Marked Point Processes

Tonglin Zhang

Outline

- Description and Definition
- Research Problems
- Testing Independence
- A Few Famous Models
- First-Order and Second-Order Analyses
- Asymptotic Frameworks

Description

A marked point process is composed of a point process and associate marks, which can be expressed as

$$\{(\mathbf{s}_i, m_i) : i = 1, \cdots, n\},$$

where $\mathbf{s}_1, \cdots, \mathbf{s}_n$ are locations and $m_1, \cdots, m_n$ are associated marks.

A Simulated Example

I did simulation:

- Generate Homogeneous Poisson point process with $\lambda(\mathbf{s}) = 1000$ on $[0, 1]^2$.
- Generate $m_i \sim^{iid} Exp(1)$ at each $\mathbf{s}_i$.
- Then, the marked point process $\{(\mathbf{s}_i, m_i)\}$ is derived.

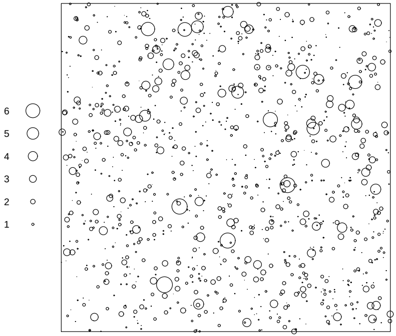


Figure : A Simulated Example of Marked Point Processes on $[0, 1]^2$.

Definition

A marked point process N can be understood as a pure point process on $\mathcal{S} \times \mathcal{M}$, where $\mathcal{S}$ is the domain of points and $\mathcal{M}$ is the domain of marks, such that

$$N(A \times \mathcal{M}) < \infty$$

if A is bounded.

Therefore, one can define the first-order intensity function as

$$\lambda(\mathbf{s}, m) = \lim_{|d\mathbf{s}\times dm|\to 0} \frac{E[N(d\mathbf{s} \times dm)]}{|d\mathbf{s}_1 \times dm_1|}.$$

The second-order intensity function is

$$\begin{aligned} &\lambda_2((\mathbf{s}_1, m_1), (\mathbf{s}_2, m_2)) \\ = &\lim_{|d\mathbf{s}_1\times dm_1|,|d\mathbf{s}_2\times dm_2|\to 0} \frac{E[N(d\mathbf{s}_1 \times dm_1)N(d\mathbf{s}_2 \times dm_2)]}{|d\mathbf{s}_1 \times dm_1||d\mathbf{s}_2 \times dm_2|}. \end{aligned}$$

Similarly, we can define the kth-order intensity function λ_k.

Based on joint intensity functions, we can define marginal intensity functions. For example, we can define

$$\lambda_s(\mathbf{s}) = \int_{\mathcal{M}} \lambda(\mathbf{s}, m) dm$$

as the marginal (first-order) intensity for points

$$\lambda_m(m) = \int_{\mathcal{S}} \lambda(\mathbf{s}, m) d\mathbf{s}$$

as the marginal (first-order) intensity for marks. Then,

$$f(m|\mathbf{s}) = \frac{\lambda(\mathbf{s}, m)}{\lambda_s(\mathbf{s})}$$

is the conditional density of marks.

In addition, we can also define

$$\lambda_{2,s}(\mathbf{s}_1, \mathbf{s}_2) = \int_{\mathcal{M}} \int_{\mathcal{M}} \lambda_2((\mathbf{s}_1, m_1), (\mathbf{s}_2, m_2)) dm_2 dm_1$$

and

$$\lambda_{2,m}(m_1, m_2) = \int_{\mathcal{S}} \int_{\mathcal{S}} \lambda_2((\mathbf{s}_1, m_1), (\mathbf{s}_2, m_2)) d\mathbf{s}_2 d\mathbf{s}_1.$$

