Estimating functions for inhomogeneous spatial point processes with incomplete covariate data

Waagepetersen, Rasmus

AALBORG UNIVERSITY

Estimating functions for inhomogeneous spatial point processes with incomplete covariate data

by

Rasmus P. Waagepetersen

R-2007-04

January 2007

Department of Mathematical Sciences
Aalborg University
Fredrik Bajers Vej 7 G ▪ DK - 9220 Aalborg Øst ▪ Denmark
Phone: +45 96 35 80 80 ▪ Telefax: +45 98 15 81 29
URL: www.math.auc.dk/research/reports/reports.htm

ISSN 1399–2503 ▪ On-line version ISSN 1601–7811

Estimating functions for inhomogeneous spatial point processes with incomplete covariate data

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

Abstract

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.

Keywords: 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 pattern X 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 $\lambda(u;\beta) = \exp(z(u)\beta^{\mathsf{T}})$ and 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 W is the observation window and $z(u)$ is the vector of spatial covariates at location $u \in W$. A spatial Poisson process is often not appropriate due to clustering not explained by the covariates. However, (1) may still be used as an estimating function for regression parameters in certain inhomogeneous cluster processes, see Waagepetersen (2006) and Møller and Waagepetersen (2006).

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. One disadvantage of this approach is the need to specify a model for the covariate process (typically involving new parameters to be estimated). In practice, the score function is often approximated by an estimating function

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

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. This is implemented in the `R` package `spatstat` (Baddeley and Turner, 2000, 2005) where for computational reasons explained in Section 2, $X \subset Q$. 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.* (2006) suggest to use instead an estimating equation 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.* (2006) demonstrate asymptotic normality of the estimate obtained using (3) under conditions valid 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.* (2006) and introduce in Section 2 Monte Carlo versions of the `spatstat` estimating equation. Asymptotic normality of the associated parameter estimates is discussed in Section 3 both for inhomogeneous Poisson processes and inhomogeneous cluster processes. Practical examples are considered in Section 4, and Section 5 contains some closing remarks.

2 Monte Carlo versions of the `spatstat` estimating function

Two types of quadrature schemes are available in `spatstat`: `grid` and `dirichlet`. In both cases, the set of quadrature points Q is the union of the observed points X and a set of dummy points D. The approximate likelihood is then formally equivalent to the likelihood of a weighted Poisson regression which can easily be maximized using standard software for generalized linear models (Berman and Turner, 1992). For the `grid` option, the observation window is divided into rectangular tiles C_v, $v \in D$, each containing exactly one dummy point $v \in D$. The quadrature weight for a quadrature point $u \in Q$ falling in a tile C_v is the area of C_v divided by the number of quadrature points falling in C_v (hence adjusting for the possible multiple occurrence of a tile C_v 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 cells for the Dirichlet tesselation generated by the quadrature points Q. A thorough account of the quadrature schemes is given in Baddeley and Turner (2000).

Our Monte Carlo versions of the `spatstat` estimating function are of the form

$$\sum_{u \in X} z(u) - \sum_{u \in X \cup D} z(u)\lambda(u;\beta)w_\beta(u) \tag{4}$$

where D is a homogeneous dummy point process of intensity $\rho > 0$ (see Section 2.1) and $w_\beta(u)$ is a weight which may depend on β. The first instance of (4) is equivalent to the `grid` option in `spatstat` but with random stratified dummy points, see Section 2.1 (this possibility is briefly mentioned in Section 4.3 of Baddeley and Turner, 2000). Consider as above rectangular tiles $C_v, v \in D$ with $v \in C_v$, $v \in D$, and let N_v denote the number of points in $X \cap C_v$. Letting $w_\beta(u) = (\rho N_u + \rho)^{-1}$,

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

is obtained. Note that if the C_u are very small, then the dummy points in $X \cap C_u$ essentially become replicates of the dummy point $u \in D$ and the last term in (5) is well approximated by $\sum_{u \in D} z(u)\lambda(u;\beta)/\rho$ as in (3).

