Estimating functions for inhomogeneous spatial point processes with incomplete covariate data

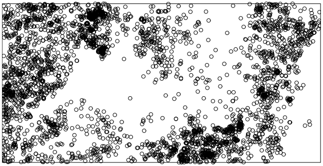
Rasmus Waagepetersen
Department of Mathematics
Aalborg University
Denmark

May 1, 2007

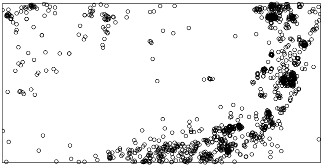
Data (Barro Colorado Island Forest Dynamics Plot)

Observation window: $S = [0, 1000] \times [0, 500]\text{m}^2$

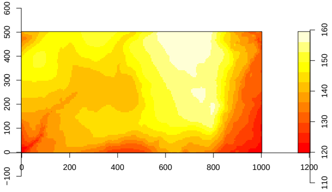
Beilschmiedia



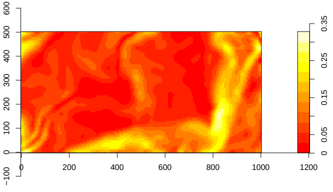
Ocotea



Elevation



Gradient norm (steepness)



Question: tree intensities related to elevation and gradient ?

Outline:

- estimation of log linear parameters for intensity function
- incomplete covariate data: deterministic approx. of score function
- Monte Carlo approx. of score
- distribution of parameter estimates
- numerical examples

Log-linear models for intensity function

$z(u) = \big(z_1(u), \ldots, z_p(u)\big)$ vector of covariates for each location u in observation window W.

E.g. $z(u) = (1, z_{\mathrm{elev}}(u), z_{\mathrm{grad}}(u))$ for rain forest example

Log-linear model for intensity function:

$$\lambda(u; \beta) = \exp(z(u)\beta^{\mathsf{T}})$$

Poisson process case: log likelihood function and derivatives

$\mathbf{x}$ observation of $\mathbf{X}$ Poisson$(W, \lambda(\cdot; \beta))$.

Density wrt. unit rate Poisson process:

$$f(\mathbf{x}; \beta) = \exp(|W| - \int_W \lambda(u; \beta)\mathrm{d}u) \prod_{u \in \mathbf{x}} \lambda(u; \beta)$$

log likelihood function:

$$l(\beta) = \sum_{u \in \mathbf{x}} z(u)\beta^\mathsf{T} - \int_W \lambda(u; \beta)\mathrm{d}u$$

Score function and Fisher information:

$$u(\beta) = \sum_{u \in \mathbf{x}} z(u) - \int_W z(u)\lambda(u; \beta)\mathrm{d}u \quad j(\beta) = \int_W z(u)^\mathsf{T} z(u)\lambda(u; \beta)\mathrm{d}u$$

$$\hat{\beta} \approx N(\beta, V) \quad V = j(\beta)^{-1}$$

Case of second-order reweighted stationary point processes

Solution of

$$u(\beta) = \sum_{u \in \mathbf{x}} z(u) - \int_W z(u)\lambda(u;\beta)\mathrm{d}u = 0$$

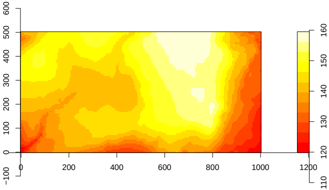
yields asymptotically normal estimate of β for wide class of non-Poisson processes satisfying certain mixing conditions.

Examples: second order reweighted stationary Poisson cluster processes and log Gaussian Cox processes (Waagepetersen, 2007, Guan and Loh, 2007).

(second order reweighted stationary: translation invariant pair correlation function).

Missing covariate data

Elevation covariate



interpolated from elevation observations on grid.

However, evaluating

$$u(\beta) = \sum_{u \in \mathbf{x}} z(u) - \int_W z(u)\lambda(u;\beta)\mathrm{d}u$$

requires $z(u)$ *observed* for any $u \in W$!

Approximations of log likelihood I

Suppose $z(u)$ observed at finite set of locations $\mathbf{Q} \subset W$.

Rathbun (1996) approximate

$$\int_W z(u)\lambda(u;\beta)\mathrm{d}u \approx \int_W \widehat{z(u)\lambda(u;}\beta)\mathrm{d}u$$

where $\widehat{z(u)\lambda(u;}\beta)$ unbiased prediction of $z(u)\lambda(u;\beta)$, $u \in W$ given $z(u)$, $u \in \mathbf{Q}$.

