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.
Conditional likelihood

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

$$\rho_2(u) = \rho_1(u) f(u; \beta)$$

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

Avoid estimation of baseline intensity using conditional likelihood: condition on $X = X_1 \cup X_2$. Probability that $u \in X$ comes from $X_1$ is

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

Conditional likelihood (Diggle and Rowlingson, 1994):

$$l(\beta) = \sum_{u \in X_1 \cap W} \log[p(u; \beta)] + \sum_{v \in X_2 \cap W} \log[1 - p(v; \beta)]$$
Second-order properties

$\text{Var} \hat{\beta}$ depends on second-order properties of $X_1$ and $X_2$.

It may be OK to model $X_1$ as inhomogeneous Poisson process since no restrictions on $\rho_1$.

May need to account for clustering in $X_2$ not explained by $\rho_1$ and $f(\cdot; \beta)$.

Assume $X_2$ second-order reweighted stationary. Then second-order moments determined by translation invariant pair correlation function $g(\cdot)$:

$$\mathbb{E} \sum_{u, v \in X_2 \cap W} \neq h(u, v) = \int_{W^2} h(u, v) \rho_2(u) \rho_2(v) g(u - v) du dv$$
Kernel density estimate of pair correlation function

\[ \hat{g}(t) = \frac{1}{2\pi t|W|} \sum_{u, v \in X_2 \cap W} \frac{k_b(\|u - v\| - t)e_{u, v}}{\hat{\rho}_1(u)f(u; \hat{\beta})\hat{\rho}_1(v)f(v; \hat{\beta})} \]

where \( k_b(\cdot) \) smoothing kernel.

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

Edge correction \( e_{u, v} \) difficult to calculate for irregular \( W \).
Consistent kernel estimation of pair correlation function

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

$$\hat{g}(t) = \frac{\sum_{u, v \in X_2 \cap W} k_b(||u - v|| - t)}{\sum_{u, v \in X_1 \cap W} f(u; \hat{\beta}) f(v; \hat{\beta}) k_b(||u - v|| - t)}$$

Numerator estimate of

$$\int_{W^2} k_b(||u - v|| - t) \rho_2(u) \rho_2(v) g(u - v) du dv \approx g(t) \int_{W^2} k_b(||u - v|| - t) \rho_2(u) \rho_2(v) du dv$$

and denominator estimate of

$$\int_{W^2} k_b(||u - v|| - t) \rho_2(u) \rho_2(v) du dv$$

Avoids estimation of $\rho_1(\cdot)$
Parametric estimation of pair correlation function

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

Again condition on \( X = X_1 \cup X_2 \). Consider \( u \neq v \) in \( X \).

Compute probabilities \( p_{22}, p_{11} \) and \( p_{12} \) that both points from \( X_2 \), both from \( X_1 \) and one from \( X_2 \) one from \( X_1 \).

E.g. probability that \( X_2 \) has points at \( u \) and \( v \) is

\[
\rho_2(u)\rho_2(v)g(u - v; \psi)du dv
\]

Hence conditional probability that \( u, v \in X_2 \) given \( u, v \in X \) is

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

(\( \rho_1(u)\rho_1(v) \) cancels in numerator and denominator).
Conditional pairwise likelihood:

\[
l_2(\beta, \psi) = \sum_{u,v \in X_2 \cap W} \log[p_{22}(u, v; \beta, \psi)] + \sum_{u,v \in X_1 \cap W} \log[p_{11}(u, v; \beta, \psi)] + 2 \sum_{u \in X_1 \cap W} \sum_{v \in X_2 \cap W} \log[p_{12}(u, v; \beta, \psi)]
\]

Joint maximization with respect to \( \beta \) and \( \psi \) numerically unstable.

Hence two-step approach where \( \beta \) estimated from Diggle-Rowlingson conditional likelihood.
Model checking

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

\[ R(s) = \sum_{u \in X_2 \cap W} k(s - u) - \sum_{v \in X_1 \cap W} k(s - v)f(v; \beta) \]

\( R(s) \) has expectation zero:

\[ \mathbb{E} \left( \sum_{u \in X_2 \cap W} k(s - u) \right) = \int_W k(s - u) \rho_2(u) du = \int_W k(s - u) \rho_1(u) f(u; \beta) du = \mathbb{E} \left( \sum_{v \in X_1 \cap W} k(s - v)f(v; \beta) \right) \]
Golden plover data

Covariate vector 
\[ \mathbf{Z}(u) = (Z_1(u), \ldots, Z_4(u)) \]: slope, altitude, percent cover of heather, percent cover of cotton grass.

Log linear regression model:

\[ f(u; \beta) = \exp(\beta \mathbf{Z}(u)^T) \]

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

Very close to inhomogeneous Poisson process.
Standardized residuals

Covariates $Z_1, \ldots, Z_4$

With extra covariate $Z_5$

$Z_5$ indicator for lower left isolated region.

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