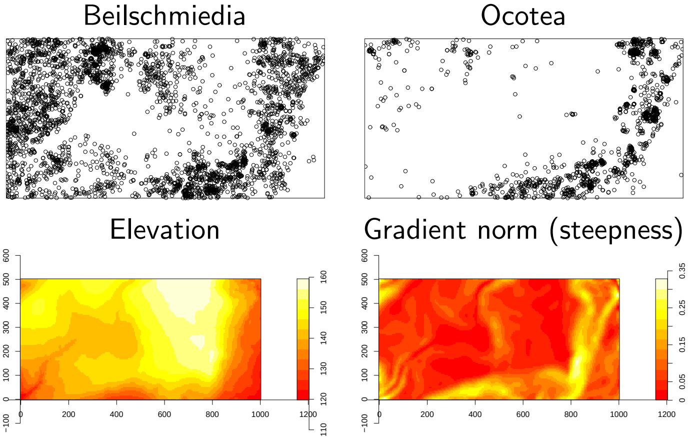
Estimating functions for inhomogeneous spatial point processes with incomplete covariate data

Rasmus Waagepetersen
Department of Mathematics
Aalborg University
Denmark

June 6, 2007

Data (Barro Colorado Island Forest Dynamics Plot)

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



Question: tree intensities related to elevation and gradient ?

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

Log-linear model for intensity function:

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

Interpretation:

$$\lambda(u; \beta)\mathrm{d}N_u = P(\text{point occurs in neighbourhood } N_u \text{ around } u)$$

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:

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

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:

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

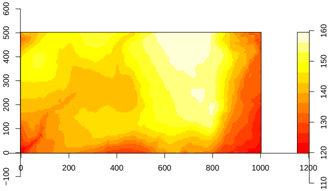
Score function and Fisher information:

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

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

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)

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

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

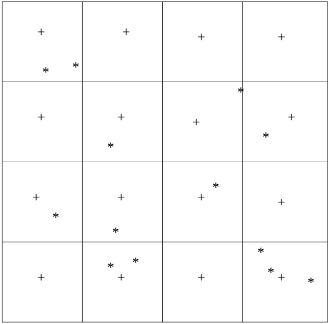
Note: $\mathbf{X}$ events of point pattern (supplied by 'nature')

$\mathbf{D}$ dummy points controlled by scientist.

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

Quadrature schemes in spatstat

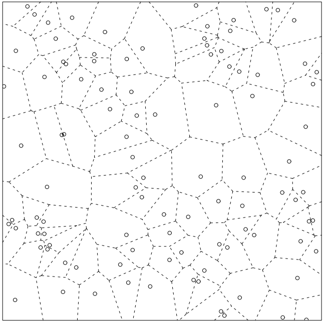
Grid



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

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

Estimating function

$$u^{\text{spat}}(\beta) = \sum_{u \in \mathbf{X}} z(u) - \sum_{u \in \mathbf{X} \cup \mathbf{D}} w(u) z(u) \lambda(u; \beta)$$

formally equivalent to score function of weighted Poisson regression (log link):

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

with weights $w(u)$ and 'observations' $y_u = 1[u \in \mathbf{X}]/w(u)$.

Hence implementation straightforward using e.g. `glm()` in `R` (Poisson family, log link).

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 (*simple* random dummy points). Assume wlog $|W| = 1$.

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

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

CLT for Monte Carlo approximation:

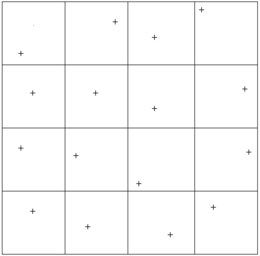
$$M^{1/2}\Big[\sum_{u \in \mathbf{D}_n} \frac{f(u)}{M} - \int_W f(u)\mathrm{d}u\Big] \xrightarrow{d} N(0, G_f)$$

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 f continuously differentiable. Then CLT

$$M\left[\sum_{u\in\mathbf{D}_n}\frac{f(u)}{M}-\int_W f(u)\mathrm{d}u\right]\xrightarrow{d} N(0,G_f)$$

where M increasing number of dummy points 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

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

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

Distribution of parameter estimate

Consider increasing intensity asymptotics: intensities

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

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

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

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

Distribution of parameter estimate

Consider increasing intensity asymptotics: intensities

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

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

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

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

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

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

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

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

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

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,

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

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

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

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

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

(either simple 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`

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

(either simple or stratified dummy points)

Monte Carlo version of `grid`: (stratified dummy points)

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

Advantage: implementation using `glm()` just as for usual `spatstat`.

Asymptotic distribution of parameter estimates I

Grid version ($k = 1/2$)

$$u_n^{\text{grid}}(\beta) = \sum_{u \in \mathbf{X}_n} z(u) - \sum_{v \in \mathbf{D}_n} \frac{n}{n^{1/2}\rho} \frac{z(v)\lambda(v;\beta) + \sum_{u \in \mathbf{X} \cap C_{v,n}} z(u)\lambda(u;\beta)}{\#(\mathbf{X} \cap C_{v,n}) + 1}$$

$$\approx u_n^{rath}(\beta) = \sum_{u \in \mathbf{X}_n} z(u) - \sum_{u \in \mathbf{D}_n} \frac{z(u) n \lambda(u;\beta)}{n^{1/2}\rho}$$

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

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

Tends to MLE asymp. cov. V if $\rho \to \infty$.

Asymptotic distribution of parameter estimates II

Dirichlet estimating function:

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

Asymptotic distribution of parameter estimates II

Dirichlet estimating function:

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

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

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

Note: normalizing sequences $n^{-k+1/2}$ and n^{-k} depend on k.

Asymptotic distribution of parameter estimates II

Dirichlet estimating function:

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

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

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

Note: normalizing sequences $n^{-k+1/2}$ and n^{-k} depend on 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 $\rho \to \infty$!

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.

	simple					stratified			
q	0.25	1	10	100		0.25	1	10	100
u^{rath}	2.47	1.51	1.06	1.01	u^{grid}	1.08	1.01	1.00	1.00
u^{dir}	2.12	1.43	1.06	1.01	u^{dir}	1.56	1.53	1.53	1.53

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

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

Perspective

- methodology available for handling missing covariate data
- implementation straightforward
- also works for cluster point processes
- random sampling schemes required

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.