To some extent, an analogue of the `dirichlet` version of `spatstat` is obtained with

$$u_d(\beta) = \sum_{u \in X} z(u) - \sum_{u \in X \cup D} z(u)\frac{\lambda(u;\beta)}{\lambda(u;\beta) + \rho} \tag{6}$$

where $w_\beta(u) = (\lambda(u;\beta)+\rho)^{-1}$ is the inverse 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 to the expected area of a Dirichlet cell in a region of constant intensity $\lambda(u;\beta)+\rho$. Although intuitively appealing, (6) yields an asymptotically suboptimal estimating function for certain choices of dummy point distributions, see Section 3.1.

Let Z be the matrix with rows $z(u)$, $u \in X \cup D$, and V the diagonal matrix with diagonal entries $\lambda(u;\beta)/(\rho N_u + \rho)$ in the case of (5) or $\rho\lambda(u;\beta)/(\lambda(u;\beta)+\rho)^2$ in the case of (6). Then we may write $u_s(\beta)$ and $j(\beta) = -\mathrm{d}u_s(\beta)/\mathrm{d}\beta$ as

$$Z^{\mathsf{T}}\big(1[u \in x] - w_\beta(u)\lambda(u;\beta)\big)_{u \in X \cup D} = Z^T V^{1/2} y \text{ and } j(\beta) = Z^{\mathsf{T}} V Z$$

where $y = V^{-1/2}\big(1[u \in x] - w_\beta(u)\lambda(u;\beta)\big)_{u \in X \cup D}$. Newton-Raphson steps for solving $u_s(\beta) = 0$ thus become equivalent to iterative weighted least squares and can be implemented using minor modifications of code for estimation in generalized linear models.

2.1 Dummy point distributions

To establish asymptotic results for the estimating functions (5) and (6) 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 $f : W \to \mathbb{R}^p$,

$$n^{1/2}\left[\sum_{u \in D_n} \frac{f(u)}{n^k\rho} - \int_W f(u)\mathrm{d}u\right] \to N(0, G_f/\rho^{1/k}) \tag{7}$$

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 (7) 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.* (2006). 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 (7) 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, (7) holds with

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

where

$$A_f(u_1, u_2) = \left[\frac{\partial f_i}{\partial u_1}\frac{\partial f_j}{\partial u_1} + \frac{\partial f_i}{\partial u_2}\frac{\partial f_j}{\partial u_2}\right]. \tag{8}$$

Note that when using (5) and stratified dummy points we naturally choose $C_{u,n} = s_{i,n}$ 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

The asymptotic distribution of parameter estimates is obtained using infill asymptotics where both the intensities of X and D tend to infinity. One may think of X as representing the accumulation of points up to a certain 'time' point n and the intensity of X is then proportional to n. The intensity of the dummy points is chosen to match the increasing intensity of observed points as n increases. More specifically we consider sequences of Poisson point processes X_n and dummy point processes D_n with intensity functions

$$\lambda_n(u; \beta^*) = n\lambda(u; \beta^*),\ \beta^* \in \mathbb{R}^p \quad \text{and} \quad \rho_n = n^k\rho \tag{9}$$

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 tends to infinity at a slower rate than the intensity of X_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

$$V = \left[\int_W z(u)^\mathsf{T} z(u)\lambda(u; \beta^*)\mathrm{d}u\right]^{-1}. \tag{10}$$

A similar expression is obtained using increasing domain asymptotics, see Rathbun and Cressie (1994) and Kutoyants (1998). Assuming (7) and following Rathbun *et al.* (2006), the solution of $u_{r,n}(\beta) = 0$ with

$$u_{r,n}(\beta) = \sum_{u \in X_n} z(u) - \sum_{u \in D_n} z(u) \frac{\lambda(u;\beta)}{n^{k-1}\rho} \tag{11}$$

is asymptotically normal with asymptotic covariance matrix

$$V^r = V + VG_gV/\rho^{1/k} \tag{12}$$

with $g(u) = z(u)\lambda(u;\beta^*)$, cf. (7). Note that this converges to V as $\rho \to \infty$.

Consider next the grid type estimating function,

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

where $C_{u,n}$ is the square cell to which u belongs (cf. Section 2.1) and $N_{u,n}$ is the number of points in $X_n \cap C_{u,n}$. In the case of stratified dummy points we assume continuously differentiable covariates $z_i(\cdot)$ to apply (7) with $k = 1/2$. The estimating function $u_{g,n}$ is then asymptotically equivalent to (11) (see Appendix B) and the asymptotic covariance matrix V^g is again given by (12).

