

Bias correction and bootstrap methods for a spatial sampling scheme

PETER HALL¹, GAVIN MELVILLE^{1,2} and ALAN H. WELSH¹

¹*Centre for Mathematics and its Applications, Australian National University, Canberra, ACT 0200, Australia. E-mail: halpstat@fac.anu.edu.au*

²*NSW Agriculture, PMB 19, Trangie, NSW 2823, Australia.*

Motivated by sampling problems in forestry and related fields, we suggest a spatial sampling scheme for estimating the intensity of a point process. The technique is related to the ‘wandering quarter’ method. In applications where the cost of identifying random points is high relative to the cost of taking measurements, for example when identification involves travelling within a large region, our approach has significant advantages over more traditional approaches such as T-square sampling. When the point process is Poisson we suggest a simple bias correction for a ‘naive’ estimator of intensity, and also discuss a more complex estimator based on maximum likelihood. A technique for pivoting, founded on a fourth-root transformation, is proposed and shown to yield second-order accuracy when applied to construct bootstrap confidence intervals for intensity. Bootstrap methods for correcting edge effects and for addressing non-Poisson point-process models are also suggested.

Keywords: boundary effect; confidence interval; edge effect; forestry; intensity estimation; pivotal statistic; Poisson process; T-square sampling; wandering quarter sampling

Introduction

The difficulty and importance of sampling spatial point processes to estimate their intensity have led to the development of a number of different approaches; see, for example, the monographs by Cliff and Ord (1981), Ripley (1981, 1988), Diggle (1983), Upton and Fingleton (1985) and Cressie (1991). One of the most successful approaches is the method of T-square sampling, which is based on measuring the distance from a random point to the nearest object and then the distance to the next nearest object. However, the need to select a number of random points, and the burden of actually locating these points in the field, reduce the attractiveness of T-square sampling in an area such as forestry where the cost of locating random points is large relative to the cost of taking measurements at these points. The ‘wandering quarter’ method proposed by Catana (1963) overcomes these difficulties by reducing the required number of random starting points. This is achieved by following the initial point-to-object distance by a chain of object-to-object distances. With the notable exception of O’Hara Hines and Hines (1989), who used data collected by wandering quarter sampling to test for spatial randomness, the method has been little used and, in the words of Diggle (1983), remains ‘an ingenious sampling procedure whose statistical potential appears not to have been tapped’. One reason for its slow adoption is that the use of longer chains of

object-to-object distances brings about an increased likelihood of boundary problems – either of reaching the end of the region in less than n steps or of encountering the side boundary.

In this paper we propose methods which overcome these difficulties. We suggest several point estimators of intensity, including one based on a simple bias correction and another on maximum likelihood. Two bootstrap methods founded on fourth-root transformations, which produce pivotal statistics, are suggested for constructing interval estimators. Another bootstrap approach is proposed for correcting errors induced by edge effects, and yet another for accommodating non-Poissonness of the sampled point process. The performance of our methods is described both theoretically and empirically, the latter including an application to data on the spatial distribution of a Western Australian plant species. For simplicity we shall often refer to the technique as the ‘wandering quarter’, although it is more general than the method of that name introduced by Catana (1963).

In principle, the wandering quarter method can also be used to count moving objects such as wild animals and whales. In practice, however, these applications can be hindered by the need to determine the object locations. A wild animal is often noted at a considerable distance from the observer, and its exact position can be difficult to determine without sophisticated electronic equipment. Furthermore, the relationship between the observer and the animal has an impact on the spatial distribution of the animal population, which consequently can be difficult to model. Also, when counting members of an animal population in a non-instantaneous way one needs to ensure that no animal is counted more than once; for some species that can be very challenging unless the animal is actually captured.

To describe a general version of our sampling procedure, suppose the objects of interest are of infinitesimal size and are located in a well-defined region. We begin the construction of a single chain by determining a fixed orientation, called the sampling direction, in which to sample, and then selecting a random point P_0 in the region. From any point P_{i-1} , the next point in the chain is the location P_i of the nearest object in the region within the planar ‘cone’ which has its apex at P_{i-1} , its axis pointing in the sampling direction, and which subtends the angle $2\theta \leq \pi$ at its vertex. That is, at each step we include in our sample the next nearest object within the sampling cone. This can be described as stepping-stone sampling: if we think of the region as a river and the objects as stepping stones, we cross the river on stones selected by starting from a random point and always moving to the next nearest stone, within the sampling cone, in the forward direction. We recover Catana’s wandering quarter method when $\theta = \pi/4$. The process is stopped either when there are no points in the cone between the current point and the boundary of the region, or after a predetermined number n of steps, whichever occurs soonest.

We record the length of each step, the orientation of each step relative to the sampling direction, and the number of steps until we stop. Thus, the data from a single chain are $\{N, X_1, \dots, X_N, \Theta_1, \dots, \Theta_N\}$, where N equals the number of steps from P_0 before stopping, X_i denotes the length of the i th step, and Θ_i equals the angle subtended to the axis of the cone by the vector representing the i th step. From these data we wish to estimate the intensity, λ , of the point process. If we stop after n steps, and if the process is Poisson, then there is no information in the Θ_i about λ , and so we may use $\{X_1, \dots, X_n\}$ instead.

We adopt two simplifications in our development. First, we consider only a single chain; results for multiple independent chains are then straightforward to obtain. Second, nearly all the available theory for distance-based methods for spatial sampling requires the simplifying assumption that the objects of interest are located at the points of a homogeneous Poisson process of intensity λ . Either, this assumption is adopted as the null hypothesis under test, as in O'Hara Hines and Hines (1989), or, under the Poisson assumption, the objective is to make inference about λ . The homogeneous Poisson process has limitations as a description of the spatial distribution of plants or animals, but it allows us to develop theoretical results which enable us both to understand the simplest case and the limitations of methodology developed specifically for it. Under more general point-process models, such as those of Bartlett and Lewis or Neyman and Scott, we may take the estimators suggested here as a starting point and construct iterative corrections via Monte Carlo experiments (see Section 3).

Let $w \leq \infty$ be the distance in the sampling direction from the initial point, P_0 , to the boundary of the region. The two cases of interest are: w large relative to n , so that we stop after n steps; and w small relative to n , so that we stop on reaching the boundary after a random number of steps, N . It is convenient to represent these cases in abstract discussion as $w = \infty$ and $w < \infty$, respectively. Our interest is primarily in the context of random-length chains ($w < \infty$), but we also discuss fixed-length chains ($w = \infty$) where relevant. For random-length chains we derive the maximum likelihood estimator and, because it is computationally complex, several simpler forms. In conjunction with these estimators we propose the use of parametric bootstrap methods, both to set confidence intervals based on random-length chains and to adjust for bias and edge effects in fixed- and random-length chains.

These methods, which are described in Section 2, enable us to overcome the difficulties posed by wandering quarter sampling when the sampled process is Poisson. Section 3 describes iterated-bootstrap techniques for correcting biases when either there are edge effects or the sampled process is modelled in a relatively complex way. In Section 4 we outline theoretical results describing properties of our methods. Section 5 addresses empirical properties, including an application to real data. Technical details are deferred to an Appendix.

