Point Processes

Tonglin Zhang
Outline

- Simulated Examples
- Interesting Problems
- Analysis under Stationarity
- Analysis under Nonstationarity
- Likelihood Analysis
- Stochastic Integral
- Asymptotic Frameworks
Simulated Examples

Poisson Processes

Let $S$ be the study area. A poisson process is derived if $N(A_1), \cdots, N(A_k)$ are iid Poisson random variable with mean $\mu(A_1), \cdots, \mu(A_k)$ if $A_1, \cdots, A_k$ are disjoint subsets of $S$.

Extensions:

- Cox process: $\mu$ is a random measure (e.g. $\log \mu$ is given by a Gaussian random field).
- Mixed poisson process (mean measure is given by $Y \mu$, where $Y$ is a nongenative random variable).
Simulated Examples

Figure: A Homogeneous Poisson Point Process on $[0, 1]^2$. 
Neyman Scott Cluster Process

A Neyman Scott cluster process has:
- A parent process and an off-spring process.
- The parent process is Poisson.
- The off-spring process is derived from the parent process: each parent point generate a Poisson number of off-spring points around it.
- Example: Let $s_1, \ldots, s_m$ be parent point. Then,
  - There are $m$ clusters.
  - Cluster 1 is around $s_1$; cluster 2 is around $s_2$; and so on.
Figure: A Homogeneous Neyman Scott Cluster Process on $[0, 1]^2$. 
Simulated Examples

How to generate an inhomogeneous Poisson point process

Let $f$ be a PDF on $s$. Then, we can

1. Generate $n \sim \text{Poisson}(b)$.
2. Generate $n$ observations independently from a PDF $f$. Let them be $s_1, \cdots, s_n$.
3. Then, $\{s_1, \cdots, s_m\}$ is a sample from Poisson point process with intensity

$$\lambda(s) = bf(s).$$
Other Processes

Compound Point Process

- Generate $n$ from a nonnegative discrete random variable.
- Generate $n$ observations independent from $f$. 
Stationarity: \( \lambda(s) \) is a constant; \( \lambda_2(s_1, s_2) = \lambda_2(s_1 - s_2) \);
\( g(s_1, s_2) = g(s_1 - s_2) \); and etc.

Isotropic: Stationarity, and \( \lambda_2(s_1, s_2) = \lambda_2(\|s_1 - s_2\|) \).

Nonstationary process: estimation of intensity functions.

Stationary temporal point process: return (or recurrence) intervals.

Second-order analysis: \( K \)-function, \( L \)-function, Pair correlation function.

Return Intervals.

Marked point processes: relationship between points and marks.
The \( K \)-Function (Ripley 1976)

Let \( \{s_1, \cdots, s_n\} \) be observed points from a stationary point process with intensity \( \lambda \). \( K \)-function describes the second-order properties, which is

\[
K(t) = 2\pi \int_0^t u g(u) du,
\]

where \( u \) is the pair correlation function.
Analysis under Stationarity

We can estimate

$$
\hat{K}(t) = \frac{1}{|A| \lambda^2} \sum_{i=1}^{n} \sum_{j \neq i} w_{ij} I(d_{ij} \leq t),
$$

where $A$ is a disc, $w_{ij}$ is the weight for edge correction, and $d_{ij} = \|s_i - s_j\|$ ($\lambda$ may be replaced as $\hat{\lambda}$ if $\lambda$ is unknown).

