Estimating functions for inhomogeneous spatial point processes with incomplete covariate data

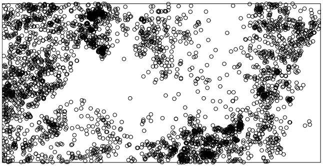
Rasmus Waagepetersen
Department of Mathematics
Aalborg University
Denmark

August 15, 2007

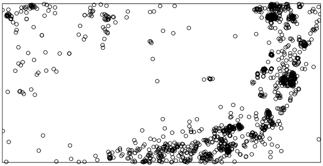
Data (Barro Colorado Island Forest Dynamics Plot)

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

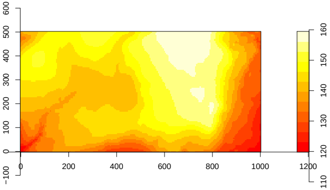
Beilschmiedia



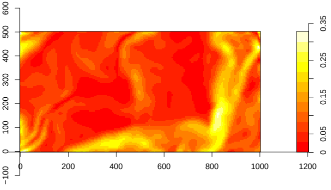
Ocotea



Elevation



Gradient norm (steepness)



Question: tree intensities related to elevation and gradient ?

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:

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

Interpretation:

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

Estimation of β

$\mathbf{X}$: point pattern observed in W

Poisson process log likelihood function:

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

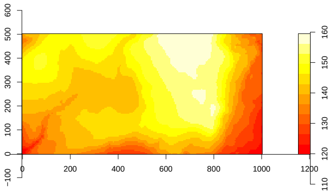
Poisson score estimating function:

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

Also applicable for *non-Poisson* point processes with intensity function $\lambda(\cdot;\beta)$ (Schoenberg, 2005, Waagepetersen, 2007)

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 $\mathbf{D}$.

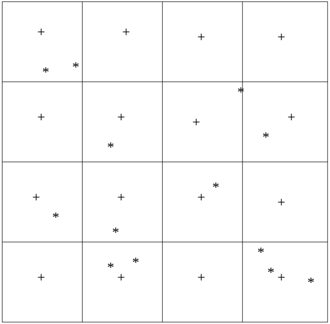
Note: $\mathbf{X}$ events of point pattern (supplied by 'nature')

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

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

Quadrature schemes in spatstat

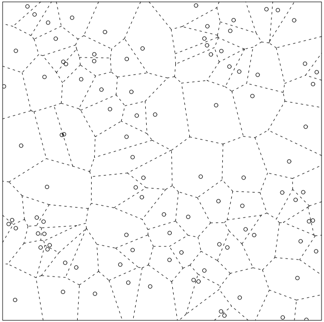
Grid



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

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):

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

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 $\mathbf{D}$ on W (*simple* random dummy points). Assume wlog $|W| = 1$.

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

$$
\begin{aligned}
u^{rath}(\beta) &= \sum_{u \in \mathbf{X}} z(u) - \frac{1}{M} \sum_{u \in \mathbf{D}} z(u)\lambda(u;\beta) \\
&= u(\beta) + \Big[\frac{1}{M} \sum_{u \in \mathbf{D}} z(u)\lambda(u;\beta) - \int_W z(u)\lambda(u;\beta)\mathrm{d}u\Big]
\end{aligned}
$$

Monte Carlo approximation of integral

Consider M *random* uniform dummy points $\mathbf{D}$ on W (*simple* random dummy points). Assume wlog $|W| = 1$.

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

$$
\begin{aligned}
u^{rath}(\beta) &= \sum_{u \in \mathbf{X}} z(u) - \frac{1}{M} \sum_{u \in \mathbf{D}} z(u)\lambda(u;\beta) \\
&= u(\beta) + \Big[\frac{1}{M} \sum_{u \in \mathbf{D}} z(u)\lambda(u;\beta) - \int_W z(u)\lambda(u;\beta)\mathrm{d}u\Big]
\end{aligned}
$$

CLT for Monte Carlo approximation:

$$
M^{1/2}\Big[\frac{1}{M} \sum_{u \in \mathbf{D}} z(u)\lambda(u;\beta) - \int_W z(u)\lambda(u;\beta)\mathrm{d}u\Big] \xrightarrow{d} N(0, G)
$$

where G asymptotic covariance matrix.

Distribution of parameter estimate (Poisson process case)

Poisson score asymptotically normal

$$u(\beta) \approx N\big(0, i(\beta)\big)$$

where $i(\beta)$ Fisher information.

