Estimation of the pair correlation function for an inhomogeneous spatial point process using a baseline point process

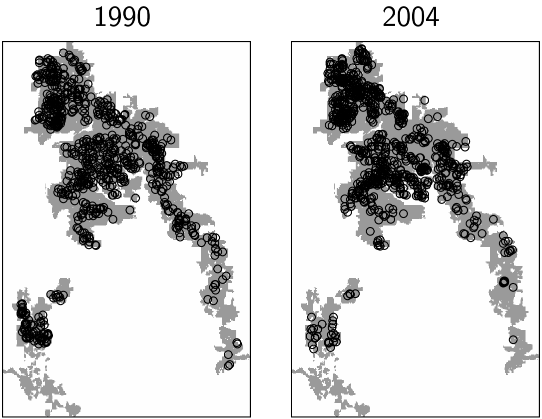
Rasmus Waagepetersen
Department of Mathematical Sciences
Aalborg University

Joint work with

Yongtao Guan
Division of Biostatistics
Yale University

May 17, 2010

Data example: golden plover birds in the Peak District



Climate change of spatial distribution of bird locations ?

Very irregular observation window W=grey area =Peak District.

Beale, Pearce-Higgins, Bonn and Illian (2008) modelled 2004 pattern using an inhomogeneous Markov point process with offset given by non-parametric estimate of intensity for 1990 pattern and topography/vegetation covariates.

Conditional likelihood

Intensity $\rho_1(\cdot)$ for 1990 baseline intensity for 2004 data:

$$\rho_2(\mathbf{u}) = \rho_1(\mathbf{u}) f(\mathbf{u}; \beta)$$

Changes in intensity function modelled via regression term $f(\mathbf{u}; \beta)$ depending on $\beta \mathbf{Z}(u)^{\mathsf{T}}$ where $\mathbf{Z}(u)$ is a covariate vector.

Avoid estimation of baseline intensity using conditional likelihood: condition on $\mathbf{X} = \mathbf{X}_1 \cup \mathbf{X}_2$. Probabibility that $u \in \mathbf{X}$ comes from $\mathbf{X}_1$ is

$$p(\mathbf{u}; \beta) = \frac{\rho_1(\mathbf{u})}{\rho_1(\mathbf{u}) + \rho_2(\mathbf{u})} = \frac{1}{1 + f(\mathbf{u}; \beta)}$$

Conditional likelihood (Diggle and Rowlingson, 1994):

$$l(\beta) = \sum_{\mathbf{u} \in \mathbf{X}_1 \cap W} \log[p(\mathbf{u}; \beta)] + \sum_{\mathbf{v} \in \mathbf{X}_2 \cap W} \log[1 - p(\mathbf{v}; \beta)]$$

Second-order properties

$\mathbb{V}\mathrm{ar}\hat{\beta}$ depends on second-order properties of $\mathbf{X}_1$ and $\mathbf{X}_2$.

It may be OK to model $\mathbf{X}_1$ as inhomogeneous Poisson process since no restrictions on ρ_1.

May need to account for clustering in $\mathbf{X}_2$ not explained by ρ_1 and $f(\cdot;\beta)$.

Assume $\mathbf{X}_2$ second-order reweighted stationary. Then second-order moments determined by translation invariant pair correlation function $g(\cdot)$:

$$\mathbb{E} \sum_{\mathbf{u},\mathbf{v} \in \mathbf{X}_2 \cap W}^{\neq} h(u,v) = \int_{W^2} h(u,v)\rho_2(\mathbf{u})\rho_2(\mathbf{v})g(\mathbf{u}-\mathbf{v})\mathrm{d}\mathbf{u}\mathrm{d}\mathbf{v}$$

Kernel density estimate of pair correlation function

$$\hat{g}(t) = \frac{1}{2\pi t|W|} \sum_{\mathbf{u},\mathbf{v} \in \mathbf{X}_2 \cap W}^{\neq} \frac{k_b(\|\mathbf{u} - \mathbf{v}\| - t) e_{\mathbf{u},\mathbf{v}}}{\hat{\rho}_1(\mathbf{u}) f(\mathbf{u}; \hat{\beta}) \hat{\rho}_1(\mathbf{v}) f(\mathbf{v}; \hat{\beta})}$$

where $k_b(\cdot)$ smoothing kernel.