2. Methodology

2.1. Estimators

Under the assumption of a homogeneous Poisson process, successive interpoint distances in our sampling scheme can be written in the form $X_i = (tZ_i)^{1/2}$, where $t = (\lambda\theta)^{-1}$, Z_i has an exponential distribution with unit mean, and the variables Z_1, Z_2, \dots are independent. Hence, when $w = \infty$ the maximum likelihood estimator of λ is given by

$$\hat{\lambda} = \left(n^{-1} \theta \sum_{i=1}^n X_i^2 \right)^{-1}. \tag{2.1}$$

In the case $w < \infty$, the maximum likelihood estimator $\hat{\lambda}_{ML}$ minimizes the function

$$\ell(\lambda) \equiv N \log \lambda - \lambda \theta \sum_{i=1}^N X_i^2 + \log \left[\int_0^\theta \exp \left\{ -\lambda \theta (\cos \omega)^{-2} \left(w - \sum_{i=1}^N X_i \cos \Theta_i \right)^2 \right\} d\omega \right]. \tag{2.2}$$

To appreciate why, put $Y_i = X_i^2$ and note that, since the variables $Z_1, \dots, Z_{n+1}, \Theta_1, \dots, \Theta_{n+1}$ are independent, the joint density of $Y_1, \dots, Y_{n+1}, \Theta_1, \dots, \Theta_{n+1}$ is

$$f_{n+1}(y_1, \dots, y_{n+1}, \theta_1, \dots, \theta_{n+1}) = (2t\theta)^{-(n+1)} \exp \left(-t^{-1} \sum_{i=1}^{n+1} y_i \right),$$

for $y_i > 0$ and $|\theta_i| \leq \theta$. The probability $p_n = P(N = n)$ equals the integral of f_{n+1} over the region defined by

$$\sum_{i=1}^n y_i^{1/2} \cos \theta_i \leq w, \quad \sum_{i=1}^{n+1} y_i^{1/2} \cos \theta_i \geq w. \tag{2.3}$$

Conditional on $N = n$, the joint density of $Y_1, \dots, Y_n, \Theta_1, \dots, \Theta_n$ equals the integral of f_{n+1}/p_n over those values of y_{n+1} and θ_{n+1} that satisfy the second identity in (2.3), for fixed y_1, \dots, y_n and $\theta_1, \dots, \theta_n$ – that is,

$$y_{n+1}^{1/2} \cos \theta_{n+1} > w - \sum_{i=1}^n y_i^{1/2} \cos \theta_i.$$

Therefore, defining $y = (y_1, \dots, y_n), \bar{\theta} = (\theta_1, \dots, \theta_n)$ and

$$g_n(y, \bar{\theta}) = \theta^{-1} \int_0^\theta \exp \left\{ -t^{-1} (\cos \omega)^{-2} \left(w - \sum_{i=1}^n y_i^{1/2} \cos \theta_i \right)^2 \right\} d\omega,$$

we see that the probability that $N = n, Y_i \in (y_i, y_i + dy_i)$ and $\Theta_i \in (\theta_i, \theta_i + d\theta_i)$ for $1 \leq i \leq n$ equals

$$f_n(y, \bar{\theta}) g_n(y, \bar{\theta}) dy_1 \dots dy_n d\theta_1 \dots d\theta_n.$$

It follows that the negative log-likelihood equals the function at (2.2), up to terms that do not depend on λ .

The integral in the likelihood, and the resulting complexity of the calculations, motivate consideration of more elementary estimators. An analogue of the estimator in (2.1) is the ‘simple estimator’,

$$\hat{\lambda}_{\text{simp}} = \left(N^{-1} \theta \sum_{i=1}^N X_i^2 \right)^{-1}. \tag{2.4}$$

As we shall show in Section 4, $\hat{\lambda}_{\text{simp}}$ is biased. However, the majority of bias is readily removed using an adjusted form of the estimator,

$$\hat{\lambda}_{\text{adj}} = \left\{ \left(N - \frac{3}{2} \right)^{-1} \theta \sum_{i=1}^N X_i^2 \right\}^{-1}. \tag{2.5}$$

2.2. Properties of estimators

First consider the case $w = \infty$. Since $t^{-1} \sum_{i \leq n} X_i^2$ has a gamma distribution with n degrees of freedom, $\tilde{\lambda} \equiv (1 - n^{-1})\hat{\lambda}$ is unbiased for λ . If (a_1, a_2) denotes an interval in which a gamma random variable with n degrees of freedom lies with probability α , then $(\hat{\lambda}a_1/n, \hat{\lambda}a_2/n)$ is an exact α -level confidence interval for λ . Standard asymptotic arguments show that both $\hat{\lambda} - \lambda$ and $\tilde{\lambda} - \lambda$ are asymptotically normally distributed with zero mean and variance λ^2/n .

The theory is more complex when $w < \infty$, and will be dealt with in detail in Section 4. We shall show there that the estimators $\hat{\lambda}_{\text{ML}}$, $\hat{\lambda}_{\text{simp}}$ and $\hat{\lambda}_{\text{adj}}$ are all asymptotically normally distributed with asymptotic variance λ^2/ν , where $\nu = \nu(w) = w/E(X \cos \Theta)$ and (X, Θ) has the distribution of a generic (X_i, Θ_i) . The quantity ν plays a role similar to that of n in the case $w = \infty$, and is asymptotic to N , in the sense that $N/\nu \rightarrow 1$ almost surely as $\nu \rightarrow \infty$. Now, $\nu \sim C\lambda^{1/2}w$ as $w \rightarrow \infty$, where C is a constant depending only on θ ; the proof will be outlined in Section 4. Therefore, the asymptotic variances of $\hat{\lambda}_{\text{ML}}$, $\hat{\lambda}_{\text{simp}}$ and $\hat{\lambda}_{\text{adj}}$ are all equal to $C^{-1}\lambda^{3/2}w^{-1}$. The latter property will play a major role in our bootstrap method.

Since $N/\nu \rightarrow 1$, the variance properties when $w < \infty$ are directly analogous to those when $w = \infty$: in asymptotic terms the variance in both cases, and for each estimator, equals λ^2 multiplied by the inverse of the number of steps taken. Note, however, that when $w < \infty$ the actual bias and variance are in general not well defined unless we condition on $N \geq 2$. Both $\hat{\lambda}_{\text{ML}}$ and $\hat{\lambda}_{\text{simp}}$ have biases of order w^{-1} , whereas $\hat{\lambda}_{\text{adj}}$ has bias of order w^{-2} , as $w \rightarrow \infty$.

2.3. Bootstrap confidence intervals when $w < \infty$

It does not seem possible to construct exact confidence intervals when $w < \infty$, although bootstrap methods give a high degree of accuracy. As a prelude to describing the methods, let $\hat{\Lambda}(w)$ denote any of the three estimators $\hat{\lambda}_{\text{ML}}$, defined by minimizing $\ell(\lambda)$ in (2.2), $\hat{\lambda}_{\text{simp}}$ and $\hat{\lambda}_{\text{adj}}$. Let the stochastic process $\hat{\Lambda}_0(\cdot)$ have the distribution of $\hat{\Lambda}(\cdot)$ when λ , the true value of the parameter, equals 1. We take $\hat{\Lambda}_0(\cdot)$ to be independent of the sample $\chi = \{N, X_1, \dots, X_N, \Theta_1, \dots, \Theta_N\}$, and allow the argument v of $\hat{\Lambda}_0(v)$ to be a function of the data, denoting v by \hat{v} to indicate this property.