Approximations of log likelihood I

Suppose $z(u)$ observed at finite set of locations $\mathbf{Q} \subset W$.

Rathbun (1996) approximate

$$\int_W z(u)\lambda(u;\beta)\mathrm{d}u \approx \int_W \widehat{z(u)\lambda(u;\beta)}\mathrm{d}u$$

where $\widehat{z(u)\lambda(u;\beta)}$ unbiased prediction of $z(u)\lambda(u;\beta)$, $u \in W$ given $z(u)$, $u \in \mathbf{Q}$.

Riemann approximation:

$$\int_W z(u)\lambda(u;\beta)\mathrm{d}u \approx \sum_{u\in\mathbf{Q}} w(u)z(u)\lambda(u;\beta)$$

where $w(u)$ quadrature weight for $u \in \mathbf{Q}$.

Approximations of log likelihood II: spatstat

Approximation of score function used in R package spatstat (Baddeley and Turner)

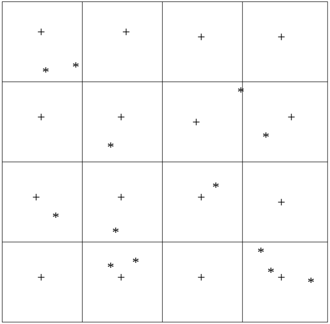
$$u(\beta) \approx u^{\mathsf{spat}}(\beta) = \sum_{u \in \mathbf{X}} z(u) - \sum_{u \in \mathbf{Q}} z(u)\lambda(u;\beta)w(u)$$

but now $\mathbf{Q} = \mathbf{X} \cup \mathbf{D}$ *includes observed points* in addition to 'dummy' points $\mathbf{D}$.

Two types of weights: `grid` or `dirichlet`

Quadrature schemes in spatstat

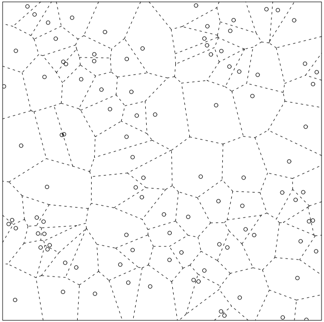
Grid



$w(u) = \frac{|C_v|}{\#(\mathbf{X} \cap C_v)+1}, \; u \in C_v$

where $W = \cup_{v \in \mathbf{D}} C_v$

Dirichlet



$w(u)$ area of *Dirichlet cell* for u in Dirichlet tesselation generated by $\mathbf{Q}$.

spatstat: relation to generalized linear models and iterative weighted least squares

Z: matrix with rows $z(u)$, $u \in \mathbf{X} \cup \mathbf{D}$

A: diagonal matrix with diagonal entries $w(u)\lambda(u;\beta)$, $u \in \mathbf{X} \cup \mathbf{D}$.

$$u^{\text{spat}}(\beta) = Z^{\mathsf{T}}\big(1[u \in x] - w(u)\lambda(u;\beta)\big)_{u \in \mathbf{X} \cup \mathbf{D}} = Z^{T} A y$$

where $y = A^{-1}\big(1[u \in x] - w(u)\lambda(u;\beta)\big)_{u \in \mathbf{X} \cup \mathbf{D}}$.

Derivative

$$\frac{\mathrm{d}}{\mathrm{d}\beta} u^{\text{spat}}(\beta) = Z^{\mathsf{T}} A Z$$

Newton-Raphson iterations

$$Z^{\mathsf{T}} A Z(\beta^{m+1} - \beta^{m}) = Z^{T} A y$$

equivalent to iterative weighted least squares (Berman and Turner, 1992) - hence implementation straightforward using e.g. `glm()` in `R`.

Distribution of parameter estimates from approximate score functions

?

Problem: hard to obtain distribution of parameter estimates from approximate score functions

Monte Carlo approximation of integral

Consider M *random* uniform dummy points on W (binomial point process $\mathbf{D}$ of intensity $M/|W|$).