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

$$u_{d,n}(\beta) = \sum_{u \in X_n} z(u) - \sum_{u \in X_n \cup D_n} z(u) \frac{\lambda(u;\beta)}{\lambda(u;\beta) + n^{k-1}\rho}. \tag{14}$$

The following result is verified in Appendix A.

Theorem 1. *Assume that the matrix Z_n with rows $z(u)$, $u \in X_n \cup D_n$, has full rank almost surely and that (7) holds. Further, let*

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

or $(k = 1)$

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

Then $n^{1/2}(\hat{\beta}_n - \beta^*) \to N(0, V^d)$ *with*

$$V^d = F^{-1}CF^{-1} + F^{-1}G_g F^{-1}/\rho^{1/k} \tag{15}$$

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 Inhomogeneous cluster processes

As an alternative to an inhomogeneous Poisson process, Waagepetersen (2006) considers a cluster process $X = X_{c \in Y}$ 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. The intensity function of X is then of log-linear form $\exp(z(u)\beta^{\mathsf{T}})$ where $\beta_1 = \log(\kappa\alpha)$ and $z_1(u) = 1$. Waagepetersen (2006) suggests to estimate the regression parameter β using the estimating function (1). In the following we consider asymptotic results when (5) or (6) is used instead.

3.3 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, 2006)

$$V^c = V + VAV/\kappa^* \tag{16}$$

where A is defined as B in Theorem 2 below but with $H(c) = \int_W z(u)\lambda(u;\beta)$ $h(u-c)\mathrm{d}u$.

Consider first the grid type estimating function (13) with stratified dummy points and continuously differentiable covariates. In analogy with the Poisson process case we then obtain the asymptotic covariance matrix

$$V^{c,g} = V^c + VG_gV/\rho^{1/k}. \tag{17}$$

For the Dirichlet type weights where $\hat{\beta}_n$ is obtained by solving $u_{d,n}(\beta) = 0$ with $u_{d,n}$ given by (14), the asymptotic distribution is given by the following theorem.

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

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

where ($k < 1$)

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

Then $n^{1/2}(\hat{\beta}_n - \beta^*) \to N(0, V^{c,d})$ with

$$V^{c,d} = F^{-1}CF^{-1} + F^{-1}BF^{-1}/\kappa^* + F^{-1}G_gF^{-1}/\rho^{1/k} \tag{18}$$

where G_g is given as in Theorem 1.

A sketch of the proof is given in Appendix A.

3.4 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 (5) or (6) 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 $\{v_1, \ldots, v_M\}$ distributed as and independent of $\{u_1, \ldots, u_M\}$. 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.

4 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 and 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 4.1 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 4.2 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. (8)) are used when computing G_g appearing in the asymptotic covariance matrices.

4.1 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 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

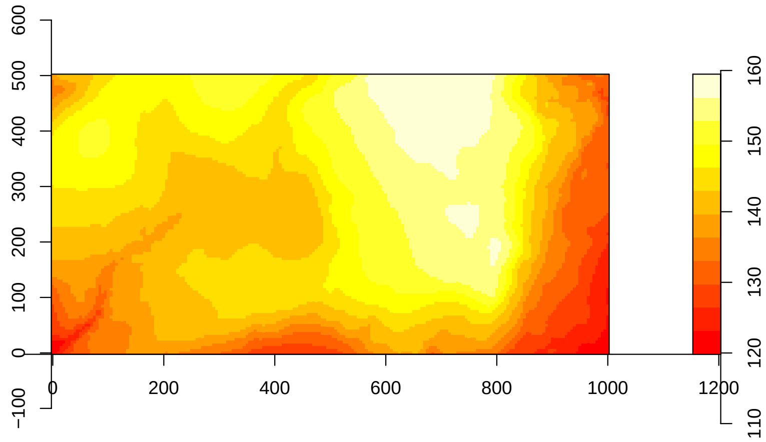

Figure 1: Elevation.

