Estimating functions for inhomogeneous spatial point processes with incomplete covariate data

Rasmus Waagepetersen
Department of Mathematical Sciences, Aalborg University
Fredrik Bajersvej 7G, DK-9220 Aalborg
rw@math.aau.dk

Summary

The R package `spatstat` provides a very flexible and useful framework for analyzing spatial point patterns. A fundamental feature is a procedure for fitting spatial point process models depending on covariates. However, in practice one often faces incomplete observation of the covariates and this leads to parameter estimation error which is difficult to quantify. In this paper we introduce a Monte Carlo version of the estimating function used in `spatstat` for fitting inhomogeneous Poisson processes and certain inhomogeneous cluster processes. For this modified estimating function it is feasible to obtain the asymptotic distribution of the parameter estimates in the case of incomplete covariate information. This allows a study of the loss of efficiency due to the missing covariate data.

Some key words: Asymptotic normality; Cluster process; Estimating function; Experimental design; Inhomogeneous point process; Missing covariate data; Poisson process.

1 Introduction

The basic model for the relation between a spatial point process and spatial covariates is an inhomogeneous Poisson process with intensity function $\lambda(\cdot;\beta)$ depending on the spatial covariates and an unknown parameter β. In this paper we focus on the log-linear model $\lambda(u;\beta) = \exp(z(u)\beta^{\mathsf{T}})$ where $z(u)$ is the vector of non-random spatial covariates at the location u. A Poisson process with log-linear intensity function was coined a modulated Poisson process in Cox (1972). Suppose the inhomogeneous Poisson point process X is observed within a bounded observation window W. The score function for β is then

$$u(\beta) = \sum_{u \in X} z(u) - \int_W z(u)\lambda(u;\beta)\mathrm{d}u \tag{1}$$

where the first term is a sum over the points in X (recall that X is a finite random subset of W). A spatial Poisson process is often not appropriate due to clustering of points not explained by the covariates or repulsion between the points. However, for non-Poisson processes with intensity function $\lambda(\cdot;\beta)$, (1) may still be useful for estimating β, see Schoenberg (2005), Waagepetersen (2007), Møller & Waagepetersen (2006), and Guan & Loh (2007).

In practice $z(\cdot)$ is often only observed at a finite set of locations so that the integral in (1) cannot be evaluated exactly. Rathbun (1996) proposes to substitute the missing covariate values by kriging predictions assuming that $\{z(u)|u \in W\}$ is a realization of a stochastic process. One disadvantage of this approach is the need to specify a model for the covariate process (typically involving new parameters to be estimated). A simpler approach is to approximate the score function by an estimating function

$$u_q(\beta) = \sum_{u \in X} z(u) - \sum_{u \in Q} z(u)\lambda(u;\beta)w(u) \tag{2}$$

obtained using numerical quadrature with quadrature points $u \in Q \subset W$ and associated weights $w(u)$ assuming that $z(u)$ is observed for u both in X and Q. Berman & Turner (1992) suggested quadrature schemes where for computational reasons explained in Section 2, $X \subset Q$. The Berman-Turner scheme and an extension of it to Markov point processes (Baddeley & Turner, 2000) is implemented in the `R` package `spatstat` (Baddeley & Turner, 2005).

It is in general not clear whether an approximate maximum likelihood estimate (MLE) obtained from (2) is consistent and asymptotically normal and how the variance matrix of the estimate differs from that of the MLE. Rathbun et al. (2007) suggest to use instead an estimating function of the form

$$u_r(\beta) = \sum_{u \in X} z(u) - \sum_{u \in D} \frac{z(u)\lambda(u;\beta)}{\rho(u)} \tag{3}$$

where D is a point process on W with intensity function $\rho(\cdot)$. Rathbun et al. (2007) demonstrate asymptotic normality of the estimate obtained using (3) under suitable conditions satisfied e.g. if D is a simple random sample (i.e. a binomial point process) independent of X.

The package `spatstat` is by far the most versatile and popular software for fitting spatial point process models. In this paper we build on the ideas in Rathbun et al. (2007) and consider in Section 2 Monte Carlo versions of the type of estimating equation used in `spatstat`. Asymptotic normality of the associated parameter estimates is discussed in Sections 3 and 4 for respectively inhomogeneous Poisson processes and inhomogeneous cluster processes. Practical examples are considered in Section 5, and Section 6 contains some closing remarks.

2 The Berman-Turner scheme with random dummy points

Consider the estimating function (2). Following Berman & Turner (1992), suppose that the set of quadrature points Q is the union of the point process X and a set of additional dummy points D chosen by the user. Then (2) can be rewritten as

$$u_q(\beta) = \sum_{u \in X \cup D} w(u) z(u) \big(y_u - \lambda(u; \beta) \big)$$

where the sum is over *both* the events in the point process X and the dummy points in D, and $y_u = 1[u \in X]/w(u)$ is the indicator that a quadrature point u is an event in X divided by the associated quadrature weight $w(u)$. Hence $u_q(\beta)$ is formally equivalent to the score of a weighted Poisson regression with weights $w(u)$ and 'observations' y_u and the equation $u_q(\beta) = 0$ can thus easily be solved using standard software for generalized linear models. Discussions and simulation studies of various dummy point and weight schemes can be found in Berman & Turner (1992), Lawson (1992), Baddeley & Turner (2000), and Wang & Lawson (2006).

The Berman-Turner scheme forms the basis of the estimation procedure `ppm` in the `R` package `spatstat` (Baddeley & Turner, 2005). This procedure allows any choice of dummy points D and quadrature weights $w(u)$. However, the most frequently used options for the weights are either `grid` or `dirichlet`. For the `grid` option, the observation window is divided into rectangular tiles. The quadrature weight for a quadrature point $u \in Q$ falling in a tile T is the area of T divided by the total number of quadrature points falling in T (hence adjusting for the possible multiple occurrences of a tile T in the quadrature sum). The advantage of this scheme is the easy calculation of the quadrature weights. For the `dirichlet` option, the weights are the areas of the tiles of the Dirichlet tesselation generated by the quadrature points Q.