Then,

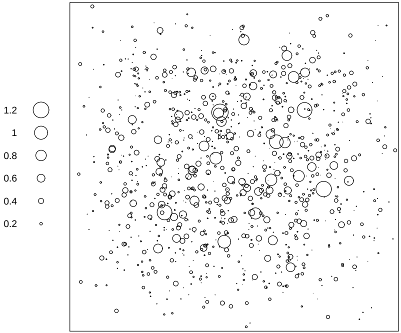
$$f_2(m_1, m_2 | \mathbf{s}_1, \mathbf{s}_2) = \frac{\lambda_2((\mathbf{s}_1, m_1), (\mathbf{s}_2, m_2))}{\lambda_{2,s}(\mathbf{s}_1, \mathbf{s}_2)}$$

is the bivariate conditional density of marks.

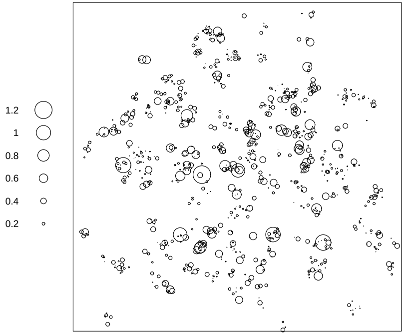
A Simulated Example

Let $\mathcal{S} = [0,1]^2$ and $\mathcal{M} = \mathbb{R}^+$. The intensity function of points is $\lambda_s(\mathbf{s}) = \kappa\beta(2,2)$. The density of marks is $f(m|\mathbf{s}) = \exp(\omega)$, where $\omega = 10(1 + \|\mathbf{s} - \mathbf{s}_0\|)$ with $\mathbf{s}_0 = (0.5, 0.5)$. We consider the marked Poisson and cluster point process, respectively. We choose $\kappa = 1000$ and the size of cluster is Poisson with mean 5.

Marked Poisson Point Process



Marked Cluster Point Process



Research Problems

Althouth many people have done research on pure point processes, only a few people have worked on marked Point processes. The most interesting problem is to investigate the relationship between points and marks, which may include

- testing independence;
- modeling relationship between points and marks;
- first-order and second-order analyses;
- return intervals; and
- parametric and nonparametric analysis.

Testing Independence

There are two ways to describe independence between marks and points. The first one uses the Janossy measure and the second one uses Intensity functions.

Definition of Independence

If the Janossy measure is used, then one uses that

$$\begin{aligned}&P((\mathbf{s}_1, m_1) \in A_1 \times B_1, \cdots, (\mathbf{s}_n, m_n) \in A_n \times B_n)\\=&P(\mathbf{s}_1 \in A_1, \cdots, \mathbf{s}_n \in A_n)P(m_1 \in B_1, \cdots, B_n \in B_n),\end{aligned}$$

for any $n \geq 1$, $A_1, \cdots, A_n \subseteq \mathcal{S}$, and $B_1, \cdots, B_n \subseteq \mathcal{M}$. This is often used as the definition of *independence*.

Definition of Separability

If the intensity functions are used, then one uses that

$$\frac{\lambda_k((\mathbf{s}_1, m_1), \cdots, (\mathbf{s}_k, m_k))}{\lambda_{k,s}(\mathbf{s}_1, \cdots, \mathbf{s}_k)\lambda_{k,m}(m_1, \cdots, m_k)}$$

is constant. This is often called *separability*, where $\lambda_{k,s}$ and $\lambda_{k,m}$ are the marginal intensity for points and marks, respectively.

Question: What is the relationship between these two definitions?

Testing Independence

I have proposed a Kolmogorov-Smirnov test to assess independence.

- The test considers the null hypothesis

$$H_0 : P(\mathbf{s} \times m \in A \times B) = P(\mathbf{s} \in A)P(\mathbf{m} \in B)$$

  for any $A \in \mathscr{B}(\mathcal{S})$ and $B \in \mathscr{B}(\mathcal{M})$.