Note particularly that, when the true parameter value is λ , the distribution of $\hat{\Lambda}(w)/\lambda$ is

identical to that of $\hat{\Lambda}_0(\lambda^{1/2}w)$. Also, since the asymptotic variance of $\hat{\Lambda}(w)$ equals $C^{-1}\lambda^{3/2}w^{-1}$, where C is as in Section 2.2, an approximate variance-stabilizing transformation is obtained by taking the fourth root of $\hat{\Lambda}$. These properties are the key to our bootstrap methods, which are as follows.

Given \hat{v} , we may approximate the distribution of $\hat{\Lambda}_0(\hat{v})$, conditional on χ , to arbitrarily high accuracy using Monte Carlo methods. Define $\hat{v} = \hat{\Lambda}(w)^{1/2}w$, and take the conditional distribution of $\hat{\Lambda}(w)^{1/4}\{\hat{\Lambda}_0(\hat{v})^{1/4} - 1\}$ (or $\hat{\Lambda}(w)^{1/4}\{\hat{\Lambda}_0(\hat{v}) - 1\}$), given the data, to be our approximation to the unconditional distribution of $T_1(\lambda) \equiv \hat{\Lambda}(w)^{1/4} - \lambda^{1/4}$ (or $T_2(\lambda) \equiv \hat{\Lambda}(w)\lambda^{-3/4} - \lambda^{1/4}$). Both T_1 and T_2 are continuous, strictly monotone decreasing functions, and so are appropriate as the basis for confidence intervals for λ . The function T_1 has the advantage that the equation $T_1(\lambda) = x$ always admits a unique solution $\lambda > 0$, regardless of the value of $x \in \mathbb{R}$. This is not true of the equation $T_2(\lambda) = x$. The approximations are second-order accurate, in that they capture terms of order $w^{-1/2}$, as $w \rightarrow \infty$, in the difference between the distribution of $\hat{\Lambda}(w)$ and its limiting form. Consequently, the errors of approximation are $O_p(w^{-1})$.

This bootstrap approximation is readily used to construct confidence intervals for λ with coverage accuracy $O(w^{-1})$. In particular, if \hat{a}_1, \hat{a}_2 are functions of the data, and of w , such that

$$P\{\hat{a}_1 \leq \hat{\Lambda}(w)^{1/4}\{\hat{\Lambda}_0(\hat{v})^{1/4} - 1\} \leq \hat{a}_2 | \chi\} = \alpha,$$

$$P\{\hat{b}_1 \leq \hat{\Lambda}(w)^{1/4}\{\hat{\Lambda}_0(\hat{v}) - 1\} \leq \hat{b}_2 | \chi\} = \alpha,$$

then

$$P\{\hat{a}_1 \leq \hat{\Lambda}(w)^{1/4} - \lambda^{1/4} \leq \hat{a}_2\} = \alpha + O(w^{-1}),$$

$$P\{\hat{b}_1 \leq \hat{\Lambda}(w)\lambda^{-3/4} - \lambda^{1/4} \leq \hat{b}_2\} = \alpha + O(w^{-1}).$$

Equivalently, if $\lambda = \tilde{\lambda}_i$ denotes the solution of the equation $\hat{a}_i = \hat{\Lambda}(w)^{1/4} - \lambda^{1/4}$, or of $\hat{b}_i = \hat{\Lambda}(w)\lambda^{-3/4} - \lambda^{1/4}$, for $i = 1$ and 2 , then the confidence interval $(\tilde{\lambda}_2, \tilde{\lambda}_1)$ covers λ with probability $\alpha + O(w^{-1})$; see Theorem 5.2.

3. Bias correction by simulation

3.1. Adjustment for bias caused by edge effects

We first suggest corrections for biases caused by the intrusion of boundaries into calculation of the estimator. Depending on the location of the initial point P_0 , there can be a significant non-zero probability that one or more of the cones that define the steps will protrude beyond the left- and right-hand boundaries before we reach the point at which we stop. Naive application of the methods suggested earlier may then lead to significant biases, which can be corrected using the bootstrap as follows.

Compute a pilot estimator $\hat{\lambda}_1$ of λ , perhaps just $\hat{\lambda}_{\text{simp}}$, and simulate from a homogeneous Poisson process with intensity $\hat{\lambda}_1$ in a region of the same size and shape as the actual

region. Without bothering with any corrections, compute the bootstrap version $\hat{\lambda}_{\text{simp}}^*$ of $\hat{\lambda}_{\text{simp}}$ for the simulated data; and by repeated simulation of the same process, calculate a Monte Carlo approximation to $\tilde{\lambda}_1 = E'(\hat{\lambda}_{\text{simp}}^*|\chi)$, where $E'(\cdot|\chi)$ denotes expectation conditional on χ and on the bootstrap version of N being at least 1. Put $\hat{\lambda}_2 = \hat{\lambda}_1^2/\tilde{\lambda}_1$. Now iterate the procedure, defining, at the j th step, first $\tilde{\lambda}_j = E'(\hat{\lambda}_{\text{simp}}^*|\chi)$, where $\hat{\lambda}_{\text{simp}}^*$ is now derived using a simulated homogeneous Poisson process with intensity $\hat{\lambda}_j$, and then $\hat{\lambda}_{j+1} = \hat{\lambda}_1\hat{\lambda}_j/\tilde{\lambda}_j$. At the j th step the bias-corrected estimator is $\hat{\lambda}_j$. Stop when numerical convergence is achieved.

3.2. Adjustment for bias caused by non-Poissonness

It is generally the case that statistical techniques proposed in the context of independent data form a basis for methods for dependent data. The present setting is no exception. While the bias-correction properties of Poisson-based estimators such as $\hat{\lambda}_{\text{adj}}$ are unlikely to be available for non-Poisson data, the estimators can be expected to be reasonably unbiased if the point process is not far from Poisson, and that is in itself a recommendation. We suggest basing general point estimators of λ on $\hat{\lambda}_{\text{simp}}$ or $\hat{\lambda}_{\text{adj}}$, modelling the departure of the sampled point process from Poissonness, and borrowing ideas from Section 3.1 so as to correct for bias.

Specifically, let $\hat{\lambda}$ be an estimator such as $\hat{\lambda}_{\text{simp}}$ or $\hat{\lambda}_{\text{adj}}$ which we wish to correct for bias, and let $\mathcal{M}(\lambda)$ be a model (e.g. Neyman–Scott) for the point process, depending on the unknown value of λ . Regard $\hat{\lambda} = \hat{\lambda}_1$ as a preliminary estimator of λ , and suppose that after j steps we have calculated an updated estimator $\hat{\lambda}_j$. Compute values $\hat{\lambda}^*$ of $\hat{\lambda}$ based on independent Monte Carlo trials of the point process under the model $\mathcal{M}(\hat{\lambda}_j)$, conditional on the data. (The edge-correction step suggested in Section 3.1 may be built into the definition of $\hat{\lambda}$. If it is, then the present Monte Carlo step effectively involves the double bootstrap.) Take $\tilde{\lambda}_j$ to be the average of these B values, and let $\hat{\lambda}_{j+1} = \hat{\lambda}_1, \hat{\lambda}_j/\tilde{\lambda}_j$ be the bias-corrected estimator of λ at the $(j + 1)$ th step.

In practice, the model would generally be chosen using prior information, acquired for example from a previous data set in a related setting obtained by direct observation rather than wandering quarter sampling. From those data, model parameters other than the intensity could be estimated. There is unlikely to be adequate information in data gathered using the wandering quarter method for estimating parameters other than intensity. For the same reason, while techniques based on the block bootstrap for spatial data (see Hall 1985; Sherman 1996; García Soidán and Hall 1997) are theoretically possible in this setting, they would generally not be feasible. However, one could treat the sequence of values X_i as a time series and either apply block bootstrap methods in that context (Carlstein 1986; Künsch 1989; Politis *et al.* 1999) or use a more structural bootstrap method (Bose 1988).