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), (5), (6), and (1), respectively. The standard errors are extracted from V^r, V^d, and V given by (12), (15), and (10), 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.

		bin.				str.			
β_2^*	Est. fct. \q	0.25	1	10	100	0.25	1	10	100
.01	u_r (u_g)	2.22	1.41	1.05	1.00	1.06	1.00	1.00	1.00
	u_d	2.21	1.41	1.05	1.00	1.06	1.01	1.01	1.01
0.1	u_r (u_g)	2.47	1.51	1.06	1.01	1.08	1.01	1.00	1.00
	u_d	2.12	1.43	1.06	1.01	1.56	1.53	1.53	1.53
1.0	u_r (u_g)	9.11	4.64	1.75	1.10	5.33	1.65	1.01	1.00
	u_d	3.68	2.52	1.47	1.09	6e6	6e6	6e6	6e6

Table 1: Asymptotic standard errors for estimates of β_2 obtained from either (3), (5), or (6) 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.

When binomial dummy points are used, the Dirichlet type estimating function u_d does in general a bit better than the Rathbun *et al.* (2006) 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 grid weights rather than Dirichlet type weights. 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.

4.2 Clustered rain forest trees

Waagepetersen (2006) 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 and κ are $(-4.99, 0.02, 5.84)$ and 8e-5. Due to clustering, the standard errors for β_2 and β_3 obtained from (16) are respectively 8.8 and 9.9 times larger than the standard errors obtained from (10) assuming an inhomogeneous Poisson process.

We now investigate a hypothetical situation where the parameter estimate is obtained using (3), (5), or (6) 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 (6) while (5) is used in the stratified case where the asymptotic covariance matrices for (3) and (5) 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 (17), (18), and (16), respectively. A similar pattern is obtained for β_3 (not shown). In the computations, $\rho = M/|W|$, β^* is the estimate obtained in Waagepetersen (2006), while varying values of κ^* given by 1, 10, or 100 times 8e-5 are considered corresponding to decreasing degree of clustering.

The results for the highly clustered case $\kappa^* = $ 8e-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 covariates plays a relatively bigger role. For binomial dummy points, u_d again does a bit 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

	κ^*	8e-5			8e-4			8e-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

Table 2: Asymptotic standard errors for estimates of β_2 obtained with (3), (6), or (5) divided by the asymptotic standard error for the MLE. Binomial dummy points are used for (3) and (6) while (5) 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} = $ 8e-5 from Waagepetersen (2006).

the covariates need not be continuously differentiable) we, for a given M, obtain larger standard errors than with stratified dummy points.

5 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 (5) seems preferable. Otherwise one may use the option of binomial dummy points and the Dirichlet type estimating function (6). 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` quadrature scheme is used.

The asymptotic results in Section 3 require an experimental design where the distribution of the dummy points is chosen so that (7) holds. In Section 2.1 we discuss binomial and stratified dummy points. Other possibilities include vertices on a randomly translated lattice (see e.g. the review in Kiêu and 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.

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 1-5 of Theorem 3 in Appendix C with $a_n = n^{k/2}$ and $c_n = n^{k-1/2}$; the case $k = 1$ follows along similar lines but with $a_n = c_n = n^{1/2}$ – the first condition of Theorem 3 is satisfied for both choices of a_n and c_n. Condition 2 holds since Z_n has full rank almost surely, see also the discussion in the first part of Section 2.

Turning to condition 3, 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 4 follows by continuity and Markovs inequality. Hence the main task is to verify condition 5.

Rewrite

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

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 central limit theorem (CLT II in Hoffmann-Jørgensen, 1994, using condition 5.22.5), $n^{-1/2}\sigma_n^{-1}\sum_{i=1}^{n}(Y_{i,n} - \mu_n)$ converges to a standard multivariate normal distribution. Note that $\lim_{n\to\infty} \sigma_n^2/n^{2k-2} \to \rho^2 C = \rho^2 \int_W z(u)^\mathsf{T} z(u)/\lambda(u;\beta^*)\mathrm{d}u$. 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