- It is enough to consider a collection of subsets in $\mathcal{A} \in \mathscr{B}(\mathcal{S})$ and $\mathcal{B} \in \mathscr{B}(\mathcal{M})$ such that the test statistic is

$$T_n = \sqrt{n} \sup_{A \in \mathcal{A}, B \in \mathcal{B}} \left| \frac{N(A \times B)}{n} - \frac{N(A \times \mathcal{M})}{n} \frac{N(\mathcal{S} \times B)}{n} \right|.$$

  The null hypothesis is rejected if T_n is large.

- Let $\kappa = E(n) = E[N(\mathcal{S} \times \mathcal{M})]$. If κ is large, then T_n weakly converges to a Brownian pillow on $\mathcal{S} \times \mathcal{M}$ whose distribution is unknown.
- However, if we define $\mathcal{A}$ by a transformation F from $\mathcal{S}$ to $\mathbb{R}$ as $A \in \mathcal{A}$ if and only if

$$A = F^{-1}((-\infty, t] : t \in \mathbb{R}),$$

  then

$$T_n \xrightarrow{D} \mathbb{W}^2,$$

  where $\mathbb{W}^2$ is the standard two-dimensional Brownian pillow on $[0, 1]^2$.
- $\mathbb{W}^2$ is an extension from the standard Brownian bridge $\mathbb{W} = \mathbb{W}^1$ on $\mathbb{R}^2$. However, the quantile function of $\mathbb{W}^2$ is unknown. I derive it using a simulation method.

There are some other methods.

- Schoenberg (2004) considers a test based on marked Poisson point processes.
- Guan (2007) proposes a stationary tests.
- Schlather, Ribeiro, and Diggle (2004) consider a random field methods.
- All of these methods contain unknown parameters or functions to be estimated.

If independence is accepted, then we can

- model points and marks separately;
- ignore the relationship between points and marks;
- predict points and marks independently.

If points and marks are dependent, then we should consider their relationship. We can use

- Intensity dependent models; and
- Location dependent models.

Intensity-Dependent Models

- Ho and Stoyan (2008, *Statistics and Probability Letters*, 1194-1199) propose (for normal marks)

$$m(\mathbf{s}) = a + b\lambda_s(\mathbf{s}) + \epsilon(\mathbf{s}),$$

where $\epsilon(\mathbf{s})$ is a white noise; and

$$\lambda_s(\mathbf{s}) = \exp(\alpha + \beta S(\mathbf{s})); m(\mathbf{s}) = S(\mathbf{s}) + \epsilon(\mathbf{s}),$$

where $S(\mathbf{s})$ is a Gaussian random field.

- Myllymaki and Penttinen (2009, *Statistica Neerlandica*, 450-473) consider an intensity-dependent marking model as

$$m(\mathbf{s})|\lambda_s(\mathbf{s}) \sim F_m(\cdot|\lambda(\mathbf{s})).$$

Location-Dependent Models

It is possible to consider a model with

$$E[m(\mathbf{s})] = f_{\theta(\mathbf{s})}(m)$$

This is also called the location dependent model. It has been considered. The idea can be thought as motivated from the GWR (geographical weighted regression).

First-Order Analysis

We can still use the method of Stochastic Integral for the first-order analysis, which is based on an expression of

$$\int_{\mathcal{S}} \int_{\mathcal{M}} f(\mathbf{s}, m) N(d\mathbf{s} \times dm).$$

For example, if $\lambda(\mathbf{s}, m) = \lambda_\theta(\mathbf{s}, m)$, then the composite likelihood function

$$\ell(\theta) = \int_{\mathcal{S}} \int_{\mathcal{M}} \log \lambda_\theta(\mathbf{s}, m) N(d\mathbf{s} \times dm) - \int_{\mathcal{S}} \int_{\mathcal{M}} \lambda_\theta(\mathbf{s}, m) dm d\mathbf{s}$$

can be used.

Second-Order Analysis

One can also consider a method to estimate $\lambda_2((\mathbf{s}_1, m_1), (\mathbf{s}_2, m_2))$. However, there is no other measurement for the second-order properties.

There are still two types of asymptotic frameworks:

- Increasing domain.
- Fixed domain.