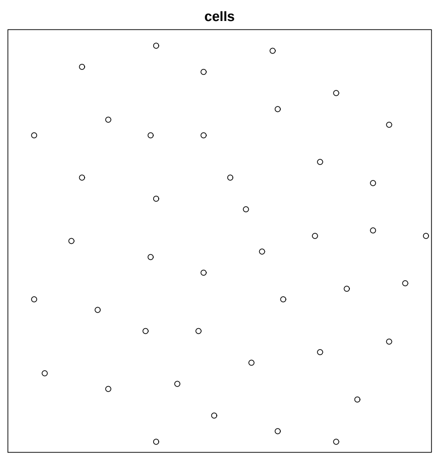
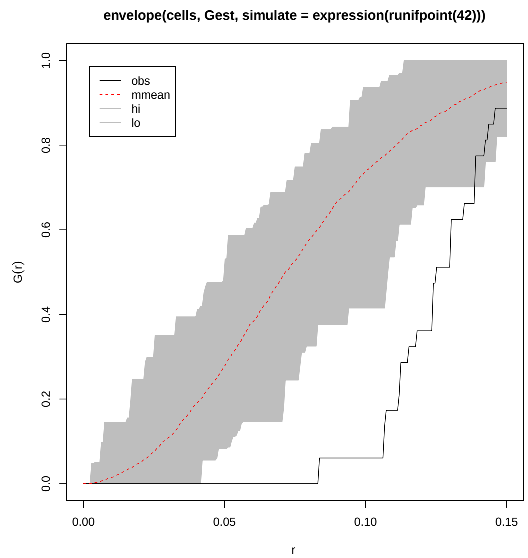
A J–function for inhomogeneous point processes

Marie-Colette van Lieshout

CWI & Eindhoven University of Technology
The Netherlands

Exploratory analysis for spatial point processes






```
plot(envelope(cells, Gest, simulate=expression(runifpoint(42))))
Generating 99 simulations by evaluating expression ...
      lty col   key   label                                           meaning
obs     1   1   obs   obs(r)           observed value of G(r) for data pattern
mmean   2   2 mmean mean(r)               sample mean of G(r) from simulations
hi      1   8    hi   hi(r) upper pointwise envelope of G(r) from simulations
lo      1   8    lo   lo(r) lower pointwise envelope of G(r) from simulations
```

Summary statistics for stationary point processes

Let X be a stationary point process on $\mathbb{R}^d$ with intensity $\rho > 0$.

Popular statistics include

$$\begin{cases} F(t) &= \mathbb{P}(X \cap B(0,t) \neq \emptyset) \\ G(t) &= \mathbb{P}^{!0}(X \cap B(0,t) \neq \emptyset) \\ K(t) &= \mathbb{E}^{!0}\left[\sum_{x \in X} 1\{x \in B(0,t)\}\right] / \rho \\ J(t) &= (1 - G(t)) \,/\, (1 - F(t)) \end{cases}$$

where $B(0,t)$ is the closed ball of radius $t \geq 0$ centred at the origin, $\mathbb{P}^{!0}$ the reduced Palm distribution of X.

Product densities and correlation functions

The **product densities** of a simple point process X satisfy

$$\mathbb{E}\left[\sum\nolimits^{\neq}_{x_1,\ldots,x_n \in X} f(x_1,\ldots,x_n)\right] =$$

$$\int \cdots \int f(x_1,\ldots,x_n)\,\rho^{(n)}(x_1,\ldots,x_n)\,dx_1\cdots dx_n$$

for all measurable $f \geq 0$ ($^{\neq}$ indicates a sum over n-tuples of distinct points).

In words: $\rho^{(n)}(x_1,\ldots,x_n)\,dx_1\cdots dx_n$ is the infinitesimal probability of finding points of X at each of $dx_1, \ldots, dx_n$.

N-point correlation functions are defined recursively by $\xi_1 \equiv 1$ and

$$\frac{\rho^{(n)}(x_1,\ldots,x_n)}{\rho(x_1)\cdots\rho(x_n)} = \sum_{k=1}^{n} \sum_{D_1,\ldots,D_k} \xi_{n(D_1)}(\mathbf{x}_{D_1})\cdots\xi_{n(D_k)}(\mathbf{x}_{D_k})$$

where $\{D_1,\ldots,D_k\}$ is a partition of $\{1,\ldots,n\}$, $\mathbf{x}_{D_j} = \{x_i : i \in D_j\}$.