In connection with grid weights, Baddeley & Turner (2000) discuss the possibility of using random dummy points where each tile contains precisely one dummy point picked at random in the tile. They, however, do not provide a detailed study of the resulting estimating function specified by (4) in the next section.

2.1 Random dummy points

In this section we define two Berman-Turner type estimating functions with random dummy points D of constant intensity ρ. Two types of dummy point processes are discussed in detail in Section 2.2. The weights are either grid weights as in `spatstat` or inverse proportional to the intensity of the quadrature point process $Q = X \cup D$.

Following Baddeley & Turner (2000), Section 4.3, consider random stratified dummy points where the observation window is divided into rectangular tiles and one dummy point is placed at random in each tile. For a dummy point $v \in D$, let T_v denote the tile containing v, and for any point $u \in Q$ let $N_u = \#(X \cap T_v)$ denote the number of events in X falling in the tile T_v containing u. We then obtain the estimating function

$$u_g(\beta) = \sum_{u \in X} z(u) - \sum_{u \in X \cup D} z(u) \frac{\lambda(u;\beta)}{\rho N_u + \rho}. \tag{4}$$

Note that if the T_v are very small, then the events in $X \cap T_v$ essentially become replicates of the dummy point $v \in D$ and the last term in (4) is well approximated by $\sum_{u \in D} z(u)\lambda(u;\beta)/\rho$ as in (3).

An analogue of the `spatstat` estimating equation with `dirichlet` weights is obtained with

$$u_d(\beta) = \sum_{u \in X} z(u) - \sum_{u \in X \cup D} z(u) \frac{\lambda(u;\beta)}{\lambda(u;\beta) + \rho} \tag{5}$$

where $\lambda(u;\beta) + \rho$ is the intensity of $X \cup D$. The analogy is based on two considerations: first, as for the `dirichlet` option in `spatstat` all quadrature points in $X \cup D$ are treated on an equal footing. Second, for a stationary point process of intensity α, the expected area of the associated typical Dirichlet cell is $1/\alpha$. Hence the weight $(\lambda(u;\beta)+\rho)^{-1}$ may be viewed as an approximation of the area of a Dirichlet cell in a region of constant intensity $\lambda(u;\beta) + \rho$. Although intuitively appealing, (5) yields an asymptotically suboptimal estimating function for certain choices of dummy point distributions, see Section 3.1. Note that (5) does not completely fall within the Berman-Turner setup since the weights depend on the parameter β. However, as pointed out by a referee, u_d can be rewritten as

$$u_d(\beta) = \sum_{u \in X \cup D} z(u)\Big(\tilde{y}_u - \frac{\exp(-\log\rho + z(u)\beta^{\mathsf{T}})}{\exp(-\log\rho + z(u)\beta^{\mathsf{T}}) + 1}\Big).$$

This is formally equivalent to the score of a logistic regression with 'observations' $\tilde{y}_u = 1[u \in X]$ and offset $-\log\rho$. Thus, $u_d(\beta) = 0$ too can straightforwardly be solved using software for generalized linear models.

2.2 Dummy point distributions

To establish asymptotic results for the estimating functions (4) and (5) we need a dummy point sampling design that ensures a central limit theorem for the Monte Carlo integration error. In Section 3 we more specifically consider sequences of dummy point processes D_n of increasing intensity $\rho_n = n^k\rho$, $\rho > 0$, $0 < k \le 1$, and require for integrable functions $f : W \to \mathbb{R}^p$,

$$n^{1/2}\left[\sum_{u \in D_n} \frac{f(u)}{n^k\rho} - \int_W f(u)\mathrm{d}u\right] \to N(0, G_f/\rho^{1/k}) \tag{6}$$

where G_f is a positive definite matrix.

Suppose D_n is a simple random sample of $n\rho|W|$ independent uniform points on W (i.e. D_n is a binomial point process of intensity $n\rho$). Then (6) holds with $k = 1$ and

$$G_f = \int_W f(u)^{\mathsf{T}} f(u)\mathrm{d}u - \frac{1}{|W|}\int_W f(u)^{\mathsf{T}}\mathrm{d}u \int_W f(u)\mathrm{d}u.$$

This is a special case of the type of dummy point distributions considered in Rathbun et al. (2007). An immediate generalization is to use independent binomial processes within regions of a fixed subdivision of W.

If the components of $f = (f_1, \ldots, f_p)$ are continuously differentiable we may achieve $k = 1/2$ in (6) using a stratified sampling design where the stratification depends on the number of dummy points. Suppose to be specific that $W = [0,a] \times [0,b]$ is rectangular. Divide W in $M_n = n^{1/2}\rho|W| = m_{1,n}m_{2,n}$, $m_{2,n} = m_{1,n}b/a$, squares $s_{i,n}$, $i = 1, \ldots, M_n$, each of sidelength $a/m_{1,n}$. We then obtain stratified dummy points $D_n = \{u_{1,n}, \ldots, u_{M_n,n}\}$ where the points $u_{i,n}$ are independent with $u_{i,n}$ uniform on $s_{i,n}$. Generalizing results in Okamoto (1976) to the multivariate case, (6) holds with

$$G_f = \frac{1}{12}\int_W A_f(u)\mathrm{d}u$$

where

$$A_f(u_1, u_2) = \left[\frac{\partial f_i}{\partial u_1} \frac{\partial f_j}{\partial u_1} + \frac{\partial f_i}{\partial u_2} \frac{\partial f_j}{\partial u_2} \right]. \tag{7}$$

Note that when using (4) and stratified dummy points we naturally let $T_{u,n} = s_{i,n}$ if $u \in D_n$ is generated in $s_{i,n}$. Stratified dummy points can easily be generated with the `spatstat` procedure `stratrand()`.

