	225 (33)	229 (25)
	Marked Pt	mean (SD)	177 (25)	185 (16)	193 (13)
		median (IQR)	173 (32)	183 (21)	192 (18)
$r = 0.14$	Splitting	mean (SD)	299 (130)	261 (52)	298 (371)
		median (IQR)	287 (184)	257 (65)	233 (114)
	Tiling	mean (SD)	366 (56)	403 (44)	476 (41)
		median (IQR)	358 (66)	398 (55)	473 (53)
	Subsets	mean (SD)	354 (53)	351 (38)	1055 (61)
		median (IQR)	346 (66)	346 (46)	1053 (79)
	Marked Pt	mean (SD)	254 (28)	304 (24)	429 (27)
		median (IQR)	251 (35)	302 (30)	427 (37)

For the Matérn cluster field, the widths look more similar across methods. Differences in widths tend to be small compared to their spread. There does not appear to be a clearly best method. However, the splitting method appears to do the worst and, in almost all cases, one of either the tiling or subsets method

had the shortest normalized confidence intervals.

Finally, with the soft core process, the marked point method appears to be superior, producing intervals with the shortest mean widths in most combinations of subregion size and values of r. The spread of interval widths is also smallest with intervals obtained by the marked point method. For this process the splitting method also does better than the tiling and subsets methods and is on par with the marked point method in some cases. Again, the intervals obtained by the splitting method had the greatest variability in widths. The intervals obtained by tiling are again the largest in most cases.

From the above comparisons, it is clear that the tiling method is the weakest method. Furthermore, it requires the most computation, with complete recalculation of weights for each resample. For the Poisson and soft core processes, the marked point method appears to be the best, with shorter and less variable interval widths. For the Matérn cluster field, there is no clear method of choice. This is not surprising since this is clearly the hardest case, and there may not be any procedure that really works well. The subsets method seems to be slightly better in that the normalized confidence intervals are shorter. More studies should be done to investigate if the superiority of the marked point method indeed holds for a wider range of processes.

5. Further Issues

In this section we investigate the difference toroidal wrapping makes to the empirical coverages of confidence intervals obtained by resampling using the marked point method. We also report results obtained from simulation studies using a combination of rectangular observation windows and subregions. This is an attempt to investigate whether resampling regions should be similar in shape to the observation window.

The tiling method was not satisfactory even for resampling Poisson processes. In this section we look at possible reasons for this inadequacy to effectively resample Poisson processes. Finally, we describe a way to improve on the very poor coverage achieved by all methods for the soft core process at small distances (see Figure 4).

5.1. Toroidal wrapping

To resample a point process observed in a window A, subregions of the same size are randomly placed in A to copy the observed pattern. If toroidal wrapping is not used, the subregions have to be placed completely within A. Then the points of the observed process will not be resampled with equal probability. This causes a bias in the estimates obtained from the resampling process, so that the center of the distribution of $\tilde{K}(r) - \hat{K}(r)$ may not be close to that of $\hat{K}(r) - K(r)$.

Thus we would expect that for two resampling schemes, identical except that one uses toroidal wrapping and the other does not, the one without toroidal wrapping will produce confidence intervals with lower empirical coverages.

We simulated 1000 realizations of the Poisson, Matérn cluster and soft core processes on the unit square. With each realization we resampled the observed pattern using the marked point method, once with toroidal wrapping and once without. Figure 5 shows the empirical coverages for the nominal 95% confidence intervals obtained without wrapping. The plots here should be compared with the plots in the top row of Figure 4. Notice that, with the exception of the clustered process resampled using the smaller subregions, the empirical coverage is considerably lower if toroidal wrapping is not used. The drop in empirical coverage increases with subregion size. This is to be expected, since the positions of large subregions within the window are confined to a smaller area. Thus, for resampling a point process, we suggest using toroidal wrapping.

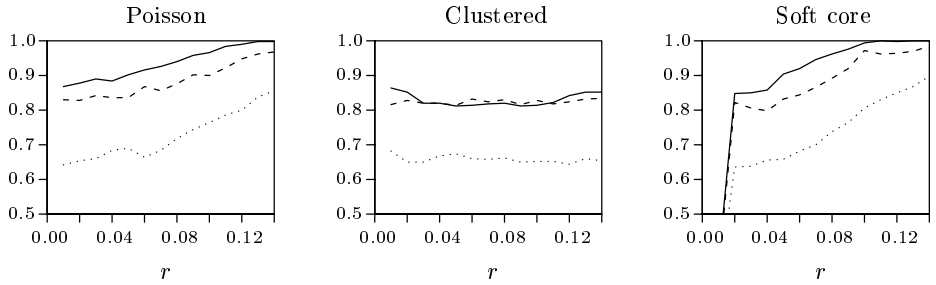

Figure 5. Plots of the empirical coverage of nominal 95% confidence intervals obtained by the marked point method without toroidal wrapping for the Poisson, Matérn cluster and soft core point processes, using subregion sizes of 0.5 (*dotted line*), 0.25 (*dashed line*) and 0.125 (*solid line*).

It is possible to resample points with equal probability without using toroidal wrapping. One way is to randomly place subregions in A and allow them to fall partly outside A. For each subregion A_i, resample only the points in $A \cap A_i$. The last subregion is chosen so that the total resampled area $\sum_i v_d(A \cap A_i)$ equals the area of A. With this method, the actual subregions used are not of the same dimensions, but points are resampled with equal probability. This method of using subregions can be easily applied with the marked point and subsets methods, since these methods do not require resampling subregions to be joined together. For the tiling method, however, it is difficult to recreate the window A by arranging subregions of different sizes. For the marked point method at least, there seems to be little difference between this method of resampling points with equal probability and the one with toroidal wrapping.