Suppose $\hat{\beta}$ solution of $u^{rath}(\beta) = 0$ where

$$u^{rath}(\beta) = u(\beta) + \Big[\frac{1}{M}\sum_{u \in \mathbf{D}} z(u)\lambda(u;\beta) - \int_W z(u)\lambda(u;\beta)\mathrm{d}u\Big]$$

Distribution of parameter estimate (Poisson process case)

Poisson score asymptotically normal

$$u(\beta) \approx N\big(0, i(\beta)\big)$$

where $i(\beta)$ Fisher information.

Suppose $\hat{\beta}$ solution of $u^{rath}(\beta) = 0$ where

$$u^{rath}(\beta) = u(\beta) + \Big[\frac{1}{M}\sum_{u \in \mathbf{D}} z(u)\lambda(u;\beta) - \int_W z(u)\lambda(u;\beta)\mathrm{d}u\Big]$$

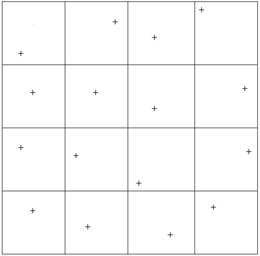
Then

$$\hat{\beta} \approx N\big(\beta, i(\beta)^{-1} + i(\beta)^{-1}\ G\ i(\beta)^{-1}\big)$$

NB: increasing intensity asymptotics: $N = \mathbb{E}\#\mathbf{X} \rightarrow \infty$! (fixed observation window W)

Stratified dummy points

Alternative: one uniformly sampled dummy point in each of M cells (stratified dummy points)



Suppose f continuously differentiable. Then CLT

$$M\left[\frac{1}{M}\sum_{u\in\mathbf{D}} f(u) - \int_W f(u)\mathrm{d}u\right] \xrightarrow{d} N(0, G)$$

(faster rate of convergence).

Asymptotic covariance matrix G depends on partial derivatives $\frac{\partial f_i}{\partial u_j}$ of f.

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

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

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

(either simple or stratified dummy points)

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

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

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

(either simple or stratified dummy points)

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

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

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

Distribution of parameter estimates (Poisson process case)

Grid: $u^{\text{grid}}(\beta)$ and $u^{\text{rath}}(\beta)$ with stratified dummy points same asymptotic distribution - hence same asymptotic distribution for parameter estimates.

Dirichlet: more subtle behaviour

- stratified dummy points: $u^{\text{dir}}(\beta)$ less efficient than $u^{\text{grid}}(\beta)/u^{\text{rath}}(\beta)$
- simple dummy points: asymptotic covariance matrix tends to that of MLE when $q = M/N$ increases ($N = \mathbb{E}\#\mathbf{X}$).

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 estimating function 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

$q = M/N$ proportion of dummy points.

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

Clustered point processes

Consider cluster process $\mathbf{X} = \cup_{c \in \mathbf{Y}} \mathbf{X}_c$ 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

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

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

Estimate $\hat{\beta}$ using $u(\beta)$, $u^{\mathrm{dir}}(\beta)$ or $u^{\mathrm{grid}}(\beta)$ still asymptotically normal but covariance matrix now depends on $\psi = (\kappa, \omega)$.

Minimum contrast estimation of ψ

Closed form expression $K(t;\psi)$ and estimate

$$\hat{K}(t) = \sum_{u,v \in \mathbf{X}}^{\neq} \frac{1[\|v-u\| \leq t]}{\lambda(u;\hat{\beta})\lambda(v;\hat{\beta})|W \cap W_{v-u}|}$$

available for K-function (extension to inhomogeneous case).

Minimum contrast estimation of ψ

Closed form expression $K(t;\psi)$ and estimate

$$\hat{K}(t) = \sum_{u,v \in \mathbf{X}}^{\neq} \frac{1[\|v-u\| \leq t]}{\lambda(u;\hat{\beta})\lambda(v;\hat{\beta})|W \cap W_{v-u}|}$$

available for K-function (extension to inhomogeneous case).

Second step for estimation of ψ:

$$\hat{\psi} = \arg\min_{\psi} \int_0^r [\hat{K}(t)^c - K(t;\psi)^c]^2 \mathrm{d}t$$

NB: only requires $z(u)$ for $u \in \mathbf{X}$.

Inference regarding clustering

So far focused on intensity parameter β.

Sometimes clustering of main interest but still need to estimate β to adjust for aggregation due to covariates.

Joint asymptotic normality of $(\hat{\beta}, \hat{\psi})$ established in W & Guan (2007) for wide class of mixing point processes including

- cluster processes
- log Gaussian Cox processes (ψ covariance parameter of Gaussian field).

References

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

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.

Waagepetersen, R. and Guan, Y. (2007). Two-step estimation for inhomogeneous spatial point processes, in preparation.