Summary statistics and product densities

Let X be a stationary point process with intensity $\rho > 0$. Then

$$\begin{cases} K(t) &= \int_{B(0,t)} \frac{\rho^{(2)}(0,x)}{\rho^2}\, dx \\ F(t) &= -\sum_{n=1}^{\infty} \frac{(-1)^n}{n!} \int_{B(0,t)} \cdots \int_{B(0,t)} \rho^{(n)}(x_1, \dots, x_n)\, dx_1 \cdots dx_n \\ G(t) &= -\sum_{n=1}^{\infty} \frac{(-1)^n}{n!} \int_{B(0,t)} \cdots \int_{B(0,t)} \frac{\rho^{(n+1)}(0, x_1, \dots, x_n)}{\rho}\, dx_1 \cdots dx_n \\ J(t) &= 1 + \sum_{n=1}^{\infty} \frac{(-\rho)^n}{n!} J_n(t) \end{cases}$$

for

$$J_n(t) = \int_{B(0,t)} \cdots \int_{B(0,t)} \xi_{n+1}(0, x_1, \dots, x_n)\, dx_1 \cdots dx_n.$$

Hence

$$J(t) - 1 \approx -\rho \left(K(t) - |B(0,t)| \right).$$

Intensity reweighting

Baddeley, Møller and Waagepetersen, SN 2000.

A point process X is **second order intensity-reweighted stationary** if the random measure

$$\Xi = \sum_{x \in X} \frac{\delta_x}{\rho(x)}$$

is second-order stationary, where δ_x denotes the Dirac measure at x.

Examples

- Poisson point processes;
- location dependent thinning of a stationary point process;
- log Gaussian Cox processes driven by a Gaussian random field with translation invariant covariance function.

Summary statistics for inhomogeneous patterns

Assume X is second order intensity-reweighted stationary. Define

$$K_{\text{inhom}}(t) = \frac{1}{|B|} \mathbb{E}\left[\sum\nolimits_{x,y \in X}^{\neq} \frac{1\{x \in B\}\, 1\,\{y \in B(x,t)\}}{\rho(x)\, \rho(y)} \right]$$

regardless of the choice of bounded Borel set $B \subset \mathbb{R}^d$ with strictly positive volume $|B|$, and using the convention $a/0 = 0$ for $a \geq 0$.

To define analogues of F and G, for given $x \in \mathbb{R}^d$ and $t \geq 0$, solve

$$t = \int_{B(x,r(x,t))} \rho(y)\, dy,$$

then set

$$\begin{cases} F_x(t) &= \mathbb{P}(d(x,X) \leq r(x,t)) \\ G_x(t) &= \mathbb{P}^{!x}(d(x,X) \leq r(x,t)) \end{cases}$$

where $d(x,X)$ denotes the shortest distance from x to a point of X.

Inhomogeneous J-function

One could define

$$J_x(t) = \frac{1 - G_x(t)}{1 - F_x(t)}.$$

Drawbacks

- $r(x,t)$ may be hard to compute;
- the definitions depend on x as well as t.

Goal: Give alternative definitions of F, G, and J for intensity-reweighted moment stationary point processes that do not depend on the choice of origin and are easy to use in practice.

Idea: Use the representation in terms of product densities!

Intensity-reweighted moment stationarity

Let X be a simple point process on $\mathbb{R}^d$ for which product densities of all orders exist and $\inf_x \rho(x) = \bar{\rho} > 0$. If the ξ_n are translation invariant, that is,

$$\xi_n(x_1 + a, \ldots, x_n + a) = \xi_n(x_1, \ldots, x_n)$$

for almost all $x_1, \ldots, x_n \in \mathbb{R}^d$ and all $a \in \mathbb{R}^d$, X is **intensity-reweighted moment stationary**.

Examples

- Poisson point processes;
- location dependent thinning of a stationary point process;
- log Gaussian Cox processes driven by a Gaussian random field with translation invariant covariance function.