4. Theoretical properties in the case $w < \infty$

First we address properties of N . Defining $V_i = X_i \cos \Theta_i$, we see that N equals the largest integer j such that $\sum_{i \leq j} V_i \leq w$. Therefore, in a renewal process where consecutive service lifetimes are the independent random variables V_1, V_2, \dots , N equals the number of items

that are replaced during the time interval $[0, w]$, and $D_w \equiv w - \sum_{i \leq N} X_i \cos \Theta_i$, appearing in formula (2.2) for ℓ , equals the ‘current life’ of the item in service at time w . Standard methods for analysing renewal processes (Cox 1962) allow us to prove that the limiting distribution of D_w , as $w \rightarrow \infty$, has density $P(V > x)/E(V)$, where V has the distribution of a generic V_i . Moments of N may be computed using standard arguments. For example, the expected value of N , as a function of w , equals the renewal function of the process. Moreover, $E(N) = \nu + O(1)$ and $\text{var}(N) = (\sigma/\mu)^2\nu + O(1)$ as $w \rightarrow \infty$, where $\mu = E(V) = \frac{1}{2}(\pi/\lambda\theta^3)^{1/2} \sin \theta$, $\sigma^2 = \text{var}(V)$ and $\nu = w/\mu$.

Using standard arguments based on an invariance principle for a sum of independent random variables, it may also be shown that the maximum likelihood estimator $\hat{\lambda}_{ML}$, defined by minimizing ℓ in (2.2), and also $\hat{\lambda}_{\text{simp}}$ and $\hat{\lambda}_{\text{adj}}$, are asymptotically normally distributed with mean λ and variance λ^2/ν . Moreover, if var' denotes variance conditional on $N \geq 2$, then $\text{var}'(\hat{\lambda}_{ML})$, $\text{var}'(\hat{\lambda}_{\text{simp}})$ and $\text{var}'(\hat{\lambda}_{\text{adj}})$ all equal $\lambda^2\nu^{-1} + O_p(w^{-2})$, as $w \rightarrow \infty$. These properties may be used as the basis of a proof that $\hat{\lambda}_{ML}$, $\hat{\lambda}_{\text{simp}}$ and $\hat{\lambda}_{\text{adj}}$ all achieve the asymptotic minimum variance bound for estimators of λ based on χ . Theorem 4.1 below gives formulae for the mean.

Let λ denote the true value of the parameter, recall the definition earlier of the random variable D_w , and put

$$W(\omega) = (\lambda\theta)^{-1}(\cos \omega)^{-2}D_w^2,$$

$$T_j = \int_0^\theta W(\omega)^j \exp\{-W(\omega)\}d\omega, \quad \tau = E(T_1/T_0).$$

Alternatively, we may replace D_w by its weak limit as $w \rightarrow \infty$, without affecting the results below. Let E' denote expectation conditional on $N \geq 2$.

Theorem 4.1. *If $\theta \in (0, \pi/2)$ is fixed and $w < \infty$ then, as $w \rightarrow \infty$,*

$$E'(\hat{\lambda}_{\text{simp}}) = \lambda + \frac{3}{2}\mu\lambda w^{-1} + O(w^{-2}), \tag{4.1}$$

$$E'(\hat{\lambda}_{\text{adj}}) = \lambda + O(w^{-2}), \tag{4.2}$$

$$E'(\hat{\lambda}_{ML}) = \lambda + \mu\lambda(\frac{3}{2} - \tau)w^{-1} + O(w^{-2}). \tag{4.3}$$

Both (4.1) and (4.3) continue to hold if E' is taken to denote expectation conditional on $N \geq 1$. Formula (4.1) shows that $\hat{\lambda}_{\text{simp}}$ tends to overestimate λ . In relative terms the bias is too large by the factor

$$1 + \frac{3}{2}\mu w^{-1} + O(w^{-2}) = \frac{E(N)}{E(N) - \frac{3}{2}} + O(w^{-2}).$$

This suggests a simple remedy: replace the divisor N by $N - \frac{3}{2}$ in the definition of $\hat{\lambda}_{\text{simp}}$ in (2.4), giving the estimator $\hat{\lambda}_{\text{adj}}$ of (2.5). Result (4.2) confirms that this does in fact remove the majority of the bias. Also, since $\tau > 0$, the maximum likelihood estimator $\hat{\lambda}_{ML}$ also alleviates

some of the bias of $\hat{\lambda}_{\text{simp}}$, although the fact that τ is not in general equal to $\frac{3}{2}$ means that bias is not reduced by an order of magnitude.

Finally, we show that the bootstrap approximation suggested in Section 2.3 is second-order accurate. We use the notation there, and so $\hat{\Lambda}(w)$ denotes any one of the estimators $\hat{\lambda}_{\text{ML}}$, $\hat{\lambda}_{\text{simp}}$ and $\hat{\lambda}_{\text{adj}}$; $\hat{\Lambda}_0(\cdot)$ is a stochastic process independent of the data, with the distribution of $\hat{\Lambda}(\cdot)$ under the assumption that $\lambda = 1$; and $\hat{v} = \hat{\Lambda}(w)^{1/2}w$. For the results below it is not essential that the version of $\hat{\Lambda}$ used to define \hat{v} be the same as the one for which we wish to compute the distributional approximation. In practice it usually would be, however. Let $\hat{a}(\alpha)$, $\hat{b}(\alpha)$ denote the solutions of

$$P[\hat{\Lambda}(w)^{1/4}\{\hat{\Lambda}_0(\hat{v})^{1/4} - 1\} \leq \hat{a}(\alpha)|\chi] = \alpha,$$

$$P[\hat{\Lambda}(w)^{1/4}\{\hat{\Lambda}_0(\hat{v}) - 1\} \leq \hat{b}(\alpha)|\chi] = \alpha,$$

respectively, and write \mathcal{A} for the class of all measurable subsets of the real line.

Theorem 4.2. *If $\theta \in (0, \pi/2)$ is fixed and $w < \infty$ then, as $w \rightarrow \infty$,*

$$\sup_{\mathcal{A} \in \mathcal{A}} |P[\hat{\Lambda}(w)^{1/4}\{\hat{\Lambda}_0(\hat{v})^{1/4} - 1\} \in \mathcal{A}|\chi] - P\{\hat{\Lambda}(w)^{1/4} - \lambda^{1/4} \in \mathcal{A}\}| = O_p(w^{-1}), \tag{4.4}$$

$$\sup_{\mathcal{A} \in \mathcal{A}} |P[\hat{\Lambda}(w)^{1/4}\{\hat{\Lambda}_0(\hat{v}) - 1\} \in \mathcal{A}|\chi] - P\{\hat{\Lambda}(w)\lambda^{-3/4} - \lambda^{1/4} \in \mathcal{A}\}| = O_p(w^{-1}), \tag{4.5}$$

$$\sup_{0 < \alpha < 1} |P\{\hat{\Lambda}(w)^{1/4} - \lambda^{1/4} \leq \hat{a}(\alpha)\} - \alpha| = O(w^{-1}), \tag{4.6}$$

$$\sup_{0 < \alpha < 1} |P\{\hat{\Lambda}(w)\lambda^{-3/4} - \lambda^{1/4} \leq \hat{b}(\alpha)\} - \alpha| = O(w^{-1}). \tag{4.7}$$