Rathbun et al. (2006): Monte Carlo approx. of integral:

$$u^{rath}(\beta) = \sum_{u \in \mathbf{X}} z(u) - \sum_{u \in \mathbf{D}} \frac{z(u)\lambda(u;\beta)}{M/|W|}$$

CLT for Monte Carlo approximation:

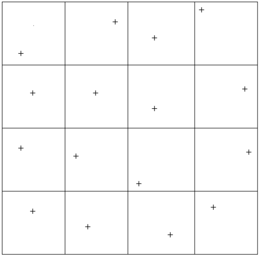
$$n^{1/2}\Big[\sum_{u \in \mathbf{D}_n} \frac{f(u)}{n\rho} - \int_W f(u)\mathrm{d}u\Big] \xrightarrow{d} N(0, G_f/\rho)$$

where

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

Stratified dummy points

Alternative: one uniformly sampled dummy point in each cell (stratified dummy points)



Suppose covariates continuously differentiable. Then CLT

$$n^{1/2}\left[\sum_{u\in\mathbf{D}_n}\frac{f(u)}{n^{1/2}\rho}-\int_W f(u)\mathrm{d}u\right]\xrightarrow{d} N(0,G_f/\rho^2)$$

where $n^{1/2}\rho$ increasing intensity of dummy point process $\mathbf{D}_n$ and

$$G_f=\frac{1}{12}\int_W A_f(u)\mathrm{d}u \quad A_f(u_1,u_2)=\left[\frac{\partial f_i}{\partial u_1}\frac{\partial f_j}{\partial u_1}+\frac{\partial f_i}{\partial u_2}\frac{\partial f_j}{\partial u_2}\right]$$

(faster rate of convergence).

Distribution of parameter estimate

Consider increasing intensity asymptotics: intensities

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

for observed $\mathbf{X}_n$ and dummy $\mathbf{D}_n$ ($k = 1$ (bin.) or $1/2$ (strat.))
ρ: controls proportion of dummy points.

Distribution of parameter estimate

Consider increasing intensity asymptotics: intensities

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

for observed $\mathbf{X}_n$ and dummy $\mathbf{D}_n$ ($k = 1$ (bin.) or $1/2$ (strat.))
ρ: controls proportion of dummy points.

$$u_n^{rath}(\beta) = \sum_{u \in \mathbf{X}_n} z(u) - \sum_{u \in \mathbf{D}_n} \frac{z(u)\lambda(u;\beta)}{n^k \rho} =$$

$$u_n(\beta) + n\Big[\int_W z(u)\lambda(u;\beta)\mathrm{d}u - \sum_{u \in \mathbf{D}_n} \frac{z(u)\lambda(u;\beta)}{n^k \rho}\Big]$$

Distribution of parameter estimate

Consider increasing intensity asymptotics: intensities

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

for observed $\mathbf{X}_n$ and dummy $\mathbf{D}_n$ ($k = 1$ (bin.) or $1/2$ (strat.))
ρ: controls proportion of dummy points.

$$u_n^{rath}(\beta) = \sum_{u \in \mathbf{X}_n} z(u) - \sum_{u \in \mathbf{D}_n} \frac{z(u)\lambda(u;\beta)}{n^k \rho} =$$

$$u_n(\beta) + n\Big[\int_W z(u)\lambda(u;\beta)\mathrm{d}u - \sum_{u \in \mathbf{D}_n} \frac{z(u)\lambda(u;\beta)}{n^k \rho}\Big]$$

Note $\mathbf{X}_n \sim \cup_{i=1}^n \mathbf{X}^i$ where $\mathbf{X}^i$ *iid* Poisson processes $\lambda(u;\beta) \Rightarrow$ CLT.

Distribution of parameter estimate

Consider increasing intensity asymptotics: intensities

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

for observed $\mathbf{X}_n$ and dummy $\mathbf{D}_n$ ($k = 1$ (bin.) or $1/2$ (strat.))
ρ: controls proportion of dummy points.

$$u_n^{rath}(\beta) = \sum_{u \in \mathbf{X}_n} z(u) - \sum_{u \in \mathbf{D}_n} \frac{z(u)\lambda(u;\beta)}{n^k\rho} =$$
$$u_n(\beta) + n\Big[\int_W z(u)\lambda(u;\beta)\mathrm{d}u - \sum_{u \in \mathbf{D}_n} \frac{z(u)\lambda(u;\beta)}{n^k\rho}\Big]$$

Note $\mathbf{X}_n \sim \cup_{i=1}^n \mathbf{X}^i$ where $\mathbf{X}^i$ *iid* Poisson processes $\lambda(u;\beta) \Rightarrow$ CLT.
Hence,

$$n^{-1/2}u_n^{rath}(\beta) \xrightarrow{d} N(0, j(\beta) + G_k/\rho^{1/k}),$$

$$n^{1/2}(\hat{\beta}_n - \beta) \xrightarrow{d} N(0,V + VG_kV/\rho^{1/k}) \quad V = j(\beta)^{-1}$$