CWI

Inhomogeneous J-function

Let X be an intensity-reweighted moment stationary point process. Set

$$J_n(t) = \int_{B(0,t)} \cdots \int_{B(0,t)} \xi_{n+1}(0, x_1, \ldots, x_n) \, dx_1 \cdots dx_n$$

and define

$$J_{\text{inhom}}(t) = 1 + \sum_{n=1}^{\infty} \frac{(-\bar{\rho})^n}{n!} J_n(t).$$

Remarks

- $J_{\text{inhom}}(t) > 1$ indicates inhibition at range t;
- $J_{\text{inhom}}(t) < 1$ suggests clustering;
- $J_{\text{inhom}}(t) - 1 \approx -\bar{\rho}\,(K_{\text{inhom}}(t) - |B(0,t)|)$;
- the definition does not depend on the choice of origin.

The generating functional

Recall that for any function $v : \mathbb{R}^d \to [0,1]$ that is measurable and identically 1 except on some bounded subset of $\mathbb{R}^d$,

$$G(v) = \mathbb{E}\left[\prod_{x \in X} v(x)\right],$$

where by convention an empty product is taken to be 1.

Properties

- the distribution of X is determined uniquely by G;
- suppose product densities of all orders exist and let $u : \mathbb{R}^d \to [0,1]$ be measurable with bounded support. Then

$$G(1-u) = 1 + \sum_{n=1}^{\infty} \frac{(-1)^n}{n!} \int \cdots \int u(x_1) \cdots u(x_n)\, \rho^{(n)}(x_1, \ldots, x_n)\, dx_1 \cdots dx_n$$

(provided the series converges).

J_{inhom} in terms of generating functionals

Write, for $t \geq 0$ and $a \in \mathbb{R}^d$,

$$u_t^a(x) = \frac{\bar{\rho}\, 1\{x \in B(a,t)\}}{\rho(x)}, \quad x \in \mathbb{R}^d.$$

Then

$$J_{\mathrm{inhom}}(t) = \frac{G^{!a}\,(1 - u_t^a)}{G\,(1 - u_t^a)}$$

where $G^{!a}$ is the generating functional of the reduced Palm distribution $\mathbb{P}^{!a}$ at a, G that of $\mathbb{P}$ itself.

Note: $G^{!a}\,(1 - u_t^a)$ and $G\,(1 - u_t^a)$ do not depend on the choice of a and lend themselves to estimation.

Remarks

For stationary X, $u_t^a(x) = 1\{x \in B(a,t)\}$, hence

$$\begin{cases} G\,(1 - u_t^a) & = \quad \mathbb{P}(X \cap B(a,t) = \emptyset) = 1 - F(t) \\ G^{!a}\,(1 - u_t^a) & = \quad \mathbb{P}^{!a}(X \cap B(a,t) = \emptyset) = 1 - G(t) \end{cases}$$

Consequently, **one retrieves the classic definition** of the J-function.

For intensity-reweighted moment stationary X, we get **counterparts of the F- and G-functions**:

$$\begin{cases} F_{\mathrm{inhom}}(t) & = \quad 1 - G\,(1 - u_t^a)\,; \\ G_{\mathrm{inhom}}(t) & = \quad 1 - G^{!a}\,(1 - u_t^a) \end{cases}$$

which do not depend on the choice of origin a and are easy to use in practice.

Example 1: The Poisson process

Let X be a Poisson point process with intensity function $\rho : \mathbb{R}^d \to \mathbb{R}^+$ that is bounded away from zero. Then

$$\rho^{(n)}(x_1, \ldots, x_n) = \prod_i \rho(x_i)$$

so X is intensity-reweighted moment stationary and

$$\begin{cases} J_{\text{inhom}}(t) & \equiv & 1 \\ F_{\text{inhom}}(t) & = & 1 - \exp\left[-\bar{\rho}\,|B(0,t)|\right] \\ G_{\text{inhom}}(t) & = & F_{\text{inhom}}(t) \end{cases}$$

Example 2: Location dependent thinning

Let X be a simple, stationary point process on $\mathbb{R}^d$ that is thinned with retention probability $p : \mathbb{R}^d \to (0, 1)$. Then

$$\rho_{\text{th}}^{(n)}(x_1, \ldots, x_n) = \rho^{(n)}(x_1, \ldots, x_n) \prod_{i=1}^{n} p(x_i)$$

so

$$\frac{\rho_{\text{th}}^{(n)}(x_1, \ldots, x_n)}{\rho_{\text{th}}(x_1) \cdots \rho_{\text{th}}(x_n)} = \frac{\rho^{(n)}(x_1, \ldots, x_n)}{\rho^n}$$

and

$$\xi_n^{\text{th}}(x_1, \ldots, x_n) = \xi_n(x_1, \ldots, x_n)$$

is translation invariant.