Let $\xi = \lambda^{3/4}(\pi/\theta^3)^{1/4}(\frac{1}{2}\sin\theta)^{1/2}$. The results noted in the first two paragraphs of this section imply that $w^{1/2}\{\hat{\Lambda}(w) - \lambda\}/\xi$ is asymptotically standard normal as $w \rightarrow \infty$. Therefore, the normal approximation to the distribution of $\hat{\Lambda}(w)$ is given by

$$P\{\hat{\Lambda}(w) \in A\} \approx P[N(0, 1) \in \{w^{1/2}(a - \lambda)/\xi : a \in A\}].$$

It may be proved as in Section A.5 that the error in this approximation is of size $w^{-1/2}$. The fact that terms in $w^{-1/2}$ make a non-vanishing contribution to the approximation is a consequence of skewness of the distribution of X . On the other hand, we see from (4.4) and (4.5) that the bootstrap method is accurate up to terms of order w^{-1} . That is, the method has implicitly corrected for the main effect of skewness, and so is second-order accurate. Results (4.6) and (4.7) show that the second-order accuracy property extends to coverage accuracy.

5. Empirical results

To explore the finite-sample performance of our methods we applied them first to simulated data which followed either a homogeneous Poisson process of intensity 2 or a cluster process, and then to real data which appeared to follow a homogeneous Poisson process.

We first address the performance of the estimators and the bootstrap confidence intervals proposed in Section 2.3, using four different cone angles. In this example we sampled in the centre of a large region, and so no edge effects occurred. In Table 1 we report the means, standard deviations and coverages of nominal 95% two-sided confidence intervals based on the two possible approximate pivotal quantities, T_1 and T_2 , and the coverages of nominal 97.5% upper and lower confidence intervals, labelled $U(T_j)$ and $L(T_j)$, respectively. True intensity is $\lambda = 2$. Each entry in the table is based on 1000 realizations. The mean ν and standard deviation τ of the numbers of observations N on which the estimates were based are approximately $(\nu, \tau) = (9.88, 1.70), (15.40, 2.12), (21.32, 2.63)$ and $(27.51, 3.34)$ in the cases $\theta = \pi/8, \pi/4, 3\pi/8$ and $\pi/2$, respectively.

The use of a cone semi-angle of $\theta = \pi/2$ turns out to be problematic for at least two reasons. First, the last two distance measurements are often outliers and make the intensity overly small. Second, if there are two objects at the same distance, up to rounding errors, then the number of steps can become infinite as the algorithm jumps from one object to the next and back again. To overcome the latter problem we used the following ‘censoring’ approach. At any step, if there were no objects within a $\pi/2$ quadrant then the end of the region was deemed to have been reached. This *ad hoc* modification affects only the last one or two steps.

When the cone angle is relatively small, the adjusted estimator has least bias and the simple estimator has greatest bias. However, when the cone angle is large, the adjusted estimator has a larger bias and, at $\pi/4$, the simple estimator has least bias. The maximum likelihood estimator is most efficient when the cone angle is small, but the adjusted estimator is more efficient when the cone angle is large. The efficiencies are virtually equal when $\theta = \pi/4$. Coverages are close to their nominal levels when the angle is small, but are below nominal levels for two-sided and for one-sided upper intervals, and above nominal levels for one-sided lower intervals, when the cone angle is large.

Table 1. Performance of point and interval estimators when no edge effects occur. Columns headed $T_i, L(T_i)$ and $U(T_i)$ give coverages of two-sided, lower one-sided and upper one-sided confidence intervals respectively, based on T_i , with respective coverages 0.95, 0.975 and 0.975. The true value of λ was 2

θ	Estimator	Mean	Std dev.	T_1	T_2	$L(T_1)$	$U(T_1)$	$L(T_2)$	$U(T_2)$
$\pi/8$	$\hat{\lambda}_{\text{simp}}$	2.20	0.686	94.7	93.6	97.5	97.2	97.0	96.6
	$\hat{\lambda}_{\text{adj}}$	1.87	0.637	96.6	95.1	98.1	98.6	97.7	97.4
	$\hat{\lambda}_{\text{ML}}$	2.08	0.640	93.0	92.0	96.0	97.0	95.5	96.5
$\pi/4$	$\hat{\lambda}_{\text{simp}}$	2.15	0.52	95.9	95.4	97.7	98.3	97.4	98.0
	$\hat{\lambda}_{\text{adj}}$	1.93	0.51	96.2	95.3	97.7	98.5	97.6	97.8
	$\hat{\lambda}_{\text{ML}}$	2.09	0.52	96.5	96.5	98.5	98.0	98.5	98.0
$3\pi/8$	$\hat{\lambda}_{\text{simp}}$	2.02	0.432	94.1	93.6	98.3	95.8	98.1	95.5
	$\hat{\lambda}_{\text{adj}}$	1.90	0.418	96.2	95.3	97.7	98.5	97.6	97.8
	$\hat{\lambda}_{\text{ML}}$	1.98	0.423	93.3	93.0	97.5	95.5	97.5	95.5
$\pi/2$	$\hat{\lambda}_{\text{simp}}$	1.84	0.345	89.3	88.9	99.4	89.9	99.4	89.6
	$\hat{\lambda}_{\text{adj}}$	1.74	0.335	96.2	95.3	97.7	98.5	97.6	97.8
	$\hat{\lambda}_{\text{ML}}$	1.81	0.346	86.5	100	86.5	86.5	100	86.5

To explore the effectiveness of the bootstrap edge correction suggested in Section 3.1, a Poisson process of intensity 2 was simulated in a region with dimensions 10×2 , the longer side being in the sampling direction. Additionally, the right-hand boundary was a straight line and the left-hand boundary was given by $y = 0.5 \sin x$. The results of making one and two iterations of the bootstrap bias correction applied to $\hat{\lambda}_{\text{simp}}$, for three different cone angles, are presented in Table 2. For the sake of comparison we also give there the means and standard deviations of $\hat{\lambda}_{\text{simp}}$. It can be seen from the table that bias and efficiency tend to increase as the cone angle increases, that the bootstrap edge correction reduces bias even when the cone angle is large, and that the effect of bias correction on efficiency is small. Only one or two iterations are required to effect most of the achievable bias correction. The mean ν and standard deviation τ of the numbers of observations N on which the estimates were based are approximately $(\nu, \tau) = (9.35, 1.76)$, $(14.46, 2.18)$ and $(19.29, 2.71)$ in the cases $\theta = \pi/8$, $\pi/4$ and $3\pi/8$, respectively.

Results are similar for the other estimators, $\hat{\lambda}_{\text{ML}}$ and $\hat{\lambda}_{\text{adj}}$. Of course, these estimators cannot be expected to have special virtues in such cases, and there is no a priori reason for preferring them to $\hat{\lambda}_{\text{simp}}$ when bootstrap bias correction is employed. We shall illustrate the application of bias correction to $\hat{\lambda}_{\text{adj}}$ later in this section.

