Estimation of the pair correlation function of a spatial point process

November 1, 2017

Rasmus Waagepetersen

Department of Mathematical Sciences
Aalborg University
Denmark
Data example: *Acalypha*

- **observation window** $W = 1000 \text{ m} \times 500 \text{ m}$
- **seed dispersal** ⇒ *clustering*
- **environment** ⇒ *inhomogeneity*

Objective: quantify dependence on environmental variables and clustering
Spatial point process

A locally finite random subset $X$ of $\mathbb{R}^2$:

$\#(X \cap A)$ finite for all bounded subsets $A \subset \mathbb{R}^2$

We let $N(A) = \#(X \cap A)$ for bounded $A \subset \mathbb{R}^2$.

Basic quantities: means and covariances

$\mathbb{E}N(A) \quad \text{Cov}[N(A), N(B)]$
Moments of a spatial point process

Fundamental characteristics of point process: mean and covariance of counts $N(A) = \#(X \cap A)$.

*Intensity measure* $\mu$:

$$\mu(A) = \mathbb{E}N(A), \quad A \subseteq \mathbb{R}^2$$

In practice often given in terms of *intensity function*

$$\mu(A) = \int_A \rho(u) \, du$$

$\rho(u) \, du$: probability of observing a point in neighbourhood of $u$. 
Pair correlation function

We further assume that covariances given in terms of intensity function and pair correlation function:

$$\text{Cov}[N(A), N(B)] = \int_{A \cap B} \rho(u) du + \int_{A} \int_{B} \rho(u) \rho(v) (g(u,v) - 1) du dv$$

= Poisson covariance + additional/less covariance due to interaction

Further interpretation of $g$: let $\rho(u|v)$ denote intensity of $X$ given $v \in X$ (‘Palm’ intensity). Then

$$g(u,v) = \frac{\rho(u|v)}{\rho(u)}$$

- i.e. how much is intensity of $u$ changed by presence of point at $v$. 
Campbell formulae

By definition of intensity function and product density and the standard proof we obtain the useful Campbell formulae:

\[ E \sum_{u \in X} h(u) = \int h(u)\rho(u)du \]

\[ E \sum_{u, v \in X} h(u, v) = \iint h(u, v)\rho(u)\rho(v)g(u, v)dudv \]
Examples of pair correlation functions

In this talk we assume $g$ is isotropic $g(v, u) = g(||v - u||)$.

Examples of $g$ for various point process models:
Reasons for estimating $g$:

- key summary of clustering/repulsion properties
- if interest focused on parametric model for intensity function

$$\rho(u; \beta) = \exp(z(u)^T \beta)$$

then $g$ needed to evaluate standard errors of parameter estimates $\hat{\beta}$
Outline of rest of talk

- Kernel estimation of pair correlation function
- Orthogonal series estimation of pair correlation function

(including choice of bandwidth and truncation of orthogonal series !)
Kernel estimate

Suppose $X$ observed in bounded window $W$.

Kernel estimate ($k_b(\cdot)$ kernel with bandwidth $b$):

$$\hat{g}(t) = \frac{1}{2\pi} \sum_{u,v \in X \cap W} \frac{k_b(t - \|v - u\|)}{\|v - u\| \rho(u) \rho(v) \|W \cap W_{v-u}\|}, \quad t > 0$$

Related to non-parametric estimation of probability densities for “distance observations” $D_{uv} = \|v - u\|$, $v, u \in X$.

However $D_{uv}$’s neither independent nor identically distributed. Need to correct for inhomogeneous intensity and edge effects.
Application of Campbell

\[ \mathbb{E}\hat{g}(t) \]

\[ = \frac{1}{2\pi} \mathbb{E} \sum_{u, v \in X \cap W} \frac{k_b(t - \|v - u\|)}{\|v - u\| \rho(u) \rho(v) \|W \cap W_{v-u}\|} \]

\[ = \frac{1}{2\pi} \int_{\mathbb{R}^2} \int_{\mathbb{R}^2} 1[u \in W, v \in W] \frac{k_b(t - \|v - u\|) \rho(u) \rho(v) g(\|v - u\|)}{\|v - u\| \rho(u) \rho(v) \|W \cap W_{v-u}\|} \, du \, dv \]

\[ = \frac{1}{2\pi} \int_{\mathbb{R}^2} \int_{\mathbb{R}^2} 1[u \in W \cap W_{-h}] \frac{k_b(t - \|h\|)}{\|h\| \|W \cap W_{-h}\|} g(\|h\|) \, du \, dh \]

\[ = \frac{1}{2\pi} \int_{\mathbb{R}^2} \frac{k_b(t - \|h\|)}{\|h\|} g(\|h\|) \, du \, dh \]

\[ = \int_{0}^{\infty} k_b(t - r) g(r) \, dr \approx g(t) \]
Choice of band width - minimize MISE

Mean integrated squared error:

\[ \text{MISE}(\hat{g}(\cdot; b)) = \int_0^R \mathbb{E}[\hat{g}(r; b) - g(r)]^2 r \, dr \]