We believe that toroidal wrapping, when used in conjunction with the marked point method (but not with the tiling or subsets method), provides a natural and effective way of adjusting for bias due to unequal sampling. Lahiri (1991) suggests accounting for unequal sampling by centering the bootstrapped estimates $\hat{K}(r)$ not at $\hat{K}(r)$ but at the average of the bootstrapped estimates, but we do not pursue that approach here.

5.2. Rectangles versus squares

When resampling a point process, it seems natural to use subregions that are similar in shape to the observation window. For example, Lahiri, Kaiser, Cressie and Hsu (1999) advocate using shape preserving subregions to subsample random fields. We report results from our simulation study using square and rectangular windows and subregions.

We used observation windows of the following sizes: $1 \times 1, 0.8 \times 1.25, 0.5 \times 2, 0.4 \times 2.5$ and 0.25×4, referred to as windows 1 to 5, respectively. All the windows have an area of 1 and represent increasingly elongated rectangular regions. For each observation window we simulated 1,000 realizations of a point process and resampled using five different subregions. The subregion sizes are $0.25 \times 0.25, 0.2 \times 0.3125, 0.125 \times 0.5, 0.1 \times 0.625$ and 0.0625×1, referred to as subregions 1 to 5, respectively. Each of these are respectively similar in shape to observation windows 1 to 5 and have area equal to 0.0625. The resampling method used is the marked point method. For some combinations of windows and subregions, the 16 subregions used to resample the process cannot be arranged to recreate the window. However, with the marked point method, we do not need to arrange the resampled points, so this does not pose a problem.

We find very little difference in empirical coverage between window sizes and subregion sizes regardless of the process considered, and especially so for the Poisson and soft core processes. There is also little difference in interval widths. Based on these criteria of empirical coverage and interval widths, it appears that, for rectangular windows, there does not seem to be any real advantage with using subregions that are similar to the observation window over using other rectangular subregions. To the extent that there is a difference, rectangular subregions may do worse with clustered processes.

Figure 6 shows empirical coverages of nominal 95% confidence intervals of $K(r)$ for the Matérn cluster process, obtained from our simulations. Each plot corresponds to simulations of the process in an observational window. The curves in each plot show empirical coverages obtained by resampling with the marked point method using subregions 1 to 5. There is a very slight drop in empirical coverage as the resampling subregions get progressively more elongated, suggesting that squarer sampling subregions may do better, regardless of the shape of

the observation window. Note that the point process considered here is isotropic. This finding may not hold for anisotropic processes.

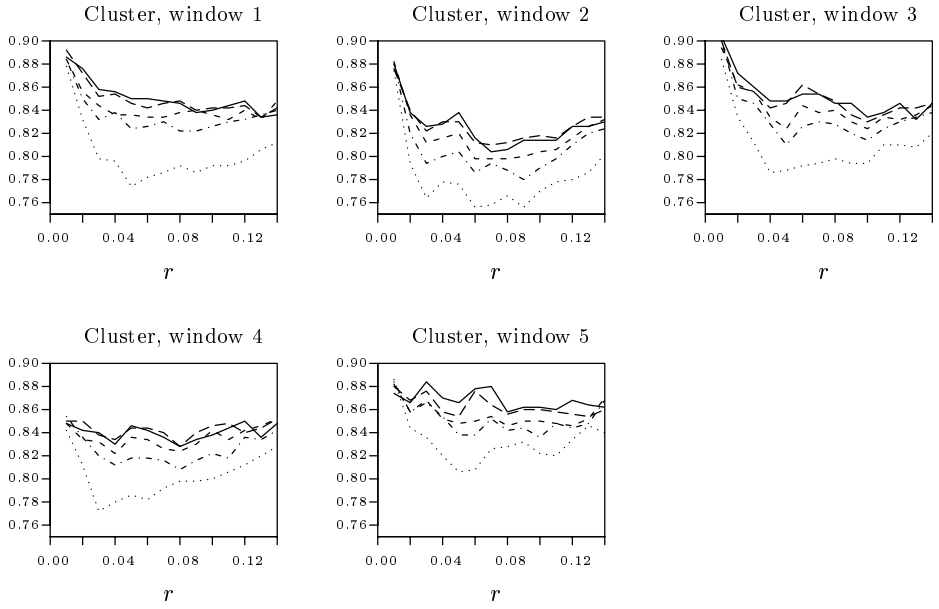

Figure 6. Empirical coverage of nominal 95% confidence intervals obtained by the marked point method for the Matérn cluster field, using subregion sizes of 0.25×0.25 (*solid line*), 0.2×0.3125 (*long dashed line*), 0.125×0.5 (*dashed line*), 0.1×0.625 (*dashed and dotted line*) and 0.0625×1 (*dotted line*). Windows 1 to 5 correspond to windows with sides $1 \times 1, 0.8 \times 1.25, 0.5 \times 2, 0.4 \times 2.5$ and 0.25×4.

5.3. The tiling method for Poisson processes

The tiling method did not do well for Poisson processes, even though for a Poisson process the distribution of points in disjoint regions are independent. With our simulations of the Poisson process on the unit square, we found that the empirical coverage of nominal 95% confidence intervals of $K(r)$ were significantly lower than 95% for all r between 0.01 and 0.14 (Figure 4). This section gives a simple explanation for this undercoverage of the basic bootstrap interval.