Problem: non-parametric estimate of $\rho_1(\cdot)$ not consistent.

Edge correction $e_{\mathbf{u},\mathbf{v}}$ difficult to calculate for irregular W.

Consistent kernel estimation of pair correlation function

Kernel estimate ($k_b(\cdot)$ smoothing kernel):

$$\hat{g}(t) = \frac{\sum^{\neq}_{\mathbf{u},\mathbf{v}\in\mathbf{X}_2\cap W} k_b(||\mathbf{u}-\mathbf{v}||-t)}{\sum^{\neq}_{\mathbf{u},\mathbf{v}\in\mathbf{X}_1\cap W} f(\mathbf{u};\hat{\beta})f(\mathbf{v};\hat{\beta})k_b(||\mathbf{u}-\mathbf{v}||-t)}$$

Numerator estimate of

$$\int_{W^2} k_b(||\mathbf{u}-\mathbf{v}||-t)\rho_2(\mathbf{u})\rho_2(\mathbf{v})g(\mathbf{u}-\mathbf{v})\mathrm{d}\mathbf{u}\mathrm{d}\mathbf{v} \approx g(t)\int_{W^2} k_b(||\mathbf{u}-\mathbf{v}||-t)\rho_2(\mathbf{u})\rho_2(\mathbf{v})\mathrm{d}\mathbf{u}\mathrm{d}\mathbf{v}$$

and denominator estimate of

$$\int_{W^2} k_b(||\mathbf{u}-\mathbf{v}||-t)\rho_2(\mathbf{u})\rho_2(\mathbf{v})\mathrm{d}\mathbf{u}\mathrm{d}\mathbf{v}$$

Avoids estimation of $\rho_1(\cdot)$

Parametric estimation of pair correlation function

Assume parametric model $g(\cdot;\psi)$.

Again condition on $\mathbf{X} = \mathbf{X}_1 \cup \mathbf{X}_2$. Consider $u \neq v$ in $\mathbf{X}$.

Compute probabilities p_{22}, p_{11} and p_{12} that both points from $\mathbf{X}_2$, both from $\mathbf{X}_1$ and one from $\mathbf{X}_2$ one from $\mathbf{X}_1$.

E.g. probability that $\mathbf{X}_2$ has points at $\mathbf{u}$ and $\mathbf{v}$ is

$$\rho_2(\mathbf{u})\rho_2(\mathbf{v})g(\mathbf{u}-\mathbf{v};\psi)\mathrm{d}\mathbf{u}\mathrm{d}\mathbf{v}$$

Hence conditional probability that $\mathbf{u},\mathbf{v} \in \mathbf{X}_2$ given $\mathbf{u},\mathbf{v} \in \mathbf{X}$ is

$$p_{22}(u,v;\beta,\psi) = \frac{\rho_2(\mathbf{u})\rho_2(\mathbf{v})g(\mathbf{u}-\mathbf{v};\psi)}{\rho_1(\mathbf{u})\rho_1(\mathbf{v}) + \rho_1(\mathbf{v})\rho_2(\mathbf{u}) + \rho_1(\mathbf{u})\rho_2(\mathbf{v}) + \rho_2(\mathbf{u})\rho_2(\mathbf{v})g(\mathbf{u}-\mathbf{v};\psi)} = \frac{f(\mathbf{u};\beta)f(\mathbf{v};\beta)g(\mathbf{u}-\mathbf{v};\psi)}{1 + f(\mathbf{u};\beta) + f(\mathbf{v};\beta) + f(\mathbf{u};\beta)f(\mathbf{v};\beta)g(\mathbf{u}-\mathbf{v};\psi)}$$