3 Asymptotic distribution of parameter estimates in the case of a Poisson process

The asymptotic distribution of parameter estimates is in this paper obtained using infill asymptotics where both the intensities of X and D tend to infinity. This type of asymptotics is useful when we wish to investigate the effect of increasing the number of dummy points within a fixed observation window. More specifically we consider sequences of Poisson point processes X_n and dummy point processes D_n with intensity functions

$$\lambda_n(u; \beta^*) = n\lambda(u; \beta^*),\ \beta^* \in \mathbb{R}^p \quad \text{and} \quad \rho_n = n^k \rho \tag{8}$$

where $\rho > 0$, $0 < k \leq 1$, and X_n and D_n are independent for each n. Note that $k < 1$ corresponds to the case where the intensity ρ_n of D_n tends to infinity at a slower rate than the intensity of X_n. One may think of X_n as representing the accumulation of points up to a certain 'time' point n and the intensity of X_n is then proportional to n.

Considering first the case of maximum likelihood estimation using (1), it is easy under infill asymptotics to show (see comments in Appendix B) that the maximum likelihood estimate is asymptotically normal with asymptotic covariance matrix

$$V = \left[\int_W z(u)^\mathsf{T} z(u) \lambda(u; \beta^*) \mathrm{d}u \right]^{-1}. \tag{9}$$

A similar expression is obtained using increasing domain asymptotics, see Rathbun & Cressie (1994) and Kutoyants (1998). Assuming (6) and following Rathbun et al. (2007), the solution of $u_{r,n}(\beta) = 0$ with

$$u_{r,n}(\beta) = \sum_{u \in X_n} z(u) - \sum_{u \in D_n} z(u) \frac{\lambda(u; \beta)}{n^{k-1}\rho} \tag{10}$$

is asymptotically normal with asymptotic covariance matrix

$$V^r = V + V G_g V / \rho^{1/k} \tag{11}$$

with $g(u) = z(u)\lambda(u;\beta^*)$, cf. (6). Note that this converges to V as $\rho \to \infty$.

Consider next the grid type estimating function,

$$u_{g,n}(\beta) = \sum_{u \in X_n} z(u) - \sum_{u \in X_n \cup D_n} z(u) \frac{n\lambda(u;\beta^*)}{n^k \rho N_{u,n} + n^k \rho} \tag{12}$$

where $T_{u,n}$ is the square tile to which u belongs (cf. Section 2.2) and $N_{u,n}$ is the number of points in $X_n \cap T_{u,n}$. In the case of stratified dummy points we assume continuously differentiable covariates $z_i(\cdot)$ to apply (6) with $k = 1/2$. The estimating function $u_{g,n}$ is then asymptotically equivalent to (10) (see Appendix B) and the asymptotic covariance matrix V^g is again given by (11).

3.1 The 'Dirichlet type' estimating function

For the 'Dirichlet type' estimating function, the estimate $\hat{\beta}_n$ is the solution of $u_{d,n}(\beta) = 0$ where

$$u_{d,n}(\beta) = \sum_{u \in X_n} z(u) - \sum_{u \in X_n \cup D_n} z(u) \frac{\lambda(u;\beta)}{\lambda(u;\beta) + n^{k-1}\rho}. \tag{13}$$

The following result is verified in Appendix A.

Theorem 1. *Assume that the matrices F and C given by*

$$F = \int_W z(u)^\mathsf{T} z(u) \mathrm{d}u \quad \textit{and} \quad C = \int_W z(u)^\mathsf{T} z(u) \frac{1}{\lambda(u;\beta^*)} \mathrm{d}u \quad (k < 1)$$

or $(k = 1)$

$$F = \int_W z(u)^\mathsf{T} z(u) \frac{\lambda(u;\beta^*)}{\rho + \lambda(u;\beta^*)} \mathrm{d}u \quad \textit{and} \quad C = \int_W z(u)^\mathsf{T} z(u) \frac{\lambda(u;\beta^*)}{(\lambda(u;\beta^*) + \rho)^2} \mathrm{d}u$$

are positive definite and that (6) *holds. Then with a probability tending to one, a solution $\hat{\beta}_n$ of* (13) *exists, and $n^{1/2}(\hat{\beta}_n - \beta^*) \to N(0, V^d)$ with*

$$V^d = F^{-1}CF^{-1} + F^{-1}G_g F^{-1}/\rho^{1/k} \tag{14}$$

where

$$g(u) = z(u) \quad (k < 1) \quad \textit{or} \quad g(u) = z(u) \frac{\lambda(u;\beta^*)}{\lambda(u;\beta^*) + \rho} \quad (k = 1).$$

Note that ρ controls the proportion of the asymptotic variance for $\hat{\beta}_n$ which is due to Monte Carlo integration error. Suppose $k = 1$ and $\rho \to \infty$. Then V^d tends to the asymptotic covariance matrix V of the MLE. In the case $k < 1$ we obtain in the limit $F^{-1}CF^{-1} \neq V$. The Dirichlet type estimating function is thus suboptimal in the case $k < 1$ even when $\rho \to \infty$.

3.2 Estimation of asymptotic covariances

Suppose in practice that the first component in β is an intercept and that an estimate $\hat{\beta}$ is obtained using (4) or (5) with M dummy points. Then the various integrals in the asymptotic covariance matrices may be estimated using that for a function $g(\cdot;\beta)$, an estimate of $\int_W g(u;\beta^*)\mathrm{d}u$ is given by

$$\sum_{u\in X_n\cup D_n} \frac{g(u;\beta^*)}{n\lambda(u;\beta^*)+n^k\rho}$$

and plugging in $X\cup D$ for $X_n\cup D_n$, $\exp(\hat{\beta}_1)$ for $n\exp(\beta_1^*)$, $\hat{\beta}_{2:p}$ for $\beta^*_{2:p}$, and $M/|W|$ for $n^k\rho$.