Suppose we simulate realizations of a Poisson process on the unit square and obtain estimates $\hat{K}(r)$ of $K(r)$. Consider resampling each realization by tiling using very small subregions, so that each subregion contains only one or no points. This creates resamples that are approximately realizations of a Poisson process, and, as the area of the subregions tend to zero, resampling in this way converges to the parametric bootstrap of generating Poisson resamples with intensity $\hat{\lambda}$.

Of course, if the process is assumed to be Poisson, $K(r)$ would be known and a parametric bootstrap is unnecessary. We do it here simply to show why the tiling method does not work even with Poisson processes.

For this subsection, we use $\hat{K}(r)$ and $\hat{K}_1(r)$ to respectively represent the general estimator of $K(r)$ and the particular estimate of $K(r)$ computed from an observed process. Suppose an observed Poisson process has estimate $\hat{K}_1(r)$. The basic bootstrap interval given at (5) assumes that the distribution of $\tilde{K}(r) - \hat{K}_1(r)$ resembles the sampling distribution of $\hat{K}(r) - K(r)$, with both distributions centered very near zero. However, for the procedure described in the previous paragraph, the resamples are approximate realizations of Poisson processes with intensity $\hat{\lambda}$ and we expect the distribution of $\tilde{K}(r)$ to resemble the distribution of $\hat{K}(r)$ instead, at least in terms of the locations of the distributions. In other words, the bootstrap estimates $\tilde{K}(r)$ behave as replicate observations of $\hat{K}(r)$ and are estimates of $K(r)$, rather than estimates of the value $\hat{K}_1(r)$ of the observed process. Thus, the distribution of $\tilde{K}(r) - \hat{K}_1(r)$ is not centered near zero like the distribution of $\hat{K}(r) - K(r)$. Instead, relative to the sampling distribution of $\hat{K}(r) - K(r)$, it is shifted by about $K(r) - \hat{K}_1(r)$.

Figure 7 shows histograms of $\tilde{K}(r) - \hat{K}_1(r)$ for $r = 0.02$ and 0.10 obtained by bootstrapping a single Poisson realization using the tiling method with subregions of size 0.01. The actual value $\hat{K}_1(r) - K(r)$ for the realization and the mean value of $\tilde{K}(r) - \hat{K}_1(r)$ are also shown (dotted and solid vertical lines respectively). Notice that these two values are about equal in size but with opposite signs. We found this property to hold for other values of r for this particular realization and for five other Poisson realizations we examined. In contrast, the mean value of $\hat{K}(r) - K(r)$ is close to zero.

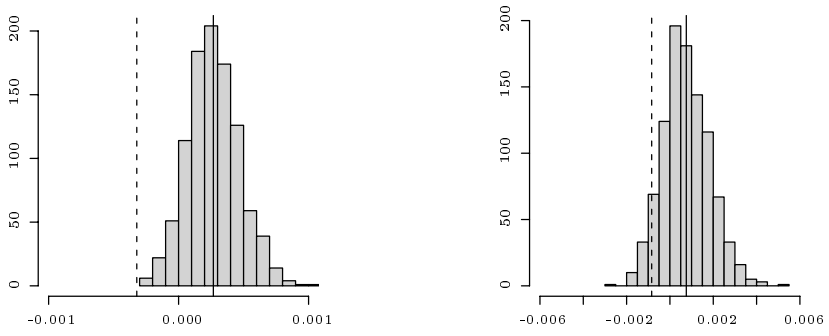

Figure 7. Histograms of $\tilde{K}(r) - \hat{K}_1(r)$ for $r = 0.02$ (*left*) and 0.10 (*right*) obtained by bootstrapping a single Poisson realization using tiling with subregions of size 0.01. Also shown are the actual value $\hat{K}_1(r) - K(r)$ (*dotted line*) for the realization and the mean value of $\tilde{K}(r) - \hat{K}_1(r)$ (*solid line*).

Figure 8 shows the empirical coverage of nominal 95% confidence intervals for $K(r)$ obtained by tiling using subregions of size 0.01, based on bootstrapping 500 realizations of a Poisson process on the unit square. A very similar curve is obtained when we resampled using the parametric bootstrap. Also included in the plot are the empirical coverages obtained using different subregion sizes. Notice that the empirical coverage is much lower than the 95% nominal level and the other curves, indicating that the tiling method does not work well with small subregions. With the tiling method using larger subregions, part of the pattern of the actual realization is preserved and the subregions will overlap more, so that the resamples are not Poisson, but the effect of the bias is still present. Furthermore, using large subregions results in too little variability in $\tilde{K}(r)$. The scale of $\tilde{K}(r) - \hat{K}(r)$ is thus smaller than that of $\hat{K}(r) - K(r)$ and the basic bootstrap interval at (5) does not work well. The implications of this finding for the parametric bootstrap with other point process models is not clear, but the results here suggest the need for caution. Some sort of bias adjustment might improve the tiling method for the Poisson case, but more study needs to be done for the case of an unknown non-Poisson process.

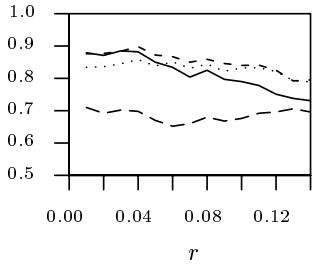

Figure 8. Plot of the empirical coverage of nominal 95% confidence intervals obtained by the tiling method using subregion size of 0.01 for a Poisson process simulated on the unit square (*long-dashed line*). Also included are the plots for subregion sizes of 0.5 (*dotted line*), 0.25 (*dashed line*) and 0.125 (*solid line*).