There is a local version, which is called the local $K$-function, which is

$$
\hat{K}_{local}(t; s_i) = \frac{1}{|A| \lambda} \sum_{j \neq i} w_{ij} I(d_{ij} \leq t).
$$
If \( N \) is a homogeneous Poisson point process, then

\[
K(t) = 2\pi \int_{0}^{t} u\,du = \pi t^2.
\]

For any \( s \) in \( S \), let

\[ U_{s,t} = \{ s' : 0 < \|s - s'\| \leq t \}. \]

If there is a point at \( s \), then

\[
E[N(U_{s,t})|s] = \pi t^2.
\]

Therefore, the L-function

\[
L(t) = \sqrt{\frac{K(t)}{\pi}} - t
\]

is useful.
If $N$ is a homogeneous cluster point process with cluster size $k$, then

$$g(u) \approx k.$$ 

Therefore,

$$K(t) \approx k\pi t^2.$$ 

Then,

$$E[N(U_{s,t})|s] \approx k\pi t^2.$$ 

There is

$$L(t) \approx (k-1)t > 0.$$ 

For any $A \subset S$, there is

$$E[N(A)] = \lambda |A|$$

and

$$V[N(A)] \approx kE[N(A)] = k\lambda |A|.$$
Pair Correlation Function

If $N$ is a Neyman cluster process, then we roughly have

$$g(u) \approx k < 1$$

for some $k$. Then,

$$L(t) < 0.$$
Pair Correlation Function

If $N$ is strong stationary, then (under for conditions for weak dependence) for any $A$, we roughly have

$$V[N(A)] = \phi E[N(A)]$$

where

$$\phi = \lambda \int_{\mathbb{R}^d} [g(u) - 1]du + 1.$$
A Simulated Example

We evaluate the value $\phi = \frac{V[N(A)]}{E[N(A)]}$ based on the following example.

- $N$ is a cluster process on $[0, 1]^2$ with cluster size $k$.
- The cluster is determined by a bivariate normal with center at its parent point and standard deviance $\sigma$.
- The parameters are $(\lambda, \sigma, k)$.
- $A$ is a rectangle at $(0.2, 0.2), (0.2, 0.8), (0.8, 0.2)$, and $(0.8, 0.8)$.
- I simulated 1000 times.
Figure: An Example for Dispersion Effect under Stationarity
**Analysis under Stationarity**

<table>
<thead>
<tr>
<th>lambda</th>
<th>sigma</th>
<th>k</th>
<th>E[N(A)]</th>
<th>V[N(A)]</th>
<th>phi</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000</td>
<td>0.005</td>
<td>4</td>
<td>359.872</td>
<td>1413.705</td>
<td>3.928</td>
</tr>
<tr>
<td>1000</td>
<td>0.005</td>
<td>2</td>
<td>360.600</td>
<td>695.630</td>
<td>1.929</td>
</tr>
<tr>
<td>1000</td>
<td>0.005</td>
<td>1</td>
<td>360.188</td>
<td>355.049</td>
<td>0.986</td>
</tr>
<tr>
<td>1000</td>
<td>0.001</td>
<td>4</td>
<td>359.594</td>
<td>1426.759</td>
<td>3.968</td>
</tr>
<tr>
<td>1000</td>
<td>0.001</td>
<td>2</td>
<td>360.050</td>
<td>710.812</td>
<td>1.974</td>
</tr>
<tr>
<td>1000</td>
<td>0.001</td>
<td>1</td>
<td>360.346</td>
<td>368.514</td>
<td>1.023</td>
</tr>
<tr>
<td>5000</td>
<td>0.005</td>
<td>4</td>
<td>1800.936</td>
<td>7131.298</td>
<td>3.960</td>
</tr>
<tr>
<td>5000</td>
<td>0.005</td>
<td>2</td>
<td>1799.848</td>
<td>3582.422</td>
<td>1.990</td>
</tr>
<tr>
<td>5000</td>
<td>0.005</td>
<td>1</td>
<td>1800.289</td>
<td>1769.029</td>
<td>0.983</td>
</tr>
<tr>
<td>5000</td>
<td>0.001</td>
<td>4</td>
<td>1800.547</td>
<td>7301.358</td>
<td>4.055</td>
</tr>
<tr>
<td>5000</td>
<td>0.001</td>
<td>2</td>
<td>1800.527</td>
<td>3561.236</td>
<td>1.978</td>
</tr>
<tr>
<td>5000</td>
<td>0.001</td>
<td>1</td>
<td>1800.356</td>
<td>1823.486</td>
<td>1.013</td>
</tr>
</tbody>
</table>
Return Intervals.