Minimize ‘leave-one-out’ estimate of MISE (Guan 2007):

\[ \text{MISE}(\hat{g}(\cdot; b)) \approx \int_0^R \hat{g}(r; b)^2 r \, dr - 2 \sum_{u,v \in X_W}^{\neq} \frac{\hat{g}^{-\{u,v\}}(\|v - u\|; b)}{2\pi \rho(u) \rho(v)|W \cap W_{v-u}|} \]

Here \( \hat{g}^{-\{u,v\}}(\cdot; b) \) is leave one pair out estimate based on \( X \setminus \{u, v\} \) and using band width \( b \).
Fast estimate of MISE

Computation of $\hat{g}^{-\{u,v\}}$: need to leave out all pairs of points $x, y \in X$ where $\{u, v\} \cap \{x, y\} \neq \emptyset$. Time-consuming to find these pairs.

Fast alternative (Jalilian and Waagepetersen, 2017):

$$\tilde{g}^{-\{u,v\}}(u, v; b) = \hat{g}(u, v; b) - 2 \frac{1}{2\pi t} \frac{k_b(t - \|v - u\|)}{\rho(u)\rho(v)|W \cap W_{v-u}|}$$

i.e. just remove contributions corresponding to distance observations $D_{uv} = D_{vu} = \|u - v\|$.

Result: both estimates consistent for MISE (increasing observation window).
Simulation study

MISE for point pattern simulations from 4 models and 5 observation windows: \( W_n = [0, n]^2, n = 1, \ldots, 5: \)

Bias of kernel estimate

Problem: kernel estimate biased for lags $r$ close to zero.

Two variants of kernel estimates for a clustered point pattern:

Left estimate is biased downwards, right estimate is biased upwards.
Bias problem difficult to fix:

- Snethlage (2000) suggested adaptive band width to reduce bias but at the expense of large variance
- Guan (2007) derived correction near zero but assuming Poisson process
Orthogonal series estimation

Orthogonal series estimate (OSE) of probability density on interval $I$:

$$f(x) = \sum_{k=1}^{\infty} \theta_k \phi_k(x) \quad \theta_k = \int_I f(z) \phi_k(z) w(z) dz = \mathbb{E} \phi_k(X) w(X) \quad X \sim f(\cdot)$$

where $\{\phi_k\}_k$ orthogonal basis for weight function $w(\cdot)$.

Suppose $X_1, \ldots, X_n$ iid sample from $f$. Then

$$\hat{\theta}_k = \frac{1}{n} \sum_{i=1}^{n} \phi_k(X_i) w(X_i)$$

unbiased estimate of $\theta_k$.

In practice truncated (hence biased) estimate is used

$$\hat{f}(x) = \sum_{k=1}^{K} \hat{\theta}_k \phi_k(x)$$
OSE for pair correlation function

Estimate $g$ on interval $]0, R[$:

$$g(r) = \sum_{k=1}^{\infty} \theta_k \phi_k(r) \quad \theta_k = \int_0^R g(r) \phi_k(r) w(r) \, dr$$

By Campbell formula we obtain unbiased estimate

$$\hat{\theta}_k = \frac{1}{2\pi} \sum_{\mathbf{u}, \mathbf{v} \in \mathcal{X}_W: \|\mathbf{v} - \mathbf{u}\| \leq R} \frac{\phi_k(\|\mathbf{v} - \mathbf{u}\|) w(\|\mathbf{v} - \mathbf{u}\|)}{\|\mathbf{v} - \mathbf{u}\| \rho(\mathbf{u}) \rho(\mathbf{v}) |\mathcal{W} \cap \mathcal{W}_{\mathbf{v} - \mathbf{u}}|}$$

Similar to estimate for probability density estimation but

- $D_{uv} = \|\mathbf{v} - \mathbf{u}\|$ neither independent nor identically distributed
- we have to correct for spatially varying intensity and edge effects using factor $\rho(\mathbf{u}) \rho(\mathbf{v}) |\mathcal{W} \cap \mathcal{W}_{\mathbf{v} - \mathbf{u}}|$
Choices of basis

Cosine ($w(r) = 1$):

$$\phi_1(r) = \frac{1}{\sqrt{R}}, \quad \phi_k(r) = \frac{\sqrt{2}}{\sqrt{R}} \cos \left( (k - 1)\pi r / R \right), \quad k \geq 2$$

Bessel ($w(r) = r$):

$$\phi_k(r) = \frac{\sqrt{2}}{R J_1(\alpha_0, k)} J_0 \left( \frac{\alpha_0, k}{R} \right), \quad k \geq 1$$
Bessel basis
In practice we use truncated estimate

\[ \hat{g}(r) = \sum_{k=1}^{K} \hat{\theta}_k \phi_k(r) \]

How to choose \( K \)?