5.4. Poisson random variable approximation for regular processes

For a regular process, the number of pairs of points separated by distance r or less is very small for small r. This produces some problems when we resample a regular process to obtain confidence intervals for $K(r)$ when r is small. For the soft core process used in our simulations, the number of point pairs within 0.01 apart was less than 10 for a large majority of cases. We found that the empirical coverage of nominal 95% confidence intervals for $K(0.01)$ of this process is less than 0.4 regardless of the resampling method (see Figure 4).

In such circumstances, we can attempt to obtain better confidence intervals by approximating the number of point pairs with a Poisson random variable with mean m_0. The outline of the method of constructing confidence intervals for m_0 is given in Example 9.2.15 of Casella and Berger (2001). For a Poisson random variable with mean m_0, the $100(1-\alpha)\%$ confidence interval for m_0 is $((1/2)\chi^2_{2m,1-\alpha/2}, (1/2)\chi^2_{2m+2,\alpha/2})$, where m is the observed value of the Poisson random variable and χ^2_{2m} is a Chi-squared random variable with $2m$ degrees of freedom. If $m = 0$, $\chi^2_{2m,1-\alpha/2}$ is taken to be 0.

Define an r-close pair as a pair of points at most distance r apart. For the soft core process, we set $m_0 = \mathrm{E}(n_r)$, where n_r is the number of r-close pairs. Since, for r small, $\lambda a \mathrm{E}(N(\mathbf{x}, r)|$ point at $\mathbf{x}) \approx 2m_0$, we have $\lambda^2 a K(r) \approx 2m_0$. Using the Poisson random variable approximation, a $100(1-\alpha)\%$ confidence interval for $K(r)$ is given by

$$\left(\frac{a}{n(n-1)} \chi^2_{2n_r,1-\alpha/2} \, , \, \frac{a}{n(n-1)} \chi^2_{2n_r+2,\alpha/2} \right). \tag{8}$$

Again, we use $n(n-1)/a$ as our estimate of $\lambda^2 a$.

The confidence interval given by (8) can be used together with a resampling method to obtain confidence intervals for $K(r)$ of a regular process. For an observed realization, the number of r-close pairs is counted. The confidence interval at (8) is used for distances r for which the number of r-close pairs is small. For any observed process, an objective way to decide on when to use (8) can be as follows: a cutoff distance r_0 is chosen to be the smallest value of r at which any observed point has at least two r-close pairs. Then, for $r < r_0$, the number of r-close pairs can be considered a count of rare independent events, suitable to be approximated by a Poisson random variable. We obtain confidence intervals using (8) for $r < r_0$ and by resampling for $r \geq r_0$.

We tested whether the Poisson variable approximation will give confidence intervals with better empirical coverage for small r. For each of 500 simulated realizations of the soft core process, we used the criterion described above to compute the cutoff distance. For the soft core process we used, the Poisson variable approximation was used in all the realizations for $r = 0.01$, in about 76% of the realizations for $r = 0.02$, and in just under 1% of the realizations for $r = 0.03$. Beyond $r = 0.03$, resampling was used for all the realizations. Thus only the empirical coverages for $r = 0.01$ to 0.03 were affected by this approximation. With the approximation, the empirical coverages for $r = 0.01$ and 0.02 were 0.98 and 0.96 respectively for all the subregion sizes, and essentially unchanged for $r = 0.03$. Thus our study suggests that the Poisson random

variable approximation can be used to obtain better confidence intervals for $K(r)$ of a process at those values of r with very few point pairs observed.

6. Application to Absorber Catalogs

In this section, we use the marked point and splitting methods in the context of absorber catalogs. We look at the coverage properties and lengths of confidence intervals obtained from simulated clustered catalogs. We also use these two methods to obtain confidence intervals of $K(r)$ for an available absorber catalog.

6.1. Absorber catalogs

Quasi-stellar objects, or QSOs, are extremely bright sources of light and are among the most distant objects known to man. Their bright and focused beams of light can be easily detected from the Earth. When the electro-magnetic spectrum of a QSO is analyzed, absorption lines, i.e., lines of missing electro-magnetic frequencies, can sometimes be detected. These are due to bodies of matter lying between the QSO and the Earth that absorb certain frequencies of the light. The characteristic pattern of the absorption lines can be used to identify the chemical elements present in the matter. These absorption lines are redshifted and the amount of redshift gives the distance of the matter from the Earth. These bodies of matter are called absorption-line systems or absorbers. These absorbers are believed to be gas clouds near distant galaxies and are observed on radial line segments, called lines of sight, between the Earth and QSOs. Of interest is the clustering of these absorbers, which indirectly gives an indication of the clustering of galaxies that are too far away to be easily observed.

Thus, an absorber catalog consists of information about the lines of sight from the Earth to QSOs, their lengths and spatial distribution, as well as the locations of absorbers on these lines. Due to physical reasons (Quashnock, Vanden Berk and York (1996)), these lines do not extend all the way to the Earth nor to the QSOs, but are about 250 to 450 h^{-1} Mpc long. (The quantity h is an unknown dimensionless constant, believed to be between 0.5 and 0.75. The distance 1 h^{-1} Mpc is equal to 3.26 million light years and corresponds to a typical distance between neighboring galaxies.) Note that there is a theoretical possibility of dependence between the point process of absorber centers and the random window, depending on the positions of the QSOs, through which the absorbers are observed. Similarities in the results of clustering obtained from absorber catalogs and galaxy surveys (Quashnock and Stein (1999), Kirshner, Oemler and Schechter (1981)) suggest that this is not a serious problem.