($\rho_1(\mathbf{u})\rho_1(\mathbf{v})$ cancels in numerator and denominator)

Conditional pairwise likelihood:

$$
\begin{aligned}
l_2(\beta, \psi) &= \sum_{\mathbf{u},\mathbf{v} \in \mathbf{X}_2 \cap W}^{\neq} \log[p_{22}(\mathbf{u}, \mathbf{v}; \beta, \psi)] + \sum_{\mathbf{u},\mathbf{v} \in \mathbf{X}_1 \cap W}^{\neq} \log[p_{11}(\mathbf{u}, \mathbf{v}; \beta, \psi)] \\
&+ 2 \sum_{\mathbf{u} \in \mathbf{X}_1 \cap W} \sum_{\mathbf{v} \in \mathbf{X}_2 \cap W} \log[p_{12}(\mathbf{u}, \mathbf{v}; \beta, \psi)]
\end{aligned}
$$

Joint maximization with respect to β and ψ numerically unstable.

Hence two-step approach where β estimated from Diggle-Rowlingson conditional likelihood.

Model checking

Smoothed residual process (analogue of residuals in Baddeley *et al.*, 2005)

$$R(\mathbf{s}) = \sum_{\mathbf{u} \in \mathbf{X}_2 \cap W} k(\mathbf{s} - \mathbf{u}) - \sum_{\mathbf{v} \in \mathbf{X}_1 \cap W} k(\mathbf{s} - \mathbf{v}) f(\mathbf{v}; \beta)$$

$R(\mathbf{s})$ has expectation zero:

$$\mathbb{E} \sum_{\mathbf{u} \in \mathbf{X}_2 \cap W} k(\mathbf{s} - \mathbf{u}) = \int_W k(\mathbf{s} - \mathbf{u}) \rho_2(u) \mathrm{d}u =$$

$$\int_W k(\mathbf{s} - \mathbf{u}) \rho_1(u) f(\mathbf{u}; \beta) \mathrm{d}u = \mathbb{E} \sum_{\mathbf{v} \in \mathbf{X}_1 \cap W} k(\mathbf{s} - \mathbf{v}) f(\mathbf{v}; \beta)$$

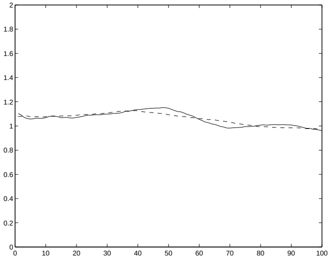
Golden plover data

Covariate vector $\mathbf{Z}(u) = (\mathbf{Z}_1(u), \ldots, \mathbf{Z}_4(u))$: slope, altitude, percent cover of heather, percent cover of cotton grass.

Log linear regression model:

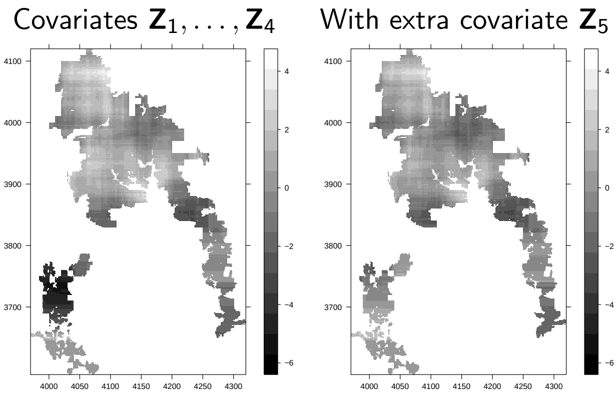
$$f(\mathbf{u}; \beta) = \exp(\beta \mathbf{Z}(\mathbf{u})^\mathsf{T})$$

Kernel estimate $\hat{g}(\cdot)$



Very close to inhomogeneous Poisson process.

Standardized residuals



$\mathbf{Z}_5$ indicator for lower left isolated region.

Significant dependence on slope (+), cotton grass (+) and indicator for lower left region (-).