Let $N$ be a stationary point process on $\mathbb{R}$. Let $\tau$ be the time for the next occurrence. If $N$ is Poisson, then the average return time of a point is

$$E(\tau) = \frac{1}{\lambda}.$$  

If $N$ is a cluster process, then we should consider the return of parent points. It is about

$$E(\tau) = \frac{1}{k\lambda}.$$  

If there is a return, then it is a return of cluster.
Assume $N$ is not stationary with first-order intensity $\lambda(s)$ and second-order intensity $\lambda_2(s_1, s_2)$. Recently, most research focuses on the second-order intensity-rewighted stationary or isotropic process, where the first is defined by

$$g(s_1, s_2) = \frac{\lambda_2(s_1, s_2)}{\lambda(s_1)\lambda(s_2)} = g(s_1 - s_2)$$

and the second is defined by

$$g(s_2, s_2) = g(\|s_1 - s_2\|).$$
For any \( A \subseteq S \), there is

\[
V[N(A)] \approx (\phi - 1)E^2[N(A)] + E[N(A)]
\]

for a certain \( \phi \).

If \( N \) is an inhomogeneous cluster process, then we roughly have

\[
V[N(A)] \approx kE[N(A)],
\]

where \( k \) is the cluster size. Therefore, an inhomogeneous cluster process is not a second-order intensity-rewighted stationary process.
If $\lambda(s)$ is given, then the $K$-function is estimated by

$$\hat{K}(t, \lambda) = \frac{1}{|A|} \sum_{i=1}^{n} \sum_{j \neq i} \frac{w_{ij} I(d_{ij} \leq t)}{\lambda(s_i) \lambda(s_j)},$$

where $A$ is a disc and $w_{ij}$ is edge-correction.
One should

- Estimate $\lambda(s)$;
- Estimate $\lambda_2(s_1, s_2)$ by estimating $g(u)$ or $K(u)$;
- It is hard to provide nonparametric estimation for both the first-order and the second-order intensity functions. Therefore, one often (e.g. Guan (2009) JASA, 1482-1491) considers:
  - A parametric analysis for $\lambda(s)$; and
  - a nonparametric analysis for $\lambda_2(s_1, s_2)$. 
A Simulated Example

We generate data from cluster process with $\lambda(s) = b\beta(2, 2)$.
Analysis under Stationarity

Table: Averages under Nonstationarity in 10,000 Simulations

<table>
<thead>
<tr>
<th>$b$</th>
<th>$\sigma$</th>
<th>$k$</th>
<th>$E[N(A)]$</th>
<th>$V[N(A)]$</th>
<th>$\phi$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000</td>
<td>0.005</td>
<td>4</td>
<td>472.797</td>
<td>1857.836</td>
<td>3.929</td>
</tr>
<tr>
<td>1000</td>
<td>0.005</td>
<td>2</td>
<td>472.561</td>
<td>935.265</td>
<td>1.979</td>
</tr>
<tr>
<td>1000</td>
<td>0.005</td>
<td>1</td>
<td>472.442</td>
<td>467.561</td>
<td>0.990</td>
</tr>
<tr>
<td>1000</td>
<td>0.001</td>
<td>4</td>
<td>471.662</td>
<td>1876.383</td>
<td>3.978</td>
</tr>
<tr>
<td>1000</td>
<td>0.001</td>
<td>2</td>
<td>472.554</td>
<td>957.339</td>
<td>2.026</td>
</tr>
<tr>
<td>1000</td>
<td>0.001</td>
<td>1</td>
<td>472.861</td>
<td>482.426</td>
<td>1.020</td>
</tr>
<tr>
<td>5000</td>
<td>0.005</td>
<td>4</td>
<td>2362.713</td>
<td>9401.313</td>
<td>3.979</td>
</tr>
<tr>
<td>5000</td>
<td>0.005</td>
<td>2</td>
<td>2362.378</td>
<td>4717.934</td>
<td>1.997</td>
</tr>
<tr>
<td>5000</td>
<td>0.005</td>
<td>1</td>
<td>2363.058</td>
<td>2343.867</td>
<td>0.992</td>
</tr>
<tr>
<td>5000</td>
<td>0.001</td>
<td>4</td>
<td>2362.672</td>
<td>9274.963</td>
<td>3.926</td>
</tr>
<tr>
<td>5000</td>
<td>0.001</td>
<td>2</td>
<td>2361.637</td>
<td>4607.397</td>
<td>1.951</td>
</tr>
<tr>
<td>5000</td>
<td>0.001</td>
<td>1</td>
<td>2363.759</td>
<td>2387.218</td>
<td>1.010</td>
</tr>
</tbody>
</table>
If $N$ is an inhomogeneous Poisson process with intensity function $\lambda_\theta(s)$, then the loglikelihood function is