We have available a catalog consisting of 276 lines of sight and 345 Carbon IV absorbers on these lines. These absorbers are selected from a larger

catalog of heavy element absorption systems drawn from the literature, using criteria described in Quashnock, Vanden Berk and York (1996) and Quashnock and Vanden Berk (1998) to produce a homogeneous catalog. An earlier version of the catalog is described in York, Yanny, Crotts, Carilli, Garrison and Matheson (1991) and an updated version is available from Daniel Vanden Berk (danvb@astro.as.utexas.edu). Calculating the weights for estimating $K(r)$ is not straightforward for an absorber catalog, since the observation region consists of many lines in $\mathbb{R}^3$ and the edge effects are huge. Loh, Quashnock and Stein (2001) describe a method to get the weights for an estimate of $K(r)$ that includes using pairs of absorbers lying on different lines of sight. We use the marked point and splitting methods to obtain standard errors for the estimates $\hat{K}(r)$.

Loh, Stein and Quashnock (2003) also developed a model to generate clustered catalogs that mimics the second-order structure and qualitatively captures higher order structure of the Vanden Berk et al. catalog. We use the same model to generate mock clustered catalogs to compare the marked point and splitting methods.

6.2. Resampling mock clustered catalogs

We compare the marked point and splitting methods using simulations of mock clustered absorbers on the lines of sight of the Vanden Berk et al. catalog. The method used to simulate the absorbers is described in Loh, Stein and Quashnock (2003). We take the true value of $K(r)$ to be the average of the estimates $\hat{K}(r)$ of $K(r)$ obtained from 10,000 mock absorber catalogs using all absorber pairs. Then, with 500 new mock catalogs we obtain nominal 95% confidence intervals for $K(r)$ using the marked point and splitting methods and find the proportion of confidence intervals that contain $K(r)$.

We assume as in Loh, Quashnock and Stein (2001) that the absorbers are balls of constant radius ζ, with an observation occurring when a ball intersects a line of sight. (This creates an inherent uncertainty in the estimates of $K(r)$. Specifically, the estimates of $K(r)$ converge to $K(u)$ for some $u \in [r-\zeta, r+\zeta]$. This uncertainty is small since the values of r considered here are much larger than ζ.) The observation window is then the union of all points in $\mathbb{R}^3$ that are at most distance ζ from some line of sight. Instead of a contiguous window, we now have as many windows as the number of lines of sight. The lines of sight are radial line segments contained in a ball S_L with the Earth as the center. This creates some ambiguity regarding how to split up the observation windows and regarding what subregions to use for resampling.

With the splitting method, we can split up the ball S_L into octants or into orange-section-like slices of the same size. We used both in our simulation study.

We also divided the ball into different numbers of slices. There are an infinite number of positions to make the cuts into octants or slices. Since the lines of sight are fixed across simulations of mock catalogs, we chose a priori the positions to make these cuts. These positions are chosen so that the octants or slices have roughly equal total length of lines. By doing this, we ensure that about the same amount of each octant or slice is probed. Specifically, let N be the number of subregions used, $v_1(L_i)$ the total length of lines in subregion i and $v_1(L)$ the total length of lines. We choose the subregions so that $\sum_{i=1}^{N} |v_1(L_i) - v_1(L)/N|$ is minimum for the given set of lines. For the splitting method with slices, we used $N = 2$, 5, 8 and 10.

With the marked point method, it is unclear what type of subregions to use for resampling. A simple way is to resample using slices about the North-South axis of the Earth. We randomly place N slices, each of volume equal to $1/N$ times the volume of S_L. The absorbers lying in these slices are resampled and their marks summed to give a resampled estimate of $\lambda^2 QK(r)$, where $Q = \pi\zeta^2 v_1(L)$. For our simulation study, we used $N = 5$, 8 and 10. It may be desirable to use subregions other than slices for resampling. With slices, whole lines are resampled. While there is nothing inherently wrong with resampling whole lines, there is also no clear reason to do so. We also use balls of fixed radius, placed randomly in a spherical shell in $\mathbb{R}^3$, to resample the absorbers. With balls, we can sometimes resample parts of lines. The center of each ball is determined randomly by simulating the spherical coordinates (R, Θ, Φ) independently, using $f_R(r) \propto r^2$ for $1500 < r < 4000$, $f_\Theta(\theta) = \cos(\theta)$ for $-\pi/2 < \theta < \pi/2$ and $f_\Phi(\phi) = 1/2\pi$ for $0 \leq \phi < 2\pi$. We continue to place more balls until the total length of lines of sight within these balls is approximately equal to the total length of the lines of sight of the original catalog. This ensures that all the resamples contain approximately the same volume of space probed by lines. We used balls of radius 300, 500 and 800 h^{-1} Mpc.

Figure 9 shows the empirical coverages of nominal 95% confidence intervals of $K(r)$ for the mock clustered catalogs, obtained by the splitting and marked point methods described above. Resampling with slices yields slightly conservative confidence intervals for moderate to large r. The empirical coverage drops below 95% for $r < 30$ h^{-1} Mpc, however. The number of slices used has little effect on the empirical coverage of the intervals. The radius of the balls used for resampling also made little difference to the empirical coverage. With resampling using balls, the coverage is roughly the same for all r, slightly below 95%.

The splitting method performs remarkably well, with octants as well as with slices. The empirical coverage is roughly 95% for all r whether octants or slices were used. The large variability in coverage with subregion size that we observed

in our simulations on the unit square does not occur here, at least for the slices we used. A possible reason for this is that the subregions are small relative to the whole region S_L and yet large compared to the distances r considered.