We also explored properties of the estimator $\hat{\lambda}_{\text{simp}}$ applied to a parent–daughter cluster process (i.e. a Neyman–Scott process). The parents in each cluster were simulated from a Poisson process with mean $4/\delta$, and each parent was given a Poisson number of daughters, the number having mean $\delta - 1$. Thus, the mean cluster size was δ and the intensity of the cluster process was $\lambda = 4$. The positions of the daughters, in the coordinate systems of their respective parents, were bivariate normal with mean zero and variance σ^2 in each coordinate. The degree of clustering was thus controlled by σ . For the detailed results reported here we took $\sigma = 0.1$ and $\delta = 2$ (model 1) or $\delta = 4$ (model 2). The process was simulated 2000 times in a 10×5 region, the shorter side being in the sampling direction. In each case, sampling was started in the centre of the left-hand edge to reduce any boundary problems. It was found empirically that for model 1 the distribution of $Z_i = \lambda \theta X_i^2$ had mean 1.26 and variance 3.05, and that for model 2 the distribution of Z_i had mean 1.58 and variance 8.31. If the point process were Poisson then the mean and variance would both equal 1.

To emulate what might happen if one were to fit an approximate model, rather than the correct model, we considered the case where the fitted model was (a) homogeneous Poisson, with no attempt made to capture clustering, or (b) an underdispersed Neyman–Scott model,

Table 2. Influence of bootstrap correction for edge effects. Columns headed Mean (i) and Std dev. (i) give the mean and standard deviation respectively after i iterations and for the estimator $\hat{\lambda}_{\text{simp}}$. The notation $i = 0$ corresponds to the uncorrected estimator $\hat{\lambda}_{\text{simp}}$. The true value of λ was 2

θ	Mean (0)	Std dev. (0)	Mean (1)	Std dev. (1)	Mean (2)	Std dev. (2)
$\pi/8$	1.95	0.81	2.04	0.85	2.05	0.86
$\pi/4$	1.90	0.56	2.08	0.58	2.06	0.58
$3\pi/8$	1.68	0.47	2.11	0.52	2.05	0.52

with correct mean cluster size δ but clustering parameter $\sigma/2$, or (c) the Neyman–Scott model using the correct mean cluster size δ and correct clustering parameter σ , or (d) an overdispersed Neyman–Scott model using the correct mean cluster size δ but clustering parameter 2σ , or (e) the Neyman–Scott model with mean cluster size $\delta/2$ but correct clustering parameter σ , or (f) the Neyman–Scott model with mean cluster size 2δ but correct clustering parameter σ . In each case we employed techniques suggested in Section 3.2 to estimate λ , but no other aspects of the model were estimated.

The results in Table 3 show that the bootstrap adjustment of the simple point estimate of intensity works extremely well when the fitting process is correct but, as one might expect, adopting an underdispersed process leads to overcorrection, and adopting an overdispersed process leads to undercorrection. The coverage of the resulting confidence intervals is generally lower than the nominal value. The one-sided intervals have coverages much closer to their nominal values than the two-sided intervals. In both one- and two-sided settings, the intervals based on T_1 generally have higher coverage than those based on T_2 and so have better coverage accuracy. The most accurate coverage for two-sided intervals, and upper intervals, is obtained by adopting an overdispersed Neyman–Scott model, whereas the best coverage for lower intervals is in most cases obtained by adopting an underdispersed Neyman–Scott model.

Finally, we applied our methodology to a real, mapped data set, which appears to follow a homogeneous Poisson process. The data represent locations of plants from a Western Australian species, and are obtainable from the authors. They were drawn from two separate

Table 3. Performance of simulation-based bias-corrected, simple estimator $\hat{\lambda}_{bc}$ for clustered point processes. We used a cone semi-angle of $\theta = \pi/4$, and employed the bootstrap to adjust for clustering. The data were generated from Neyman–Scott processes with clustering parameter $\sigma = 0.1$ and intensity $\lambda = 4$. The first column gives the model number. For models 1 and 2 the mean cluster size was $\delta = 2$ and $\delta = 4$, respectively. The fitted process used for the bootstrap was either a homogeneous Poisson process (indicated by ‘Poisson’ in the second column) or a Neyman–Scott process with the parameter values given in the second column. The third and fourth columns give average values (i.e. empirical approximations to expected values) of $\hat{\lambda}_{simp}$ and $\hat{\lambda}_{adj}$, respectively. The fifth column gives the standard deviation of $\hat{\lambda}_{adj}$. Notation for all other columns is as for Table 1

No.	Fitted process	$\hat{\lambda}_{simp}$	$\hat{\lambda}_{bc}$	$SD(\hat{\lambda}_{bc})$	T_1	T_2	$L(T_1)$	$U(T_1)$	$L(T_2)$	$U(T_2)$
1	Poisson	3.78	3.45	1.51	84.2	80.5	93.8	90.4	90.2	90.3
	$(\delta, \sigma/2)$	3.78	4.24	1.85	83.3	81.7	96.9	86.3	95.5	86.2
	(δ, σ)	3.78	4.01	1.72	88.8	86.0	95.9	92.9	93.6	92.5
	$(\delta, 2\sigma)$	3.78	3.75	1.61	89.2	86.8	92.7	96.4	90.8	96.0
	$(\delta/2, \sigma)$	3.78	3.78	1.64	84.0	81.9	89.1	94.9	87.1	94.8
	$(2\delta, \sigma)$	3.78	4.37	1.87	85.3	83.3	98.7	86.3	97.2	86.2
2	Poisson	3.26	2.92	1.54	77.4	72.2	84.4	92.9	79.3	92.9
	$(\delta, \sigma/2)$	3.26	4.30	2.34	78.8	76.9	97.8	80.9	96.1	80.8
	(δ, σ)	3.26	3.91	2.06	86.9	84.4	97.0	90.0	94.9	89.5
	$(\delta, 2\sigma)$	3.26	3.45	1.79	91.5	87.6	94.9	96.4	91.2	96.4
	$(\delta/2, \sigma)$	3.26	3.39	1.76	84.8	81.3	89.4	95.3	86.0	95.3
	$(2\delta, \sigma)$	3.26	4.53	2.55	77.9	75.8	97.7	79.3	95.5	80.3

geographical regions. Although our methodology is intended for field surveys rather than mapped data, the use of mapped data allows us to draw repeated samples and hence further evaluate our methodology.

For both regions we applied the estimator $\hat{\lambda}_{\text{adj}}$, which performed well and was not computationally burdensome. Performance is summarized in Table 4. Coverage accuracies of intervals based on T_1 are better than for intervals based on T_2 , for both regions. Values of the mean ν and standard deviation τ of the numbers of observations N are approximately $(\nu, \tau) = (5.54, 1.82)$ and $(5.79, 1.93)$ for regions A and B, respectively.

Appendix: Outline proofs

A.1. Notes on the proof of Theorem 4.1

Our arguments include Taylor expansions of smooth functions of random variables, where we assume that the expected value of the function, conditional on $N \geq 1$, may be well approximated by a term-by-term expectation of the Taylor expansion – see, for example, (A.1). We may justify such steps by considering separately the cases where the argument in the expansion is small, and where it is large. For example, in (A.1) we may define $Q = N^{-1} \sum_{i \leq N} (Z_i - 1)$, and for $0 < \epsilon < \frac{1}{2}$ consider the cases $|Q| \leq w^{-\epsilon}$ and $|Q| > w^{-\epsilon}$. The latter event has probability $O(w^{-C})$ for all $C > 0$, and a Taylor expansion is valid when the former event holds. By choosing ϵ appropriately, and going to a larger number of terms than actually shown in the expansion, we may justify formulae such as (A.1). In some of our arguments we shall do no more than identify values of coefficients in expansions, rather than compute the concise orders of terms. The expansions are invariably in powers of w^{-1} , as may be verified using the aforesaid methods.