The above equation may also be used to estimate the integrals for G_g in the case of binomial dummy points. For stratified dummy points a consistent estimate of the ijth entry of G_g may be obtained using an additional set of dummy points $\tilde{D}=\{v_1,\dots,v_M\}$ distributed as and independent of $D=\{u_1,\dots,u_M\}$. Extending Okamoto (1976) to the multivariate case, the estimate is

$$\frac{1}{2}\sum_{l=1}^{M}(g_i(u_l)-g_i(v_l))(g_j(u_l)-g_j(v_l)).$$

Of course, averaging Monte Carlo estimates of $\int_W g(u)\mathrm{d}u$ based on the two sets of dummy points, the variance is halved. Similarly, we can replace the last term in (4) by

$$\frac{1}{2}\Big(\sum_{u\in X\cup D} z(u)\frac{\lambda(u;\beta)}{\rho N_u+\rho}+\sum_{u\in X\cup\tilde{D}} z(u)\frac{\lambda(u;\beta)}{\rho N_u+\rho}\Big)$$

in which case we should replace the last term in (11) by $VG_gV/(2\rho^{1/k})$.

4 Asymptotic distribution of parameter estimates in the case of an inhomogeneous cluster process

As an alternative to an inhomogeneous Poisson process, Waagepetersen (2007) considers a cluster process $X=\cup_{c\in Y}X_c$ where the X_c are clusters of 'offspring' associated with 'mother' points c in a stationary Poisson point process Y of intensity $\kappa>0$. Given Y, the clusters X_c are independent Poisson processes with intensity functions

$$\lambda_c(u)=\alpha\exp(z_{2:p}(u)\beta_{2:p}^{\mathsf{T}})h(u-c)$$

where $\alpha > 0$, $z_{2:p}(u) = (z_2(u), \ldots, z_p(u))$ is a vector of spatially varying covariates, $\beta_{2:p} = (\beta_2, \ldots, \beta_p)$ is a vector of regression parameters, and h is a probability density determining the spread of offspring points around c. Often h is given by a Gaussian density or a uniform distribution on a disc in which case an inhomogeneneous version of the so-called modified Thomas process or the Matérn cluster process is obtained.

The intensity function of X is of log-linear form $\exp(z(u)\beta^\mathsf{T})$ where $\beta_1 = \log(\kappa\alpha)$ and $z_1(u) = 1$. Waagepetersen (2007) suggests to estimate the regression parameter β using the estimating function (1). In the following we consider asymptotic results when (4) or (5) is used instead. Parameters of the density h may as suggested in Waagepetersen (2007) be estimated using a minimum contrast method based on the generalization of the K-function to the inhomogeneous case (Baddeley et al., 2000).

4.1 Asymptotic results

To obtain asymptotic results in the case of inhomogeneous cluster processes we consider a sequence of cluster processes X_n with increasing mother intensities $n\kappa^*$ and dummy point processes D_n of intensities $n^k\rho$. The intensity function of X_n is $n\lambda(u;\beta^*) = n\exp(z(u)\beta^*)$ with $\beta_1^* = \log(\kappa^*\alpha^*)$. The asymptotic covariance matrix in the case of completely observed covariates is (Waagepetersen, 2007)

$$V^c = V + VAV/\kappa^* \tag{15}$$

where V is given by (9), the last term in (15) is due to the clustering, and

$$A = \int J(c)^\mathsf{T} J(c)\mathrm{d}c \quad \text{with } J(c) = \int_W z(u)\lambda(u;\beta^*)h(u-c)\mathrm{d}u.$$

Consider first the grid type estimating function (12) with stratified dummy points and continuously differentiable covariates. In analogy with the Poisson process case we then obtain the asymptotic covariance matrix

$$V^{c,g} = V^c + VG_gV/\rho^{1/k}. \tag{16}$$

For the Dirichlet type weights where $\hat{\beta}_n$ is obtained by solving $u_{d,n}(\beta) = 0$ with $u_{d,n}$ given by (13), the asymptotic distribution is given by the following theorem.

Theorem 2. *Define the matrices F and C as in Theorem 1 and suppose that the conditions of Theorem 1 are satisfied. Moreover, let*

$$B = \int H(c)^\mathsf{T} H(c)\mathrm{d}c$$

where $(k<1)$

$$H(c)=\int_W z(u)h(u-c)\mathrm{d}u \quad \textit{or} \quad H(c)=\int_W \frac{z(u)\lambda(u;\beta^*)}{\lambda(u;\beta^*)+\rho}h(u-c)\mathrm{d}u \quad (k=1).$$

Then with a probability tending to one, a solution $\hat\beta_n$ *of* (13) *exists, and* $n^{1/2}(\hat\beta_n-\beta^*)\to N(0,V^{c,d})$ *with*

$$V^{c,d}=F^{-1}CF^{-1}+F^{-1}BF^{-1}/\kappa^*+F^{-1}G_gF^{-1}/\rho^{1/k} \tag{17}$$

where G_g *is given as in Theorem 1.*

A sketch of the proof is given in Appendix A. Note that (17) is obtained from (14) by adding the term $F^{-1}BF^{-1}/\kappa^*$ due to the clustering.

5 Comparison of asymptotic variances in a specific example

In this section we investigate the efficiency of the various estimating functions by evaluating their corresponding asymptotic covariance matrices for a specific example of spatial covariates. Figure 1 shows elevation $z_2(u)$ on a 5 by 5 m square grid covering a 500×1000 m^2 rain forest research plot W at Barro Colorado Island in Panama, see Condit et al. (1996); Condit (1998); Hubbell & Foster (1983). The elevations are in fact interpolated from data on a coarser grid but for sake of the example we here consider them as 'true' elevation observations.

In the following Section 5.1 we evaluate asymptotic variances in the case of a Poisson process with covariate vector $z(u)=(1,z_2(u))$ fixing $\beta_1^*=0$ and letting $\beta_2^*=0.01$, 0.1 or 1.0. In Section 5.2 a third covariate z_3 is used where z_3 is the norm of the gradient obtained from the elevation map. In all the examples, asymptotic covariance matrices are computed by approximating the integrals with Riemann sums corresponding to the 5 by 5 m grid. With stratified dummy points, numerical approximation of the partial derivatives of g (cf. (7)) is used when computing G_g appearing in the asymptotic covariance matrices.