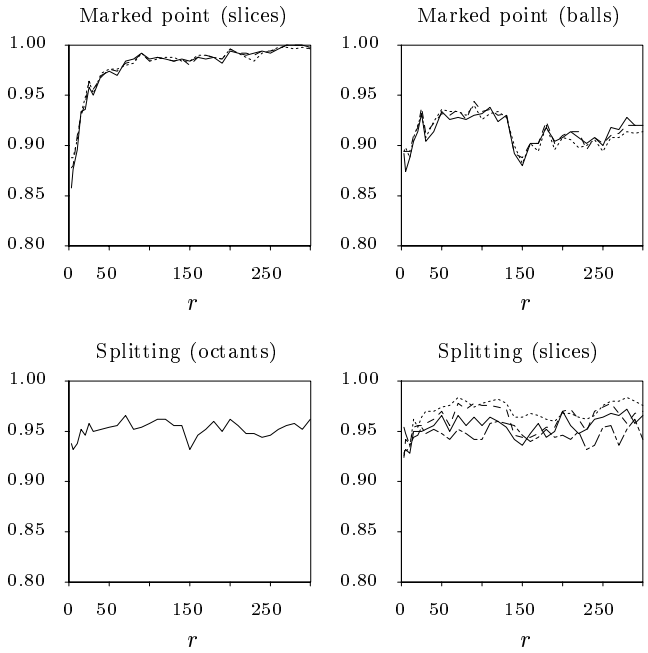

Figure 9. Plots of the empirical coverage of nominal 95% confidence intervals obtained by the marked point and splitting methods for 500 mock clustered catalogs simulated on the lines of sight of the Vanden Berk et al. catalog. The number of slices used are 2 (*dashed and dotted line*, splitting method only), 5 (*solid line*), 8 (*dashed line*) and 10 (*dotted line*), while the balls used have radii 300 (*dotted line*), 500 (*dashed line*) and 800 (*solid line*) h^{-1} Mpc.

Table 4 shows descriptive statistics of the interval widths. The interval widths for the splitting method using two slices are very much longer than those for the other methods. Resampling with the marked point method using balls results in lengths that are quite a bit smaller than those obtained by the splitting method or by resampling using slices. Recall that the marked point method using balls gave confidence intervals that were below 95%. The intervals obtained by the splitting method are slightly shorter than those by resampling using slices. With the exception of using two slices, there seems to be little difference in lengths within each method.

If we normalize the coverage of the confidence intervals obtained by resampling to 95%, using the method described in Subsection 4.3, we find that the

lengths of intervals are about the same whether slices or balls were used for resampling (see Table 5). Notice also that these lengths in Table 5 tend to be slightly shorter than the lengths of intervals obtained by the splitting method (Table 4).

Table 4. Descriptive statistics of the widths of nominal 95% confidence intervals for $K(r)$ of mock clustered catalogs, obtained by the marked point method (with slices and balls) and the splitting method (with slices and octants). These numbers have been divided by 10^4.

		r (h^{-1} Mpc)		
		50	100	150
Marked point method				
number of slices				
5	mean (SD)	73 (19)	1389 (301)	5627 (1093)
	median (IQR)	71 (25)	1377 (380)	5511 (1357)
8	mean (SD)	73 (18)	1378 (286)	5561 (1036)
	median (IQR)	71 (24)	1364 (378)	5493 (1307)
10	mean (SD)	72 (17)	1348 (273)	5417 (975)
	median (IQR)	71 (22)	1330 (361)	5287 (1291)
radius of balls				
300	mean (SD)	52 (9)	819 (160)	3016 (533)
	median (IQR)	51 (12)	796 (189)	2936 (664)
500	mean (SD)	52 (10)	827 (163)	3059 (578)
	median (IQR)	51 (13)	808 (196)	2946 (723)
800	mean (SD)	52 (10)	828 (173)	3067 (616)
	median (IQR)	51 (13)	806 (216)	2963 (790)
Splitting method				
number of slices				
2	mean (SD)	269 (208)	4372 (3358)	16598 (12688)
	median (IQR)	232 (284)	3626 (4831)	13438 (17793)
5	mean (SD)	75 (30)	1201 (500)	4619 (2017)
	median (IQR)	71 (41)	1144 (623)	4238 (2337)
8	mean (SD)	76 (48)	1221 (610)	4872 (2523)
	median (IQR)	69 (29)	1135 (534)	4430 (1925)
10	mean (SD)	83 (64)	1369 (936)	5211 (2855)
	median (IQR)	72 (29)	1198 (530)	4516 (1795)
octants				
	mean (SD)	68 (23)	1074 (376)	4178 (1387)
	median (IQR)	65 (28)	1033 (462)	4000 (1703)

Thus, with our model for generating mock catalogs, the splitting method gives more accurate coverage although at the cost of somewhat wider intervals even after normalization. Considering that in Section 4 we found the marked

point method to perform very well in most situations, we suggest using it together with the splitting method. This application to absorber catalogs should be further studied, preferably with more realistic models for the absorber catalog.