Monte Carlo versions of `spatstat` (Waagepetersen, 2007)

D point process of dummy points of intensity ρ. Monte Carlo version of `dirichlet`

$$u^{\mathsf{dir}}(\beta) = \sum_{u \in \mathbf{X}} z(u) - \sum_{u \in \mathbf{X} \cup \mathbf{D}} z(u)\lambda(u;\beta)\frac{1}{\lambda(u;\beta)+\rho}$$

(either binomial or stratified dummy points)

Monte Carlo versions of spatstat (Waagepetersen, 2007)

D point process of dummy points of intensity ρ. Monte Carlo version of `dirichlet`

$$u^{\text{dir}}(\beta) = \sum_{u \in \mathbf{X}} z(u) - \sum_{u \in \mathbf{X} \cup \mathbf{D}} z(u)\lambda(u;\beta)\frac{1}{\lambda(u;\beta) + \rho}$$

(either binomial or stratified dummy points)

Monte Carlo version of `grid`:

$$u^{\text{grid}}(\beta) = \sum_{u \in \mathbf{X}} z(u) - \sum_{u \in \mathbf{X} \cup \mathbf{D}} z(u)\lambda(u;\beta)\frac{1}{\rho(\#(\mathbf{X} \cap C_{v(u)}) + 1)}$$

($v(u) = v$ if $u \in C_v$, only stratified dummy points)

Advantage: implementation (IWLS) just as for usual `spatstat` except that weights depend on β for `dirichlet`.

Asymptotic distribution of parameter estimates I (Poisson process case)

Grid version

$$u^{\text{grid}}(\beta) = \sum_{u \in \mathbf{X}} z(u) - \sum_{v \in \mathbf{D}} \frac{z(u)\lambda(u;\beta) + \sum_{u \in \mathbf{X} \cap C_v} z(v)\lambda(v;\beta)}{\rho(\#(\mathbf{X} \cap C_v) + 1)}$$

Hence assuming continuously differentiable covariates

$$n^{-1/2} u_n^{\text{grid}}(\beta) \sim n^{-1/2} u_n^{\text{rath}}(\beta) \quad n \to \infty$$

and asymptotic covariance matrix becomes

$$V + V G_{1/2} V / \rho^2$$

where $g(u) = z(u)\lambda(u;\beta)$. Tends to MLE asymp. cov. V if $\rho \to \infty$.

Asymptotic distribution of parameter estimates II

Dirichlet estimating function:

$$u_n^{\text{dir}}(\beta) = \sum_{u \in \mathbf{X}_n} z(u) - \sum_{u \in \mathbf{X}_n \cup \mathbf{D}_n} z(u) \frac{\lambda(u;\beta)}{\lambda(u;\beta) + n^{k-1}\rho}$$

$$\sum_{u \in \mathbf{X}_n} z(u) \left(\frac{n^{k-1}\rho}{\lambda(u;\beta) + n^{k-1}\rho} \right) - \sum_{u \in \mathbf{D}_n} z(u) \frac{\lambda(u;\beta)}{\lambda(u;\beta) + n^{k-1}\rho}$$

Asymptotic distribution of parameter estimates II

Dirichlet estimating function:

$$u_n^{\mathrm{dir}}(\beta) = \sum_{u \in \mathbf{X}_n} z(u) - \sum_{u \in \mathbf{X}_n \cup \mathbf{D}_n} z(u) \frac{\lambda(u;\beta)}{\lambda(u;\beta) + n^{k-1}\rho}$$

$$\sum_{u \in \mathbf{X}_n} z(u) \left(\frac{n^{k-1}\rho}{\lambda(u;\beta) + n^{k-1}\rho} \right) - \sum_{u \in \mathbf{D}_n} z(u) \frac{\lambda(u;\beta)}{\lambda(u;\beta) + n^{k-1}\rho}$$

$$n^{-k+1/2} u_n^{\mathrm{dir}}(\beta) \xrightarrow{d} N(0, \rho^2 C_k + \rho^{2-1/k} H_k)$$

$$n^{-k} j_n^{\mathrm{dir}}(\beta) = -n^{-k} \frac{\mathrm{d}}{\mathrm{d}\beta} u_n^{\mathrm{dir}}(\beta) \xrightarrow{p} \rho F_k$$