A.2. Outline derivation of (4.1)

By Taylor expansion,

$$\lambda^{-1} \hat{\lambda}_{\text{simp}} = 1 - N^{-1} \sum_{i=1}^N (Z_i - 1) + \left\{ N^{-1} \sum_{i=1}^N (Z_i - 1) \right\}^2 + \dots,$$

Table 4. Performance of the bootstrap-corrected, adjusted estimator $\hat{\lambda}_{\text{adj}}$ for plant data. We used a cone semi-angle of $\theta = \pi/4$, and employed the bootstrap correction for edge effects. Notation for columns is as in the case of Table 1. Intensities were $\lambda = 0.0193$ for region A and $\lambda = 0.0232$ for region B

Region	Mean	Std dev.	T_1	T_2	$L(T_1)$	$U(T_1)$	$L(T_2)$	$U(T_2)$
A	0.0210	0.0277	85.4	84.4	89.0	96.4	88.7	95.8
B	0.0239	0.0130	81.7	81.2	93.1	88.6	92.7	88.6

whence it may be shown that

$$E'(\lambda^{-1}\hat{\lambda}_{\text{simp}}) = 1 - E' \left\{ N^{-1} \sum_{i=1}^N (Z_i - 1) \right\} + \nu^{-1} + O(\nu^{-2}). \tag{A.1}$$

A heuristic argument leading to the ν^{-1} term on the right-hand side of (A.1) is that $E' \{ N^{-1} \sum_{i \leq N} (Z_i - 1) \}^2$ is approximately equal to its value when N is replaced by ν , that is, to ν^{-1} (assuming temporarily that ν is an integer), plus terms of order ν^{-2} . This may be made rigorous using the methods leading to (A.3). Also, defining $\mu = E(V)$, $\sigma^2 = \text{var}(V)$, $w_i = (w - i\mu)/(i\sigma^2)^{1/2}$,

$$S_i = (i\sigma^2)^{-1/2} \sum_{j=1}^i (X_j \cos \Theta_j - \mu), \quad \Delta_i = \frac{X_i \cos \Theta_i - \mu}{\{(i-1)\sigma^2\}^{1/2}},$$

and F_i and f_i to be the distribution and density functions respectively of S_i , we have

$$E' \left\{ N^{-1} \sum_{i=1}^N (Z_i - 1) \right\} = \nu^{-1} E' \left\{ \sum_{i=1}^N (Z_i - 1) \right\} - \nu^{-2} E' \left\{ (N - \nu) \sum_{i=1}^N (Z_i - 1) \right\} + O(\nu^{-3}), \tag{A.2}$$

$$\begin{aligned} E' \left\{ \sum_{i=2}^N (Z_i - 1) \right\} &= \sum_{i=2}^{\infty} E \left\{ (t^{-1} X_i^2 - 1) I \left(\sum_{j=1}^i X_j \cos \Theta_j \leq w \right) \right\} \\ &= \sum_{i=2}^{\infty} E \{ (t^{-1} X_i^2 - 1) P(S_{i-1} \leq w_i - \Delta_i | \Delta_i) \} \\ &= \sum_{i=2}^{\infty} E \{ (t^{-1} X_i^2 - 1) \{ F_{i-1}(w_i) - \Delta_i f_{i-1}(w_i) + \frac{1}{2} \Delta_i^2 f'_{i-1}(w_i) + \dots \} \} \\ &= - \sum_{i=2}^{\infty} E \{ (t^{-1} X_i^2 - 1) \Delta_i \} f_{i-1}(w_i) + O(\nu^{-1}) \end{aligned} \tag{A.3}$$

as $w \rightarrow \infty$. Here we define $\sum_{2 \leq i \leq N}$ to equal zero if $N = 1$. Using a non-uniform approximation to f_i in the central limit theorem for S_i , and writing ϕ for the standard normal density, we may prove that

$$\sum_{i=2}^{\infty} \{ (i-1)\sigma^2 \}^{-1/2} f_{i-1}(w_i) \sim \int_0^{\infty} (x\sigma^2)^{-1/2} \phi \left(\frac{w - x\mu}{(x\sigma^2)^{1/2}} \right) dx \sim \mu^{-1}.$$

A longer argument will show that the left-hand side equals $\mu^{-1} + O(\nu^{-1})$. Moreover, $\mu_1 \equiv E \{ (t^{-1} X^2 - 1)(X \cos \Theta - \mu) \} = \frac{1}{2} E(V)$. It may also be proved, using the argument in (A.3), that

$$E' \left\{ (N - \nu) \sum_{i=1}^N (Z_i - 1) \right\} = O(1)$$

as $w \rightarrow \infty$. Heuristically, this is true because $E' \{ \sum_{i \leq N} (Z_i - 1) | N \} = \text{const.} + O_p(N^{-1} + \nu^{-1})$ as $w \rightarrow \infty$. Combining the results from (A.2) down, we deduce that

$$E' \left\{ N^{-1} \sum_{i=1}^N (Z_i - 1) \right\} = -(\mu\nu)^{-1} \mu_1 + O(w^{-2}) = -w^{-1} \mu_1 + O(w^{-2}).$$

From this formula and (A.1) we see that

$$E'(\lambda^{-1} \hat{\lambda}_{\text{simp}}) = 1 + (\mu + \mu_1)w^{-1} + O(w^{-2}) = 1 + \frac{3}{2}\mu w^{-1} + O(w^{-2}), \quad (\text{A.4})$$

which proves (4.1).

A.3. Outline derivation of (4.2)

The main departure from the proof of (4.1) is that (A.1) should be replaced by

$$E'(\lambda^{-1} \hat{\lambda}_{\text{adj}}) = E'(1 - \frac{3}{2}N^{-1}) - E' \left\{ N^{-1} \sum_{i=1}^N (Z_i - 1) \right\} + \nu^{-1} + O(\nu^{-2}).$$

The first term on the right-hand side equals $1 - \frac{3}{2}\nu^{-1} + O(\nu^{-2})$ or, equivalently, $1 - \frac{3}{2}\mu w^{-1} + O(w^{-2})$, whence we obtain instead of (A.4) the result

$$E'(\lambda^{-1} \hat{\lambda}_{\text{adj}}) = 1 + (\mu + \mu_1 - \frac{3}{2}\mu)w^{-1} + O(w^{-2}) = 1 + O(w^{-2}),$$

which is equivalent to (4.2).

A.4. Outline derivation of (4.3)

Define the random variable Δ by $\hat{\lambda} = \hat{\lambda}_{\text{simp}}(1 + \Delta)$, where $\hat{\lambda}_{\text{ML}}$ minimizes $\ell(\lambda)$, given in (2.2). Put

$$\hat{W}(\omega) = \hat{\lambda}_{\text{simp}} \theta(\cos \omega)^{-2} \left(w - \sum_{i=1}^N X_i \cos \Theta_i \right)^2,$$

$$I_j = \int_0^\theta \hat{W}(\omega)^j \exp\{-\hat{W}(\omega)\} d\omega$$

and $J_j = I_j/I_0$. Then,

$$\int_0^\theta \exp\{-(1 + \Delta)\hat{W}(\omega)\} d\omega = I_0 \sum_{i=0}^{\infty} \frac{(-1)^i}{i!} \Delta^i J_i,$$

whence it follows that

$$\ell(\hat{\lambda}_{ML}) = N \sum_{i=2}^{\infty} \frac{(-1)^i}{i} \Delta^i - \log \left(\sum_{i=0}^{\infty} \frac{(-1)^i}{i!} \Delta^i J_i \right) + T,$$

where T denotes a random variable not depending on Δ . Differentiating with respect to Δ , equating to zero and solving for Δ , we see that $\Delta = -J_1 N^{-1} + O_p(w^{-2})$. It may now be shown that

$$E'(\hat{\lambda}_{ML}) = E'(\hat{\lambda}_{simp}) + \lambda E'(\Delta) + O(w^{-2}) = E'(\hat{\lambda}_{simp}) - \lambda E'(J_1) \nu^{-1} + O(w^{-2}). \tag{A.5}$$

Finally, $E'(J_1) = \tau + O(w^{-1})$, and so (4.3) follows from (4.1) and (A.5).