Table 5. Descriptive statistics of the widths of normalized confidence intervals for $K(r)$ of mock clustered catalogs, obtained by the marked point method. These numbers have been divided by 10^4.

		r (h^{-1} Mpc)		
		50	100	150
Marked point method				
number of slices				
5	mean (SD)	60 (15)	1088 (238)	3600 (724)
	median (IQR)	58 (21)	1078 (304)	3535 (942)
8	mean (SD)	60 (15)	1045 (220)	3590 (681)
	median (IQR)	59 (20)	1024 (277)	3529 (866)
10	mean (SD)	59 (14)	1019 (207)	3644 (664)
	median (IQR)	58 (18)	1004 (270)	3546 (859)
radius of balls				
300	mean (SD)	58 (11)	1094 (218)	3608 (642)
	median (IQR)	57 (14)	1070 (257)	3509 (799)
500	mean (SD)	58 (11)	989 (194)	3648 (680)
	median (IQR)	57 (15)	967 (232)	3513 (857)
800	mean (SD)	58 (12)	1099 (226)	3663 (745)
	median (IQR)	56 (15)	1075 (292)	3542 (950)

6.3. Standard errrors for the Vanden Berk et al. catalog

Here, we apply the variants of the marked point and splitting methods used in Subsection 6.2 to the Vanden Berk et al. absorber catalog. Just like in Subsection 6.2, we find little difference between the intervals when we vary the radius of the balls used for resampling or the number of slices used.

Figure 10 shows plots of $\hat{K}(r)$ divided by $4\pi r^3/3$ together with pointwise nominal 95% confidence intervals obtained by the marked point (left column) and splitting methods (right column). The resampling was done with balls of radius 300 h^{-1} Mpc (top plot) and with 8 randomly placed slices (bottom plot). The intervals obtained by resampling with balls tend to be slightly shorter. This agrees with the finding in Subsection 6.2 (Table 4) when resampling mock clustered catalogs. For the splitting method we show plots for splitting the ball S_L into octants (top plot) and into 8 slices (bottom plot).

Confidence bands for $K(r)$ can be obtained from the bootstrap estimates $\tilde{K}(r)$ using an empirical approach outlined in Davison and Hinkley (1997, p.154): arrange the R sets of $\tilde{K}(r)$ into rows and, for each r, rank $\tilde{K}(r)$. Let n be the

number of rows with at least one rank $\leq k$ or $\geq R+1-k$. The kth and $(R+1-k)$th ordered values of $\tilde{K}(r)$ then yield a $100(1-n/R)\%$ confidence band for $K(r)$. Figure 11 shows approximate 94% confidence bands for $K(r)$ obtained in this way using bootstrap estimates from resampling with the marked point method using balls of radius 300 h^{-1} Mpc.

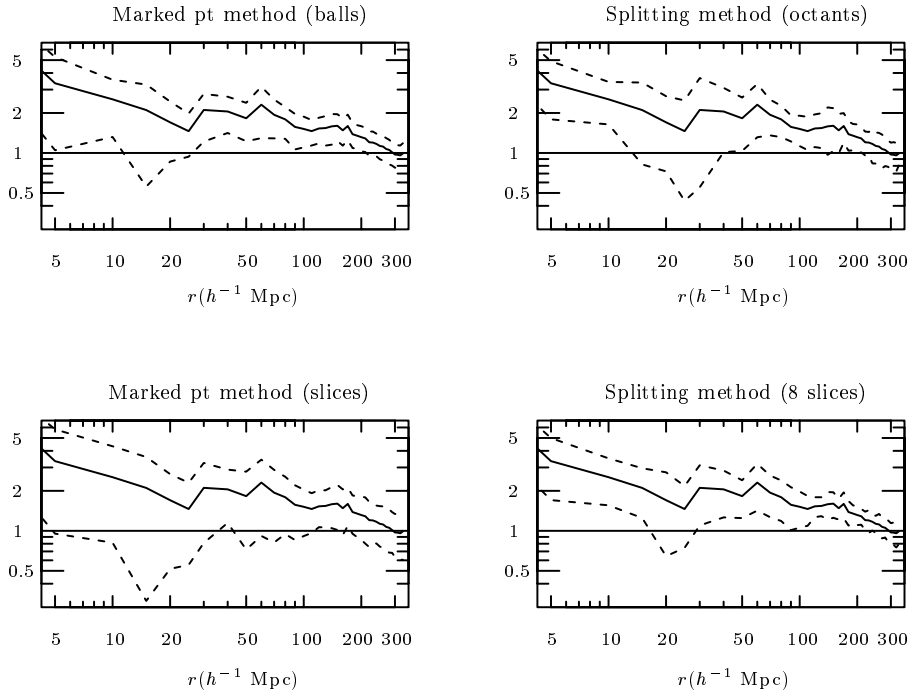

Figure 10. Plots of $\hat{K}_I(r)$ divided by $4\pi r^3/3$ together with pointwise nominal 95% confidence intervals obtained by the marked point method, with balls of radius 300 h^{-1} Mpc and with 8 randomly placed slices, and the splitting method using octants and 8 slices.

We find evidence for clustering on scales up to about 100 h^{-1} Mpc and possibly even up to 150 h^{-1} Mpc. However, this is far from conclusive. The Sloan Digital Sky Survey is an ongoing project to map about a quarter of the sky and is expected to find as many as 100,000 QSOs. With a dataset of this size, we will be able to make stronger statements regarding the existence of clustering on these scales. In agreement with the cosmological principle, which states that the universe is homogeneous on large scales, we find no evidence for clustering at scales larger than 200 h^{-1} Mpc.

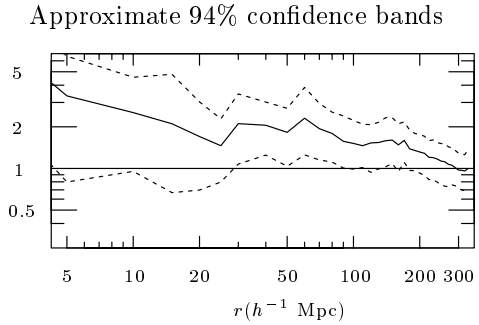

Figure 11. Plot of $\hat{K}_I(r)$ divided by $4\pi r^3/3$ together with approximate nominal 94% confidence bands (see text).