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

as $\mathbb{V}\mathrm{ar} n^{1/2} \sum_{u \in D_n} z(u)\rho^2/(\lambda(u;\beta^*)n^k\rho)$ 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 Theorem 3. 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 $\mathbb{V}\mathrm{ar}\sum_{u\in X^i} z(u)n^{k-1}\rho/(\lambda(u;\beta)+n^{k-1}\rho)$ is computed using conditioning on Y^i,

$$\sigma_n^2 = \mathbb{V}\mathrm{ar}\sum_{u\in X^i}\frac{z(u)n^{k-1}\rho}{\lambda(u;\beta^*)+n^{k-1}\rho} =$$

$$\mathbb{E}\mathbb{V}\mathrm{ar}[\sum_{u\in X^i}\frac{z(u)n^{k-1}\rho}{\lambda(u;\beta^*)+n^{k-1}\rho}|Y^i] + \mathbb{V}\mathrm{ar}\mathbb{E}[\sum_{u\in X^i}\frac{z(u)n^{k-1}\rho}{\lambda(u;\beta^*)+n^{k-1}\rho}|Y^i] =$$

$$n^{2k-2}\int_W z(u)^\mathsf{T} z(u)\frac{\rho^2\lambda(u;\beta^*)}{(\lambda(u;\beta^*)+n^{k-1}\rho)^2}\mathrm{d}u + n^{2k-2}\rho^2\int H_n(c)^\mathsf{T} H_n(c)\mathrm{d}c/\kappa^*$$

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 Theorem 3 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 (11) and (13). Assuming that the covariates $z_i(u)$ are continuously differentiable and since the sidelength of $C_{u,n}$ 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

The following results are adapted from unpublished lecture notes by Professor Jens L. Jensen, University of Aarhus. The use below of normalizing sequences a_n and c_n where a_n may differ from c_n is not standard in the literature on asymptotics for estimating functions. However, we need this to deal with the case $k < 1$ in the proofs of Theorems 1 and 2.

Consider a parametrized family of probability measures P_θ, $\theta \in \mathbb{R}^p$, and a sequence of estimating functions $u_n : \mathbb{R}^p \to \mathbb{R}^p$, $n \geq 1$. The 'true' parameter value is denoted θ^*. The conditions for consistency and asymptotic normality are that there exist sequences a_n and c_n such that

1. $c_n a_n^{-2} \to 0$,
2. $u_n(\theta) = 0$ has almost surely a unique solution $\hat{\theta}_n$ for each n,
3. $j_n(\theta^*)/a_n^2 \to F$ in probability for a positive definite matrix F,
4. For all $c > 0$, $\sup_{\|\theta - \theta^*\| a_n^2/c_n \leq c} |j_n(\theta) - j_n(\theta^*)|_{\max}/a_n^2 = 0$ in probability,
5. The normalized score function $u_n(\theta^*)/c_n$ is asymptotically zero-mean normal with covariance matrix Σ,

where probabilities are computed under P_θ^* and $|A|_{\max} = \max_{ij} |a_{ij}|$ for a matrix $A = [a_{ij}]$. The second condition is assumed for ease of exposition and can be relaxed in view of the third condition. The following theorem ensures a_n^2/c_n consistency of $\hat{\theta}_n$ and asymptotic normality.

Theorem 3. *Under the conditions stated above, for each $\epsilon > 0$, there exists a $c > 0$ such that*

$$P_{\theta^*}(\|\hat{\theta}_n - \theta^*\| a_n^2/c_n < c) > 1 - \epsilon$$

whenever n is sufficiently large. Moreover,

$$(\hat{\theta}_n - \theta^*) a_n^2/c_n \to N(0, F^{-1}\Sigma F^{-1}).$$

A proof of the result can be found in Waagepetersen (2007).

References

Baddeley, A. & Turner, R. (2000). Practical maximum pseudolikelihood for spatial point patterns (with discussion). *Australian and New Zealand Journal of Statistics* **42**(3), 283–322.

Baddeley, A. & Turner, R. (2005). Spatstat: an R package for analyzing spatial point patterns. *Journal of Statistical Software* **12**(6), 1–42, URL: `www.jstatsoft.org`, ISSN: 1548-7660.

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

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

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.

Hoffmann-Jørgensen, J. (1994). *Probability with a view towards statistics*. Chapman & Hall.

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* (eds. S. L. Sutton, T. C. Whitmore and A. C. Chadwick), Blackwell Scientific Publications, 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*. Number 134 in Lecture Notes in Statistics, Springer, New York.

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. (2006). Modelling the effects of partially observed covariates on Poisson process intensity. *Biometrika*, to appear.

Waagepetersen, R. (2006). An estimating function approach to inference for inhomogeneous Neyman-Scott processes. *Biometrics*, to appear.

Waagepetersen, R. (2007). A note on asymptotic results for estimating functions. Available at `www.math.aau.dk/~rw/`.