$$
\ell(\theta) = \sum_{i=1}^{n} \log \lambda_\theta(s_i) - \int_{S} \lambda_\theta(u) du.
$$

If $N$ is not Poisson, the above is the composite loglikelihood function, which can also provide a consistent estimator.
Let $f(s)$ be a function. We can use Stochastic Integral method as

$$U = \sum_{i=1}^{n} f(s_i) = \int_{S} f(s)N(ds).$$

Then,

$$E(U) = \int_{S} f(s)\lambda(s)ds$$

and

$$E(U^2) = \int_{S} \int_{S} f(s)f(s')\lambda_2(s,s')ds'ds + \int_{S} f^2(s)\lambda(s)ds.$$ 

Therefore,

$$V(U) = \int_{S} \int_{S} f(s)f(s')[g(s,s')-1]\lambda(s)\lambda(s')ds'ds + \int_{S} f^2(s)\lambda(s)ds.$$
For a bivariate function $f(s, s')$, there is

$$U = \int_S \int_S f(s, s') N(ds) N(s').$$

Then,

$$E(U) = \int_S \int_S f(s, s') \lambda_2(s, s') ds ds' + \int_S f(s, s) \lambda(s) ds.$$
In addition, there is

\[
E(U^2) = \int_S \int_S \int_S \int_S f(s, s') f(s'', s''') \lambda_4(s, s', s'', s''') ds ds' ds'' ds'''
+ \int_S \int_S \int_S [2f(s, s') f(s'', s'') + f(s, s') f(s, s'')]
\lambda_3(s, s', s'') ds ds' ds''
+ \int_S \int_S [f(s, s) f(s', s') + f(s, s') f(s, s')]
\lambda_2(s, s') ds ds'
+ \int_S f^2(s, s) \lambda(s) ds.
\]
For $\ell(\theta)$, there is

$$
\ell(\theta) = \int_S \log \lambda_\theta(s) N(ds) - \int_S \lambda_\theta(u) du.
$$

Then, we have

$$
\frac{\partial \ell(\theta)}{\partial \theta_j} = \int_S \frac{1}{\lambda_\theta(s)} \frac{\partial \lambda_\theta(s)}{\partial \theta_j} N(ds) - \int_S \frac{\partial \lambda_\theta(u)}{\partial \theta_j} du.
$$

Thus,

$$
E\left[ \frac{\partial \ell(\theta)}{\partial \theta_j} \right] = 0.
$$
There are two different types of asymptotic frameworks:

- Increasing domain: $S$ increases to $\mathbb{R}^d$ but intensity functions do not vary (focused by most research).

- Fixed domain: $S$ does not vary but intensity functions goes to infinity (e.g. $\lambda(s) = b\lambda_0(s)$ as $b \to \infty$, rarely discussed).