References

Braun, W. J. and Kulperger, R. J. (1998). A bootstrap for point processes. *J. Statist. Comput. Simulation* **60**, 129-155.

Bühlmann, P. and Künsch, H. R. (1995). The blockwise bootstrap for general parameters of a stationary time series. *Scand. J. Statist.* **22**, 35-54.

Casella, G. and Berger, R. L. (2001). *Statistical Inference.* 2nd edition. Wadsworth and Brooks/Cole, California.

Davison, A. C. and Hinkley, D. V. (1997). *Bootstrap Methods and their Applications.* Cambridge University Press, Cambridge.

Götze, F. and Künsch, H. R. (1996). Second-order correctness of the blockwise bootstrap for stationary observations. *Ann. Statist.* **24**, 1914-1933.

Hall, P. (1985). Resampling a coverage pattern. *Stochastic Process. Appl.* **20**, 231-246.

Hinkley, D. V. and Shi, S. (1989). Importance sampling and the nested bootstrap. *Biometrika* **76**, 435-446.

Kirshner, R. P., Oemler, A. and Schechter, P. L. (1981). A million cubic megaparsec void in Boötes? *Astrophys. J. Lett.* **248**, 57-60.

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

Lahiri, S. N. (1991). Second order optimaility of stationary bootstrap. *Statist. Probab. Lett.* **11**, 335-341.

Lahiri, S. N. (1992). Edgeworth correction by 'moving block' bootstrap for stationary and non-stationary data. In *Exploring the Limits of Bootstrap* (Edited by R. LePage and L. Billard), 183-214. Wiley, New York.

Lahiri, S. N. (1993). On the moving block bootstrap under long range dependence. *Statist. Probab. Lett.* **18**, 405-413.

Lahiri, S. N., Kaiser, M. S., Cressie, N. and Hsu, N.-J. (1999). Prediction of spatial cumulative distribution functions using subsampling. *J. Amer. Statist. Assoc.* **94**, 86-110.

Liu, R. Y. and Singh, K. (1992). Moving blocks jackknife and bootstrap capture weak dependence. In *Exploring the Limits of Bootstrap* (Edited by R. LePage and L. Billard), 225-248. Wiley, New York.

Loh, J. M., Quashnock, J. M. and Stein, M. L. (2001). A measurement of the three-dimensional clustering of C IV absorption-line systems on scales of 5-1000 h^{-1} Mpc. *Astrophys. J.* **560**, 606-616.

Loh, J. M., Stein, M. L. and Quashnock, J. M. (2003). Estimating the large-scale structure of the universe using QSO Carbon IV absorbers. *J. Amer. Statist. Assoc.* **98**, 522-532.

Nguyen, X. X. and Zessin, H. (1979). Ergodic theorems for spatial processes. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **48**, 133-158.

Ohser, J. (1983). On estimators for the reduced second moment measure of point processes. *Mathematische Operationsforschung und Statistik. series Statistics* **14**, 63-71.

Politis, D. N. and Romano, J. P. (1992a). A circular block-resampling procedure for stationary data. In *Exploring the Limits of Bootstrap* (Edited by R. LePage and L. Billard), 263-270. Wiley, New York.

Politis, D. N. and Romano, J. P. (1992b). A general resampling scheme for triangular arrays of α-mixing random variables with application to the problem of spectral density estimation. *Ann. Statist.* **20**, 1985-2007.

Politis, D. N. and Romano, J. P. (1994). Large sample confidence regions based on subsamples under minimal assumptions. *Ann. Statist.* **22**, 2031-2050.

Quashnock, J. M. and Stein, M. L. (1999). A measure of clustering of QSO heavy-element absorption-line systems. *Astrophys. J.* **515**, 506-511.

Quashnock, J. M. and Vanden Berk, D. E. (1998). The form and evolution of the clustering of QSO heavy-element absorption-line systems. *Astrophys. J.* **500**, 28-36.

Quashnock, J. M., Vanden Berk, D. E. and York, D. G. (1996). High-redshift superclustering of quasi-stellar object absorption-line systems on 100 h^{-1} Mpc scales. *Astrophys. J. Lett.* **472**, 69-72.

Ripley, B. D. (1988). *Statistical Inference for Spatial Processes.* Wiley, New York.

Stein, M. L. (1993). Asymptotically optimal estimation for the reduced second moment measure of point processes. *Biometrika* **80**, 443-449.

Stoyan, D., Kendall, W. S. and Mecke, J. (1995). *Stochastic Geometry and Its Applications.* 2nd edition. John Wiley, New York.

Stoyan, D. and Stoyan, H. (1994). *Fractals, Random Shapes and Point Fields.* John Wiley, New York.

York, D. G., Yanny, B., Crotts, A., Carilli, C., Garrison, E. and Matheson, L. (1991). An inhomogeneous reference catalogue of identified intervening heavy element systems in spectra of QSOs. *Monthly Notices Roy. Astronom. Soc.* **250**, 24-49.

Department of Statistics, Columbia University, MC 4403, 2990 Broadway, New York, NY 10027, U.S.A.

E-mail: meng@stat.columbia.edu

Department of Statistics, University of Chicago, 5734 University Ave, Chicago, IL 60637, U.S.A.

E-mail: stein@galton.uchicago.edu

(Received April 2002; accepted May 2003)