5.1 Poisson process case

In the case where D_n is a binomial process of intensity $n\rho$, we let $\rho=q\int_W\exp(z(u)\beta^*)\mathrm{d}u/|W|$ for values of $q=0.25$, 1, 10, or 100, so that the number of dummy points $n\rho|W|$ is q times the expected number of observed

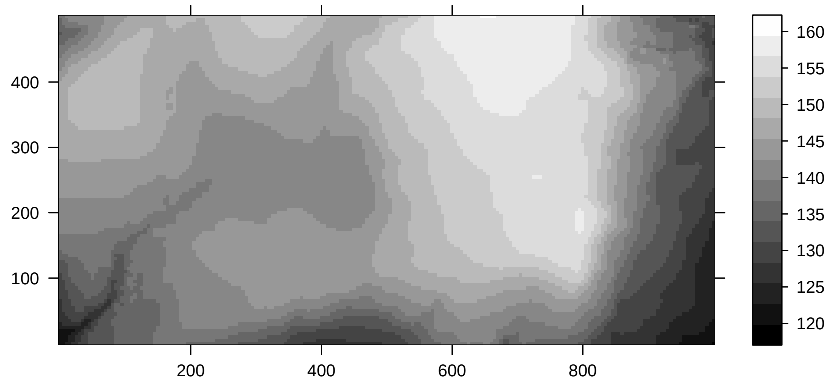

Figure 1: Elevation.

points. For stratified dummy points where $k = 1/2$, the proportion of dummy points in $X_n \cup D_n$ depends on n. To consider realistic values of ρ we imagine an n corresponding to an expected number $N = 1000 = n \int_W \exp(z(u)\beta^*)\mathrm{d}u$ of observed points and for various values of q choose $\rho = qN/(n^{1/2}|W|)$ so that the number of dummy points $M = n^{1/2}\rho|W| = qN$.

Table 1 shows ratios of asymptotic standard errors for the estimate $\hat{\beta}_2$ obtained from u_r, u_g, u_d or u given by (3), (4), (5), and (1), respectively. The standard errors are extracted from V^r, V^d, and V given by (11), (14), and (9), respectively. We consider u_g only in the case of stratified dummy points and recall that in this case the asymptotic covariance matrix V^g coincides with V^r.

Table 1: *Asymptotic standard errors for estimates of β_2 obtained from either* (3), (4), *or* (5) *divided by the asymptotic standard error for the MLE. The numbers of either binomial or stratified dummy points is q times the expected number of observed points and the 'true' parameter value β_2^* is either 0.01, 0.1, or 1.0.*

		Binomial					Stratified			
$\beta_2^* \setminus q$		0.25	1	10	100		0.25	1	10	100
.01	u_r	2.22	1.41	1.05	1.00	u_g	1.06	1.00	1.00	1.00
	u_d	2.21	1.41	1.05	1.00	u_d	1.06	1.01	1.01	1.01
0.1	u_r	2.47	1.51	1.06	1.01	u_g	1.08	1.01	1.00	1.00
	u_d	2.12	1.43	1.06	1.01	u_d	1.56	1.53	1.53	1.53
1.0	u_r	9.11	4.64	1.75	1.10	u_g	5.33	1.65	1.01	1.00
	u_d	3.68	2.52	1.47	1.09	u_d	$6{\times}10^6$	$6{\times}10^6$	$6{\times}10^6$	$6{\times}10^6$

When binomial dummy points are used, the Dirichlet type estimating function u_d performs better than the Rathbun et al. (2007) type estimating function u_r. For stratified dummy points on the other hand, the performance of u_d quickly deteriorates as β_2^* increases and already with $\beta_2^* = 0.1$ the standard errors become at least 53% larger than the MLE standard errors regardless of the value of ρ. Hence, in the case of stratified dummy points it is clearly preferable to use u_g rather than u_d. All of the Monte Carlo estimating equations perform less well as β_2^* and hence the variability of the intensity function increases. Note the potentially substantial increase in the standard errors which, depending on ρ and β_2^*, may occur due to missing covariate data.

5.2 Clustered rain forest trees

Waagepetersen (2007) fits an inhomogeneous cluster process with covariate vector $(1, z_2(u), z_3(u))$ to the positions of 3604 rain forest trees observed in the Barro Colorado Island research plot. The parameter estimates obtained for β_1, β_2, and β_3 are -4.99 (0.15), 0.02 (0.02), and 5.84 (2.53) (standard errors given in the parantheses). The cluster density h is given by a Gaussian density with standard deviation ω and the minimum contrast estimates of the clustering parameters κ and ω are $8{\times}10^{-5}$ and 20 (the asymptotic distribution of these parameters is a topic of current research). Due to clustering, the standard errors for β_1, β_2, and β_3 obtained from (15) are upto 10 times larger than the standard errors obtained from (9) assuming an inhomogeneous Poisson process.

We now investigate a hypothetical situation where the parameter estimate is obtained using (3), (4), or (5) assuming that Figure 1 does represent the true elevation map. We consider varying numbers $M = 450, 800, 1800$ of either binomial or stratified dummy points. For the binomial dummy points we consider (3) and (5) while (4) is used in the stratified case where the asymptotic covariance matrices for (3) and (4) coincide. Table 2 shows ratios between standard errors for estimates of β_2 extracted from $V^{c,r} = V^{c,g}$, $V^{c,d}$, and V^c given by (16), (17), and (15), respectively. A similar pattern is obtained for β_3 (not shown). In the computations, $\rho = M/|W|$, β^* is the estimate obtained in Waagepetersen (2007). Varying values of κ^* given by 1, 10, or 100 times $8{\times}10^{-5}$ are considered corresponding to decreasing degree of clustering, and $\alpha^* = \exp(\beta_1^*)/\kappa^*$ to ensure a constant expected number of points.