A.5. Outline derivation of (4.5)

Recall that $Z_i = \lambda \theta X_i^2$, and let (Z, Θ) denote a generic (Z_i, Θ_i) . Since the random variables Z_i and Θ_i are independent and have absolutely continuous distributions with all moments finite, we may develop a local, non-uniform Edgeworth expansion of the distribution of

$$S^T = (S_1, S_2) = \left(\sum_{i=1}^n Z_i, \sum_{i=1}^n Z_i^{1/2} \cos \Theta_i \right).$$

Indeed, writing M for the covariance matrix of $(Z, Z \cos \Theta)^T$, defining

$$T = n^{-1/2} M^{-1/2} \{S - E(S)\},$$

and letting f be the density of T , we may show that the Edgeworth expansion has the property

$$\sup_{t \in \mathbb{R}^2} (1 + \|t\|^k) \left| f(t) - \phi_2(t) - \sum_{r=1}^m n^{-r/2} \pi_r(t) \phi_2(t) \right| = O(n^{-(m+1)/2})$$

as $n \rightarrow \infty$, for any $k > 0$ and any integer $m \geq 1$. Here, ϕ_2 is the bivariate standard normal density, and π_r is a polynomial of degree $3r$ with parity identical to that of r – see, for example, Bhattacharya and Rao (1976, Section 19).

Assume temporarily that $\lambda = 1$. Using methods introduced in Section A.2, we may express the joint distribution of N , $\sum_{i \leq N} X_i^2$ and $\sum_{i \leq N} X_i \cos \Theta_i$ relatively simply in terms of the joint distribution of S_1 , S_2 and $S_2 + Z_{n+1}^{1/2} \cos \Theta_{n+1}$, and thereby prove that for the density g of $v^{1/2} \{\hat{\Lambda}_0(v) - 1\}$,

$$g(t) = \sum_{r=0}^m v^{-r/2} g_r(t) + R_{m2}(t), \tag{A.6}$$

where the functions g_r do not depend on v , g_0 is the density of a normal distribution, and the remainder R_{m2} satisfies, for $j = 2$,

$$\sup_{-\infty < t < \infty} (1 + |t|^k) |R_{mj}(t)| = O(v^{-(m+1)/2}) \quad (\text{A.7})$$

as $v \rightarrow \infty$.

Now we allow λ to be any fixed positive number. Since the distribution of $\hat{\Lambda}(w)/\lambda$ is identical to that of $\hat{\Lambda}_0(\lambda^{1/2}w)$, then, taking $v = \lambda^{1/2}w$, we have immediately from (A.6) that the density $g_{[\lambda]}$ of $(\lambda^{1/2}w)^{1/2}\{\hat{\Lambda}(w)\lambda^{-1} - 1\}$ satisfies

$$g_{[\lambda]}(t) = \sum_{r=0}^m \lambda^{-r/4} w^{-r/2} g_r(t) + R_{m3}(t), \quad (\text{A.8})$$

where the functions g_r are as in (A.6) and the remainder term satisfies (A.7) with $j = 3$. The same argument shows that if $\hat{g}_{[\lambda]}$ is the conditional density of $\hat{v}^{1/2}\{\hat{\Lambda}_0(\hat{v}) - 1\}$, given χ , where $\hat{v} = \hat{\Lambda}(w)^{1/2}w$, then

$$\hat{g}_{[\lambda]}(t) = \sum_{r=0}^m \hat{\Lambda}(w)^{-r/4} w^{-r/2} g_r(t) + R_{m4}(t), \quad (\text{A.9})$$

for the same functions g_r and a stochastic remainder $R_{m4}(t)$ that satisfies (A.7), for $j = 4$, with O replaced by O_p on the right-hand side. Taking $m = 1$, subtracting (A.8) and (A.9), and noting that $\hat{\Lambda}(w) - \lambda = O_p(w^{-1/2})$ as $w \rightarrow \infty$, which is a consequence of the mean and variance properties noted in Section 4, we see that

$$\sup_{-\infty < t < \infty} (1 + |t|^k) |\hat{g}_{[\lambda]}(t) - g_{[\lambda]}(t)| = O_p(w^{-1}),$$

from which (4.5) follows.

The derivation of (4.4) is similar. Results (4.6) and (4.7) may be established by modifying methods for proving such formulae in more conventional cases (see Hall 1992, Chapter 3) along the lines of the argument for (4.5).

References

- Bhattacharya, R.N. and Rao, R.R. (1976) *Normal Approximation and Asymptotic Expansions*. New York: Wiley.
- Bose, A. (1988) Edgeworth correction by bootstrap in autoregressions. *Ann. Statist.*, **16**, 1709–1722.
- Carlstein, E. (1986) The use of subsample values for estimating the variance of a general statistic from a stationary time series. *Ann. Statist.*, **14**, 1171–1179.
- Catana, A.J. (1963) The wandering quarter method of estimating population density. *Ecology*, **44**, 349–360.
- Cliff, A.D. and Ord, J.K. (1981) *Spatial Point Processes: Models and Applications*. London: Pion.
- Cox, D.R. (1962) *Renewal Theory*. London: Methuen.
- Cressie, N. (1991) *Statistics for Spatial Data*, 2nd edition. New York: Wiley.
- Diggle, P.J. (1983) *Statistical Analysis of Spatial Point Patterns*. London: Academic Press.
- García Soidán, P.H. and Hall, P. (1997) On sample reuse methods for spatial data. *Biometrics*, **53**, 273–281.
- Hall, P. (1985) Resampling a coverage pattern. *Stochastic Process. Appl.*, **20**, 231–246.
- Hall, P. (1992) *The Bootstrap and Edgeworth Expansion*. New York: Springer-Verlag.

- Künsch, H.R. (1989) The jackknife and the bootstrap for general stationary observations. *Ann. Statist.*, **17**, 1217–1241.
- O'Hara Hines, R.J. and Hines, W.G.S. (1989) Repeated sampling of spatial point distributions. *Comm. Statist. Theory Methods*, **18**, 2599–2614.
- Politis, D.N., Romano, J.P. and Wolf, M. (1999) *Subsampling*. New York: Springer-Verlag.
- Ripley, B.D. (1981) *Spatial Statistics*. New York: Wiley.
- Ripley, B.D. (1988) *Statistical Inference for Spatial Point Processes*. Cambridge: Cambridge University Press.
- Sherman, M. (1996) Variance estimation for statistics computed from spatial lattice data. *J. Roy. Statist. Soc. Ser. B*, **58**, 509–523.
- Upton, G.J.G. and Fingleton, B. (1985) *Spatial Data Analysis by Example. Volume 1: Point Patterns and Quantitative Data*. New York: Wiley.

Received April 2000 and revised July 2001