Asymptotic distribution of parameter estimates II

Dirichlet estimating function:

$$u_n^{\mathrm{dir}}(\beta) = \sum_{u \in \mathbf{X}_n} z(u) - \sum_{u \in \mathbf{X}_n \cup \mathbf{D}_n} z(u) \frac{\lambda(u;\beta)}{\lambda(u;\beta) + n^{k-1}\rho}$$

$$\sum_{u \in \mathbf{X}_n} z(u) \left(\frac{n^{k-1}\rho}{\lambda(u;\beta) + n^{k-1}\rho} \right) - \sum_{u \in \mathbf{D}_n} z(u) \frac{\lambda(u;\beta)}{\lambda(u;\beta) + n^{k-1}\rho}$$

$$n^{-k+1/2} u_n^{\mathrm{dir}}(\beta) \xrightarrow{d} N(0, \rho^2 C_k + \rho^{2-1/k} H_k)$$

$$n^{-k} j_n^{\mathrm{dir}}(\beta) = -n^{-k} \frac{\mathrm{d}}{\mathrm{d}\beta} u_n^{\mathrm{dir}}(\beta) \xrightarrow{p} \rho F_k$$

Hence

$$n^{1/2}(\hat{\beta}_n - \beta) \xrightarrow{d} N(0, F_k^{-1} C_k F_k^{-1} + F_k^{-1} H_k F_k^{-1} / \rho^{1/k})$$

Case $k = 1/2$: $F_k^{-1} C_k F^{-1}$ differs from V even when $\rho \to \infty$!

Cluster processes

Consider cluster process $\mathbf{X} = \mathbf{X}_{c \in \mathbf{Y}}$ where $\mathbf{Y}$ stationary Poisson point process of intensity $\kappa > 0$.

Given $\mathbf{Y}$, clusters $\mathbf{X}_c$ are independent Poisson processes with intensity functions

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

Intensity function of $\mathbf{X}$ is then of log-linear form $\exp(z(u)\beta^{\mathsf{T}})$ where $\beta_1 = \log(\kappa\alpha)$ and $z_1(u) = 1$.

Estimates $\hat{\beta}_n$ using Monte Carlo `spatstat` still asymptotically normal (consider increasing 'mother intensity' $\kappa_n = n\kappa$).

Asymptotic covariance=Asymptotic covariance for Poisson + extra term due to clustering.

Numerical example: Poisson process

Case of Poisson process with covariate vector $(1, z_{\text{elev}})$ $\beta_{\text{elev}} = 0.1$.

Ratios of asymptotic standard errors for estm. funct. estimate $\hat{\beta}_{\text{elev}}$ relative to asymptotic standard error for MLE.

	bin.				str.			
Est. fct. \ q	0.25	1	10	100	0.25	1	10	100
u^{rath} (u^{grid})	2.47	1.51	1.06	1.01	1.08	1.01	1.00	1.00
u^{dir}	2.12	1.43	1.06	1.01	1.56	1.53	1.53	1.53

$q = \#\mathbf{D}/\#(\mathbf{X} \cup \mathbf{D})$ proportion of dummy points.

u^{dir} slightly better than Rathbuns Monte Carlo approximation u^{rath} but not useful in case of stratified dummy points.

Numerical example: Clustered point process

Ratios of standard errors for $\hat{\beta}_{\text{elev}}$ (relative to complete covariate data case) in case of clustered point process with varying values of 'mother' intensity and varying numbers M of dummy points.

	κ	8e-5			8e-4			8e-3	
	M	450	800	1800	450	800	1800	450	800
u^{rath}	(bin.)	1.06	1.03	1.01	1.44	1.26	1.12	2.49	1.98
u^{dir}	(bin.)	1.01	0.99	0.97	1.35	1.20	1.08	2.32	1.86
u^{grid}	(str.)	1.00	1.00	1.00	1.04	1.01	1.00	1.17	1.06

In highly clustered case $\kappa = 8\text{e-}5$ not big loss of efficiency due to missing covariate data: term due to clustering V_{clust} dominating in asymptotic covariance matrix

$$V + V_{\text{incompl}} + V_{\text{clust}}$$

Perspective

- methodology available for handling missing covariate
- implementation straightforward
- random sampling schemes required
- need to take this into account in future experiments

References:

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

Waagepetersen, R. (2007) Estimating functions for inhomogeneous spatial point processes with incomplete covariate data, submitted.