Example 2: Location dependent thinning (ctd)

If the retention probabilities are bounded away from zero by $\bar{p}$,

$$\begin{cases} J^{\rm th}_{\rm inhom}(t) & = & 1 + \sum_{n=1}^{\infty} \frac{(-\rho\,\bar{p})^n}{n!} J_n(t) \\ 1 - F^{\rm th}_{\rm inhom}(t) & = & \mathbb{E}\left[(1-\bar{p})^{n(X \cap B(0,t))} \right] \\ 1 - G^{\rm th}_{\rm inhom}(t) & = & \mathbb{E}^{!0}\left[(1-\bar{p})^{n(X \cap B(0,t))} \right] \end{cases}$$

where $J_n(t)$ refers to the underlying point process X.

Example 3: Log Gaussian Cox process

The intensity function of the driving random measure is of the form

$$\Lambda(x) = \exp\left[Z(x)\right]$$

where Z is a Gaussian field with mean function μ and covariance function $\sigma^2 r(\cdot)$.

Assume that μ is continuous and bounded and r translation invariant. Then the Cox process X defined by Λ is well-defined and

$$\rho^{(n)}_{\mathrm{th}}(x_1, \ldots, x_n) = \mathbb{E}\left[\prod_{i=1}^{n} \Lambda(x_i)\right] = \mathbb{E}\left[e^{\sum_{i=1}^{n} Z(x_i)}\right].$$

Consequently, X is intensity-reweighted moment stationary with

$$\rho(x) = \exp\left[\mu(x) + \sigma^2/2\right]$$

bounded away from zero.

Example 3: Log Gaussian Cox process (ctd)

Write $\bar{\mu} = \inf_{x \in \mathbb{R}^d} e^{\mu(x)}$ and $Y(x) = Z(x) - \mu(x)$. Then Y is stationary and

$$\begin{cases} 1 - F_{\text{inhom}}(t) & = & \mathbb{E}_Y \left[\exp \left[-\bar{\mu} \int_{B(a,t)} e^{Y(x)} dx \right] \right] \\ 1 - G_{\text{inhom}}(t) & = & \mathbb{E}_Y \left[\frac{e^{Y(a)}}{e^{\sigma^2/2}} \exp \left[-\bar{\mu} \int_{B(a,t)} e^{Y(x)} dx \right] \right] \end{cases}$$

regardless of the choice of a. Hence

$$J_{\text{inhom}}(t) = \frac{\mathbb{E}_Y \left[e^{Y(0)} \exp \left[-\bar{\mu} \int_{B(0,t)} e^{Y(x)} dx \right] \right]}{\mathbb{E}_Y \left[e^{Y(0)} \right] \mathbb{E}_Y \left[\exp \left[-\bar{\mu} \int_{B(0,t)} e^{Y(x)} dx \right] \right]}.$$

Estimation

Let $W \subset \mathbb{R}^d$ be compact with non-empty interior. Suppose X is observed in W and $\bar{\rho} = \inf_{x \in W} \rho(x) > 0$.

Let $L \subseteq W$ be a finite point grid. Set

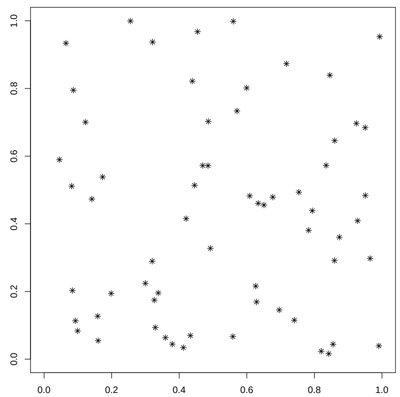
$$\begin{cases} 1 - \widehat{F_{\text{inhom}}}(t) & = & \frac{\sum_{l_k \in L \cap W_{\ominus t}} \prod_{x \in X \cap B(l_k, t)} \left[1 - \frac{\bar{\rho}}{\rho(x)}\right]}{\# L \cap W_{\ominus t}} \\ 1 - \widehat{G_{\text{inhom}}}(t) & = & \frac{\sum_{x_k \in X \cap W_{\ominus t}} \prod_{x \in X \setminus \{x_k\} \cap B(x_k, t)} \left[1 - \frac{\bar{\rho}}{\rho(x)}\right]}{\# X \cap W_{\ominus t}} \end{cases}$$