The results for the highly clustered case $\kappa^* = 8{\times}10^{-5}$ indicate that the increase in the parameter standard error due to the incompletely observed covariates is rather small and less than 1 % if 1800 dummy points are used. As the amount of clustering decreases, the incomplete observation of the

Table 2: *Asymptotic standard errors for estimates of* β_2 *obtained with* (3), (5), *or* (4) *divided by the asymptotic standard error for the estimated obtained with* (1). *Binomial dummy points are used for* (3) *and* (5) *while* (4) *is used with stratified dummy points. The number of dummy points is* M *and the asymptotic variances are evaluated with* $\beta^* = (-4.99, 0.02, 5.84)$ *and* κ^* *equal to* 1, 10, *or* 100 *times* $\hat{\kappa} = 8{\times}10^{-5}$ *from Waagepetersen (2007).*

	κ^*	$8{\times}10^{-5}$			$8{\times}10^{-4}$			$8{\times}10^{-3}$		
	M	450	800	1800	450	800	1800	450	800	1800
u_r	(bin.)	1.06	1.03	1.01	1.44	1.26	1.12	2.49	1.98	1.52
u_d	(bin.)	1.01	0.99	0.97	1.35	1.20	1.08	2.32	1.86	1.45
u_g	(str.)	1.00	1.00	1.00	1.04	1.01	1.00	1.17	1.06	1.01

covariates plays a relatively bigger role. For binomial dummy points, u_d again does better than u_r and curiously, the standard errors obtained with u_d and $M = 800$ or $M = 1800$ are in fact a bit smaller than with completely observed covariates. This is because the diagonal entries in the second term (due to clustering) of $V^{c,d}$ are smaller than those of the second term in V^c. As one might expect, with binomial dummy points (for which the covariates need not be continuously differentiable) we, for a given M, obtain larger standard errors than with stratified dummy points.

6 Discussion

The Monte Carlo versions of the `spatstat` estimating function can be implemented in much the same manner as the current `spatstat` estimating function. At the same time it is feasible to derive the asymptotic distribution of the associated parameter estimates. If the assumption of continuously differentiable covariates is tenable, the choice of stratified dummy points combined with the grid type estimating function (4) is preferable. Otherwise one may use the option of binomial dummy points and the Dirichlet type estimating function (5). One concern is the loss of efficiency which occurs with the Dirichlet type weights in the case $k < 1$ and $\rho \to \infty$ in Theorem 1. This may, however, also be an issue with the original `spatstat` estimating function when the `dirichlet` weights are used.

The asymptotic results in Section 3 require an experimental design where the distribution of the dummy points is chosen so that (6) holds. In Section 2.2 we discuss binomial and stratified dummy points. Other possibilities include vertices on a randomly translated lattice (see e.g. the review in Kiêu & Mora, 2006) or so-called scrambled nets (Owen, 1997). A central limit theorem is currently not available in the case of a randomly translated lattice whereas Loh (2003) establishes a central limit theorem for scrambled nets. From a theoretical point of view, scrambled nets offer better convergence rates than stratified dummy points but the implementation is much less straightforward. The description of scrambled nets is moreover quite technical and omitted here for brevity.

Acknowledgment

This research was funded by the Danish Natural Science Research Council, grant 272-06-0442, 'Point process modelling and statistical inference'. Adrian Baddeley and Stephen Rathbun are gratefully acknowledged for enlightening comments regarding respectively the `grid` option in `spatstat` and infill asymptotics. The detailed and constructive comments by two referees improved considerably the presentation of the paper.

A Proof of Theorem 1

Recalling the notation introduced in Section 2 and 3 we here give a proof of Theorem 1 and sketch a proof of Theorem 2.

Proof of Theorem 1:

In the following identify X_n with a union of n independent Poisson processes X^i each of intensity $\lambda(\cdot;\beta^*)$. Let

$$j_{d,n}(\beta) = -\frac{\mathrm{d}}{\mathrm{d}\beta} u_{d,n}(\beta) = n^k \sum_{u \in X_n \cup D_n} z(u)^\mathsf{T} z(u) \frac{\rho n \lambda(u;\beta)}{(n\lambda(u;\beta) + n^k \rho)^2}.$$

We consider first the case $k < 1$ and verify the conditions 1-3 in Appendix C with $a_n = n^{1-k}$; the case $k = 1$ follows along similar lines but with $a_n = 1$.

Turning to condition 1, note that

$$n^{-k} j_{d,n}(\beta^*) = \frac{1}{n} \sum_{u \in X_n} \frac{z(u)^\mathsf{T} z(u) \rho \lambda(u;\beta^*)}{(\lambda(u;\beta^*) + n^{k-1}\rho)^2} + \frac{1}{n} \sum_{u \in D_n} \frac{z(u)^\mathsf{T} z(u) \rho \lambda(u;\beta^*)}{(\lambda(u;\beta^*) + n^{k-1}\rho)^2}$$

The last term has mean value of order n^{k-1} and hence converges to zero in probability by Markovs inequality. The first term converges to ρF by the strong law of large numbers replacing $\sum_{u \in X_n}$ by $\sum_{i=1}^n \sum_{u \in X^i}$. Condition 2 follows by continuity and Markovs inequality. Hence the main task is to verify condition 3.

Rewrite

$$u_{d,n}(\beta^*) = V_n - W_n =$$
$$\left(V_n - n^k \int_W z(u) \frac{\rho\lambda(u;\beta^*)}{\lambda(u;\beta^*) + n^{k-1}\rho} \mathrm{d}u\right) - \left(W_n - n^k \int_W z(u) \frac{\rho\lambda(u;\beta^*)}{\lambda(u;\beta^*) + n^{k-1}\rho} \mathrm{d}u\right)$$

where the two terms

$$V_n = \sum_{u \in X_n} z(u) \frac{n^{k-1}\rho}{\lambda(u;\beta^*) + n^{k-1}\rho} \quad \text{and} \quad W_n = \sum_{u \in D_n} z(u) \frac{\lambda(u;\beta^*)}{\lambda(u;\beta^*) + n^{k-1}\rho}$$

are independent. Note

$$V_n \sim \sum_{i=1}^{n} Y_{i,n} \quad \text{where} \quad Y_{i,n} = \sum_{u \in X^i} z(u) \frac{n^{k-1}\rho}{\lambda(u;\beta^*) + n^{k-1}\rho}.$$