(problem analogous to band width selection for kernel estimate)

Option: choose \( K \) that minimizes estimate of mean integrated squared error
Mean integrated squared error (MISE)

\[
\text{MISE}(\hat{g}) = \int_0^R \mathbb{E}[\hat{g}(r) - g(r)]^2 w(r) dr
\]

For OSE,

\[
\text{MISE}(\hat{g}) \equiv \sum_{k=1}^{K} \left[ \mathbb{E}(\hat{\theta}_k)^2 - 2\theta_k^2 \right]
\]

Idea: replace \( \mathbb{E}(\hat{\theta}_k)^2 \) by \((\hat{\theta}_k)^2\) and replace \( \theta_k^2 \) by (asymptotically) unbiased estimate \( \hat{\theta}_k^2 = \)

\[
\frac{1}{4\pi^2} \sum_{u,v,u',v' \in \mathcal{X} \cap W: \|v-u\| \leq R, \|v'-u'\| \leq R} \frac{\phi_k(\|v-u\|) \phi_k(\|v'-u'\|) w(\|v-u\|) w(\|v'-u'\|)}{\|v-u\| \|v'-u'\| \rho(u) \rho(v) \rho(u') \rho(v')} |W \cap W_{v-u}||W \cap W_{v'-u'}|
\]
Asymptotic results

Assume $X$ observed on increasing sequence of windows $W_n$.

$\hat{g}_n$ is estimate based on $X \cap W_n$ with truncation $K_n$.

Need a number of ‘standard’ conditions on point process and smoothing scheme (e.g. increasing $K_n$ but $K_n/|W_n| \to 0$).
Asymptotic results: consistency

Crucial assumption (since we divide by \(\|v - u\|\) in \(\hat{\theta}_k\)):

\[ g(r) \frac{w(r)}{r} \leq C \]

This is OK for Bessel basis (\(w(r) = r\)) if \(g\) bounded.

For cosine basis (\(w(r) = 1\)) we need \(g(r) \to 0\) as \(r \to 0\) (e.g. determinantal point process).

We then have consistency in MISE:

\[ \text{MISE}(\hat{g}_n) = \int_0^R \mathbb{E} [\hat{g}_n(r) - g(r)]^2 w(r) dr \to 0 \]
Under suitable ($\alpha$) mixing and moment conditions we have CLT (Biscio and Waagepetersen, 2017) for spatial averages of the form

\[ S_n = \frac{1}{|W_n|} \sum_{u,v \in X \cap W_n, \|v-u\| \leq R} \frac{f_n(v-u)}{\rho(u) \rho(v)} e_n(v-u) \]

\[ e_n(v-u) = \frac{|W_n|}{|W_n \cap W_{n,v-u}|} \]

whenever $f_n$ is bounded.

This immediately gives CLT for $\hat{\theta}_{k,n}$. 
Asymptotic results: normality

The OSE of $g$ can be rewritten as

$$
\hat{g}_n(r) = \frac{1}{2\pi |W_n|} \sum_{u,v \in X \cap W_n: \|u-v\| \leq R} \frac{f_n(v-u)}{\rho(u)\rho(v)e_n(v-u)}
$$

where

$$
f_n(h) = \frac{w(h)h_n(h)}{\|h\|} \quad h_n(h) = \sum_{k=1}^{K_n} \phi_k(\|h\|) \phi_k(r).
$$

Here $f_n(h)$ not bounded but we assume $h_n$ is $O(K_n^\omega)$ for some $\omega > 0$.

Then

$$
[\text{Var} \hat{g}_n(r)]^{-1/2} \left( \hat{g}_n(r) - \mathbb{E} \hat{g}_n(r) \right) \xrightarrow{D} N(0, 1).
$$

provided

$$
\liminf_{n \to \infty} \frac{\text{Var} \hat{g}_n(r)}{K_n^{2\omega} |W_n|} > 0
$$
Conclusions from simulation studies

We conducted simulation studies for Poisson, Thomas and VarGamma processes (clustered) and determinantal processes (repulsive).

Performance in terms of MISE:

Poisson and cluster processes: orthogonal series estimates as good as or better than kernel estimates. In particular Bessel basis estimators superior to all other estimators.

Determinantal: kernel estimators as good as or better than orthogonal series estimators. Bessel basis estimators inferior.
Application to tropical tree locations

Acalypha diversifolia
- kernel: $\hat{g}_c(r; \hat{b} = 1.35)$
- Bessel: $\hat{K} = 7$, refined
- cosine: $\hat{K} = 11$, refined

Lonchocarpus heptaphyllus
- kernel: $\hat{g}_c(r; \hat{b} = 9.64)$
- Bessel: $\hat{K} = 7$, refined
- cosine: $\hat{K} = 6$, refined

Capparis frondosa
- kernel: $\hat{g}_c(r; \hat{b} = 5.13)$
- Bessel: $\hat{K} = 7$, refined
- cosine: $\hat{K} = 49$, refined
- cosine: $K = 7$, refined

Procedure for choosing $K$ failed for last example?
Conclusion/future work

Orthogonal series estimators have potential:

- for Poisson and clustered point processes: large reduction in MISE.

Issues:

- still bias for small distances.

- global basis functions - estimate \( \hat{g}(r) \) depends on chosen upper range \( R \).

Options:

- Wavelet basis ?
- non-orthogonal bases (frames) ?
References


Thanks for your attention !