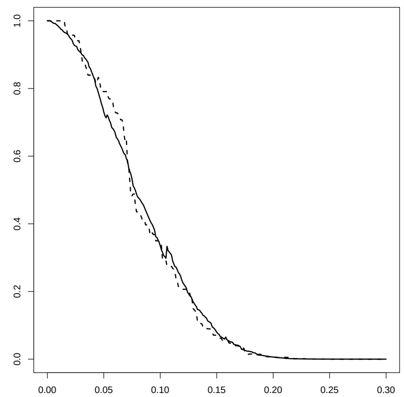
where $W_{\ominus t} = \{x \in W : B(x,t) \subseteq W\}$.

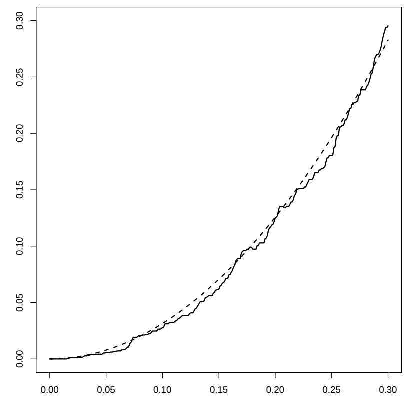
Remarks

- $1 - \widehat{F_{\text{inhom}}}(t)$ is unbiased, $1 - \widehat{G_{\text{inhom}}}(t)$ ratio-unbiased;
- if ρ is unknown, plug in its kernel estimator.