Let $\mu_n = \mathbb{E}Y_{i,n} = n^{k-1} \int_W z(u)\rho\lambda(u;\beta^*)/(\lambda(u;\beta^*) + n^{k-1}\rho)\mathrm{d}u$ and

$$\sigma_n^2 = \mathbb{V}\mathrm{ar}Y_{i,n} = n^{2k-2} \int_W z(u)^\mathsf{T} z(u) \frac{\rho^2\lambda(u;\beta^*)}{(\lambda(u;\beta^*) + n^{k-1}\rho)^2} \mathrm{d}u.$$

Then by the Lindeberg-Feller central limit theorem (e.g. Proposition 2.27 in Van der Vaart, 1998), $n^{-1/2}\sigma_n^{-1} \sum_{i=1}^n (Y_{i,n} - \mu_n)$ converges to a standard multivariate normal distribution. Note that $\lim_{n\to\infty} \sigma_n^2/n^{2k-2} = \rho^2 C = \rho^2 \int_W z(u)^\mathsf{T} z(u)/\lambda(u;\beta^*)\mathrm{d}u$. Hence

$$n^{-k+1/2} \sum_{i=1}^{n} (Y_{i,n} - \mu_n) = n^{-k+1/2}(V_n - n^k \int_W z(u) \frac{\rho\lambda(u;\beta^*)}{\lambda(u;\beta^*) + n^{k-1}\rho} \mathrm{d}u)$$

converges to $N(0, \rho^2 C)$.

Considering W_n,

$$\begin{aligned} n^{-k+1/2} W_n =& n^{-k+1/2} \sum_{u \in D_n} z(u) \frac{\lambda(u;\beta^*)}{\lambda(u;\beta^*) + n^{k-1}\rho} = \\ & n^{1/2} \sum_{u \in D_n} \frac{z(u)\rho}{n^k\rho} - n^{1/2} \sum_{u \in D_n} \frac{z(u) n^{k-1}\rho^2}{(\lambda(u;\beta^*) + n^{k-1}\rho) n^k \rho} \end{aligned}$$

where the last term converges to zero in probability since

$$\lim_{n\to\infty} \mathbb{V}\mathrm{ar}\, n^{1/2} \sum_{u \in D_n} \frac{z(u) n^{k-1}\rho^2}{(\lambda(u;\beta^*) + n^{k-1}\rho) n^k \rho} =$$
$$\lim_{n\to\infty} n^{2k-2} \mathbb{V}\mathrm{ar}\, n^{1/2} \sum_{u \in D_n} \frac{z(u)\rho^2}{\lambda(u;\beta^*) n^k \rho} = 0$$

as $\mathbb{V}\text{ar}\, n^{1/2} \sum_{u \in D_n} z(u)\rho^2/(\lambda(u;\beta^*)n^k\rho)$ converges to a constant.

Hence $n^{-k+1/2}\big(W_n - n^k \int_W z(u) \frac{\rho\lambda(u;\beta^*)}{\lambda(u;\beta^*)+n^{k-1}\rho}\mathrm{d}u\big)$ is asymptotically normal with covariance matrix $\rho^{2-1/k}G_z$ and we obtain that $n^{-k+1/2}u_{d,n}(\beta^*)$ converges to $N(0,\rho^2 C) + N(0,\rho^{2-1/k}G_z)$. Theorem 1, case $k < 1$, thus follows from the results in Appendix C. The proof for $k = 1$ proceeds in a similar manner.

Proof of Theorem 2:

The proof of Theorem 2 is analogous to the proof of Theorem 1 except that we obtain a different asymptotic covariance matrix for $u_{d,n}(\beta^*)$ identifying X_n with a superposition of independent cluster processes X^i where X^i has intensity function $\lambda(\cdot;\beta^*)$ and consists of offspring for mothers in a stationary Poisson process Y^i of intensity κ^*.

Assume $k < 1$. The variance $\mathbb{V}\text{ar} \sum_{u \in X^i} z(u)n^{k-1}\rho/(\lambda(u;\beta) + n^{k-1}\rho)$ is computed using conditioning on Y^i,

$$
\begin{aligned}
\sigma_n^2 =& \mathbb{V}\text{ar} \sum_{u \in X^i} \frac{z(u)n^{k-1}\rho}{\lambda(u;\beta^*) + n^{k-1}\rho} = \\
& \mathbb{E}\mathbb{V}\text{ar}[\sum_{u \in X^i} \frac{z(u)n^{k-1}\rho}{\lambda(u;\beta^*) + n^{k-1}\rho}|Y^i] + \mathbb{V}\text{ar}\mathbb{E}[\sum_{u \in X^i} \frac{z(u)n^{k-1}\rho}{\lambda(u;\beta^*) + n^{k-1}\rho}|Y^i] = \\
& n^{2k-2}\int_W z(u)^\mathsf{T} z(u)\frac{\rho^2\lambda(u;\beta^*)}{(\lambda(u;\beta^*) + n^{k-1}\rho)^2}\mathrm{d}u + n^{2k-2}\rho^2 \int H_n(c)^\mathsf{T} H_n(c)\mathrm{d}c/\kappa^*
\end{aligned}
$$

where

$$
H_n(c) = \int_W z(u)\frac{\lambda(u;\beta^*)}{\lambda(u;\beta^*) + n^{k-1}\rho}h(u-c)\mathrm{d}u
$$

Following the proof of Theorem 1 it follows that $n^{-k+1/2}u_{d,n}(\beta^*)$ is asymptotically zero mean normal with covariance matrix

$$
\rho^2 C + \rho^2 \int H(c)^\mathsf{T} H(c)\mathrm{d}c/\kappa^* + \rho^{2-1/k}G_g.
$$

The asymptotic variance in the case $k = 1$ is obtained in a similar manner.

B Asymptotic equivalence of estimating functions

The asymptotic distribution of parameter estimates obtained with the estimating functions (1) and (3) can be derived along the lines of the proofs in

Appendix A using the general results in Appendix C. The basic steps are to establish asymptotic normality of $n^{-1/2}$ times the estimating function and convergence of n^{-1} times minus the derivative of the estimating function.