Example 1: The Poisson process



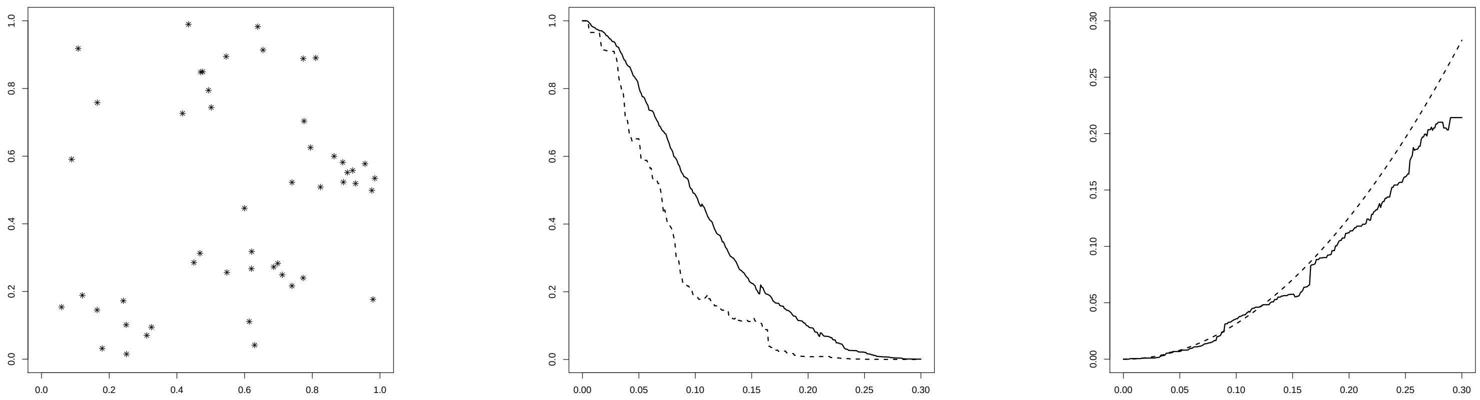




Intensity function

$$\rho(x, y) = 100\, e^{-y}.$$

Example 2: Log Gaussian Cox process



Mean function μ satisfying

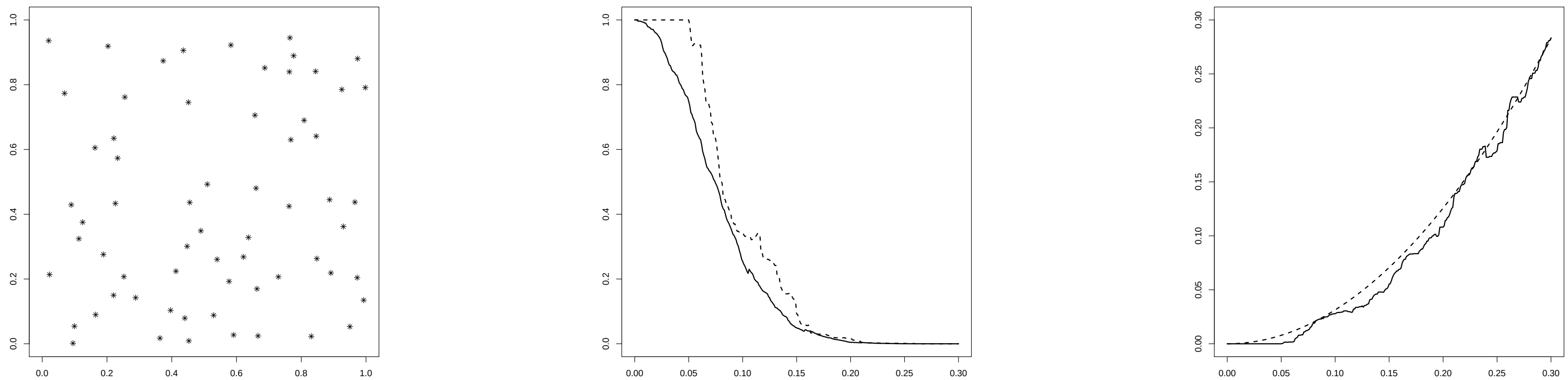
$$e^{\mu(x,y)} = 100\, e^{-y-1/2},$$

$\sigma^2 = 1$, and

$$r(t) = \exp\left[-t/0.143\right].$$

Hence $\rho(x,y) = 100\, e^{-y}$.

Example 3: Location dependent thinning



Hard core process with conditional intensity

$$\beta\, 1\{d(x, X \setminus \{x\} > R\}$$

for $\beta = 200$, $R = 0.05$. For retention probability

$$p(x, y) = e^{-y},$$

$\rho(x, y) = \rho\, e^{-y}$ so that $\rho(x, y)/\bar{\rho}$ is equal to that of the previous examples.

Pakistani earthquakes

Pakistan is regularly affected by earthquakes due to the subduction of the Indo-Australian continental plate under the Eurasian plate.

There are two convergence zones: One crosses the country from its South-West border with the Arabic Sea to Kashmir in the North-East, the other crosses the Northern part of the country in the East-West direction.

Two major earthquakes were recorded during 1973–2008:

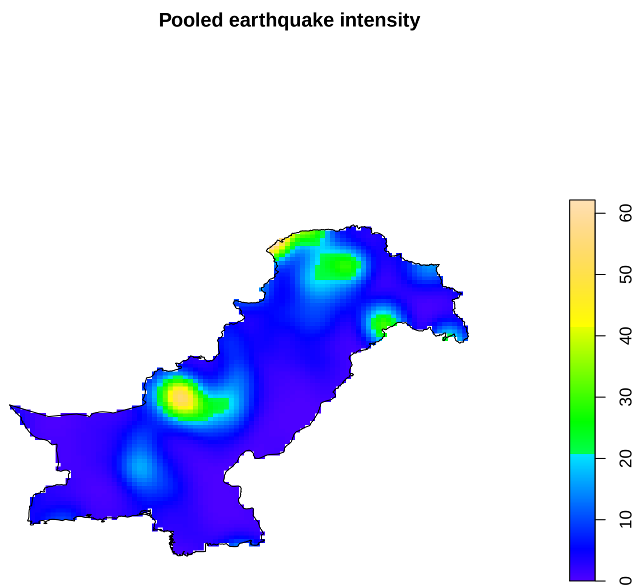
- 1997: magnitude 7.3 along the SW to NE zone; about seventy casualties;
- 2005: magnitude 7.6 in Kashmir; devastating with at least $86,000$ casualties.

Restrict to shallow quakes (depth less than 70 km) of magnitude at least 4.5 as deeper and weaker ones may not be felt.