Consider now $u_{r,n}$ and $u_{g,n}$ given by (10) and (12). Assuming that the covariates $z_i(u)$ are continuously differentiable and since the sidelength of $T_{u,n}$ is a constant times $n^{-k/2}$, it follows that $n^{-1/2}(u_{g,n}(\beta^*) - u_{r,n}(\beta^*))$ tends to zero in probability and the two terms thus have the same weak limit. Similarly, $j_{g,n}(\beta^*)/n$ has the same limit in probability as $j_{r,n}(\beta^*)/n$ where $j_{g,n}$ and $j_{r,n}$ denote the derivatives of $-u_{g,n}$ and $-u_{r,n}$. Hence, the parameter estimates obtained from $u_{g,n}$ and $u_{r,n}$ are identically distributed asymptotically.

C Some general asymptotic results for estimating functions

Consider a parametrized family of probability measures P_θ, $\theta \in \mathbb{R}^p$, and a sequence of estimating functions $u_n : \mathbb{R}^p \to \mathbb{R}^p$, $n \geq 1$, with negated derivatives j_n. The 'true' parameter value is denoted θ^* and for a matrix $A = [a_{ij}]$, $\|A\|^2 = \sum_{i,j} a_{ij}^2$. Suppose that there exist a sequence $a_n \neq 0, n \geq 1$, and positive definite matrices F and Σ so that

1. $\|a_n j_n(\theta^*)/n - F\| \to 0$ in probability,
2. for all $c > 0$, $\sup_{\theta:\sqrt{n}\|\theta-\theta^*\|\leq c} \|a_n j_n(\theta)/n - a_n j_n(\theta^*)/n\| \to 0$ in probability, and
3. the normalized score function $a_n u_n(\theta^*)/\sqrt{n}$ is asymptotically zero-mean normal with covariance matrix Σ.

It then follows from Corollary 2.5 and Theorem 2.8 in Sørensen (1999) that with a probability tending to one, there exists a solution $\hat{\theta}_n$ of $u_n(\theta) = 0$, and

$$\sqrt{n}(\hat{\theta}_n - \theta^*) \to N(0, F^{-1}\Sigma F^{-1}).$$

References

Baddeley, A. & Turner, R. (2000). Practical maximum pseudolikelihood for spatial point patterns. *Australian and New Zealand Journal of Statistics* **42**, 283–322.

Baddeley, A. & Turner, R. (2005). Spatstat: an R package for analyzing spatial point patterns. *Journal of Statistical Software* **12**, 1–42. URL: `www.jstatsoft.org`, ISSN: 1548-7660.

Baddeley, A. J., Møller, J. & Waagepetersen, R. (2000). Non- and semi-parametric estimation of interaction in inhomogeneous point patterns. *Statistica Neerlandica* **54**, 329–350.

Berman, M. & Turner, R. (1992). Approximating point process likelihoods with GLIM. *Applied Statistics* **41**, 31–38.

Condit, R. (1998). *Tropical Forest Census Plots*. Berlin, Germany and Georgetown, Texas: Springer-Verlag and R. G. Landes Company.

Condit, R., Hubbell, S. P. & Foster, R. B. (1996). Changes in tree species abundance in a neotropical forest: impact of climate change. *Journal of Tropical Ecology* **12**, 231–256.

Cox, D. R. (1972). The statistical analysis of dependencies in point processes. In *Stochastic Point Processes*, P. A. W. Lewis, ed. New York: Wiley, pp. 55–66.

Guan, Y. & Loh, J. M. (2007). A thinned block bootstrap procedure for modeling inhomogeneous spatial point patterns. *Journal of the American Statistical Association*, to appear.

Hubbell, S. P. & Foster, R. B. (1983). Diversity of canopy trees in a neotropical forest and implications for conservation. In *Tropical Rain Forest: Ecology and Management*, S. L. Sutton, T. C. Whitmore & A. C. Chadwick, eds. Blackwell Scientific Publications, pp. 25–41.

Kiêu, K. & Mora, M. (2006). Precision of stereological planar area predictors. *Journal of Microscopy* **222**, 201–211.

Kutoyants, Y. A. (1998). *Statistical inference for spatial Poisson processes*. No. 134 in Lecture Notes in Statistics. New York: Springer.

Lawson, A. B. (1992). GLIM and normalising constant models in spatial and directional data analysis. *Computational Statistics and Data Analysis* **13**, 331–348.

Loh, W.-L. (2003). On the asymptotic distribution of scrambled net quadrature. *Annals of Statistics* **31**, 1282–1324.

Møller, J. & Waagepetersen, R. P. (2006). Modern statistics for spatial point processes. *Scandinavian Journal of Statistics*, to appear.

Okamoto, M. (1976). Asymptotic normality in Monte Carlo integration. *Mathematics of Computation* **30**, 831–837.

Owen, A. (1997). Scrambled net variance for integrals of smooth functions. *Annals of Statistics* **25**, 1541–1562.

Rathbun, S. L. (1996). Estimation of Poisson intensity using partially observed concomitant variables. *Biometrics* **52**, 226–242.

Rathbun, S. L. & Cressie, N. (1994). Asymptotic properties of estimators for the parameters of spatial inhomogeneous Poisson processes. *Advances in Applied Probability* **26**, 122–154.

Rathbun, S. L., Shiffman, S. & Gwaltney, C. J. (2007). Modelling the effects of partially observed covariates on Poisson process intensity. *Biometrika* **94**, 153–165.

Schoenberg, F. P. (2005). Consistent parametric estimation of the intensity of a spatial-temporal point process. *Journal of Statistical Planning and Inference* **128**, 79–93.

Sørensen, M. (1999). On asymptotics of estimating functions. *Brazilian Journal of Probability and Statistics* **13**, 111–136.

Van der Vaart, A. W. (1998). *Asymptotic statistics*. Cambridge: Cambridge University Press.

Waagepetersen, R. (2007). An estimating function approach to inference for inhomogeneous Neyman-Scott processes. *Biometrics* **63**, 252–258.

Wang, Y. & Lawson, A. B. (2006). An evaluation of triangulation weights in fitting spatial Poisson process models. Preprint.