CWI

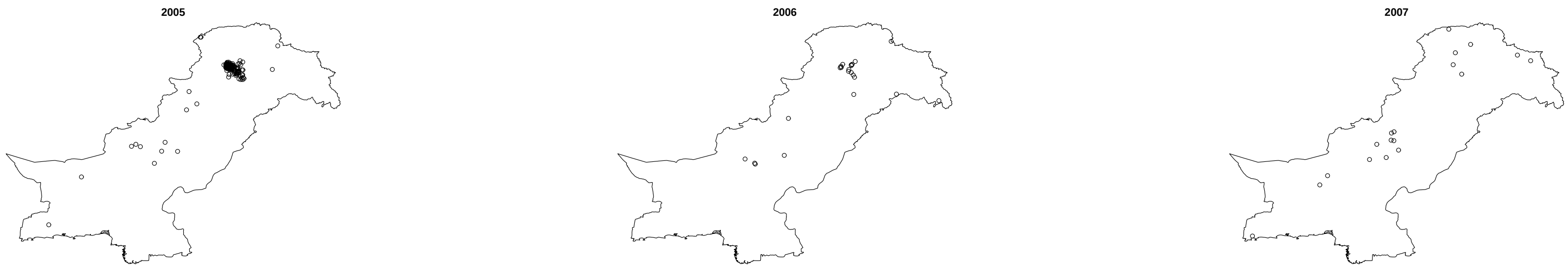
Pakistani earthquakes: intensity function

- To avoid edge effects, include earthquakes within a distance of about one degree from the Pakistan border;
- aggregate into a single pattern, but exclude the major earthquake years;
- calculate the kernel estimator of intensity (isotropic Gaussian kernel with standard deviation 0.5, that is, approximately 50 km).



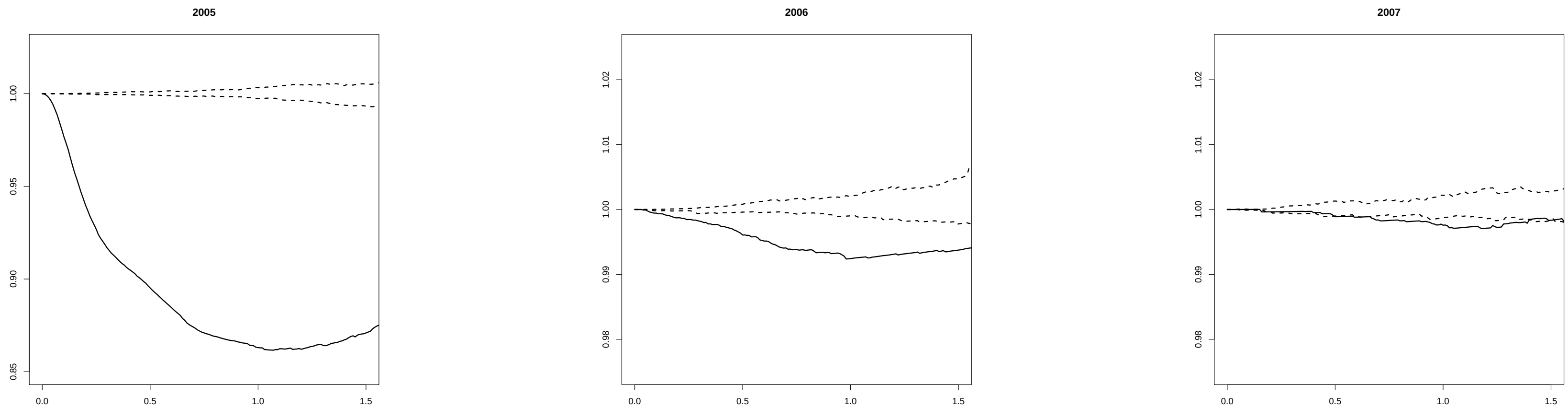
Earthquake data 2005–2007

Note: a KPSS test indicates the intensity pattern persists over the years.



Question: can the data be explained by a series of inhomogeneous Poisson processes?

Inhomogeneous J-functions 2005–2007



$\widehat{J_{\mathrm{inhom}}}$ for the locations of shallow earthquakes of magnitude at least 4.5 in 2005–2007 with upper and lower envelopes based on 19 independent realisations of an inhomogeneous Poisson process.

Conclusion: evidence of clustering over and beyond that explained by the spatial variation.

Summary and extensions

- We defined three new summary statistics for intensity-reweighted moment stationary point processes;
- calculated them explicitly for the three representative classes;
- derived minus sampling estimators;
- presented simulation examples;
- applied them to earthquake data;
- extension to marked point processes is straightforward;
- the statistics can be extended to space-time point processes cf. Gabriel and Diggle, SN 2009.

CWI