Lecture 2: Poisson point processes: properties and statistical inference

Jean-François Coeurjolly

http://www-ljk.imag.fr/membres/Jean-Francois.Coeurjolly/
Definition, properties and simulation

Statistical inference for the homogeneous Poisson point process

Statistical inference for the inhomogeneous Poisson point process
Poisson point processes

2. Classical way of thinking of a Poisson point process

• ∀ m ≥ 1, ∀ bounded and disjoint B₁, ..., Bₘ ⊂ S , the r.v. X_{B₁}, ..., X_{Bₘ} are independent.

• N(B) ∼ P(∫₃ B ρ(u)du) for any bounded B ⊂ S.

1. Classical definition : X ∼ Poisson(S, ρ)

X is a Poisson point process on S ⊂ R^d |S| < ∞, with intensity function ρ (loc. integrable) if for any h : Nlf → R^+, B ⊂ S s.t. ∫₃ B ρ(u)du < ∞

E(h(X_B)) = ∑_{n≥0} e^{-∫₃ B ρ(u)du} n! ∫₃ B ... ∫₃ B h({x₁, ..., xₙ}) ∏_{i=1}^{n} ρ(x_i)dx_i.

Theorem

∀ S ⊂ R^d, X ∼ Poisson(S, ρ) exists and is uniquely determined by v(B) = exp(− ∫₃ B ρ(u)du), for any bounded B ⊂ S.
A few properties of Poisson point processes

Proposition: if \( \mathbf{X} \sim \text{Poisson}(S, \rho) \)

1. \( \mathbb{E} N(B) = \text{Var} N(B) = \int_B \rho(u) du \) which equals \( \rho |B| \) when \( \rho(\cdot) = \rho \) (homogeneous case, i.e. stationary and isotropic case if \( S = \mathbb{R}^d \)).

2. For any \( u, v \in S \), \( \rho^{(2)}(u, v) = \rho(u)\rho(v) \) (also valid for \( \rho^{(k)}, k \geq 1 \))

3. and if \( |S| < \infty \), \( \mathbf{X} \) admits a density w.r.t. Poisson\((S, 1)\) given by

\[
f(\mathbf{x}) = e^{|S| - \int_S \rho(u) du} \prod_{u \in \mathbf{x}} \rho(u).
\]

4. Slivnyak-Mecke Theorem: for any non-negative function \( h : S \times N_{lf} \rightarrow \mathbb{R}^+ \), then

\[
\mathbb{E} \sum_{u \in \mathbf{X}} h(u, \mathbf{X} \setminus u) = \int_S \mathbb{E}(h(u, \mathbf{X})) \rho(u) du.
\]
Exercises

Exercise 1: Proof of 2-4

- For property 2. Let $A, B \subset S$ and $\alpha^{(2)}(A \times B) = \mu(A)\mu(B)$.

- For property 3. Let $Y \sim \text{Poisson}(S, 1)$ and prove that 
  $\mathbb{E}(h(X)) = \mathbb{E}(f(Y)h(Y))$ for any $h : N_\text{lf} \times \mathbb{R}^+$. 

- For property 4. Assume $|S| < \infty$ and the use the integral definition.

Exercise 2.
Let $X$ be a homogeneous Poisson point process on $S = [0, 1]^2$.

- Compute the average number of $R$-closed neighbours of $X$ in $S$, ie. average number of points which have at least a distinct neighbour in $X$ at distance $\leq R$.

- Extend Slivnyak-Mecke Formula to compute expectations of the form $\sum_{u,v}^\neq h(u, v, X \setminus \{u, v\})$.

- To what corresponds the average of $\sum_{u,v}^\neq 1(d(u, X \setminus \{u, v\} \leq R), d(v, X \setminus \{u, v\} \leq R))$. Compute this expectation using the extended SM formula.
Simulation of a Poisson p.p. on a bounded domain

Let $W$ be the window of simulation.

- **Homogeneous case** is straightforward: $N(W) \sim \mathcal{P}(\rho|W|)$; let $n$ the realization; generate $x_1,\ldots,x_n$ independent uniform points in $W$; $x = \{x_1,\ldots,x_n\}$.

- **Inhomogeneous case**: a thinning procedure can be efficiently done if $\rho(u) \leq c$; let $y$ be a simulation of a Poisson$(c,W)$; delete a point $u \in y$ with prob. $1 - \rho(u)/c$.

Find the (most probable) related realization in $W = [-1, 1]^2$ for the intensities: $\rho(u) = \beta e^{-u_1^2 - 0.5u_3^2}$; $\rho = 200$; $\rho(u) = \beta e^{2\sin(4\pi u_1 u_2)}$ ($\beta$ is adjusted s.t. the mean number of points in $S$, $\int_W \rho(u)du = 200$.)
Statistical inference for a Poisson point process

- **Inference**: consists in estimating
  - $\rho$ : homogeneous case.
  - $\rho(\cdot; \theta)$ : parametric inhomogeneous function.
  - $\rho(u)$ : non parametric estimation

  - **Note**: All these estimates can be used even if the spatial point process is not Poisson (wait for a few slides ...)

- **Goodness-of-fit tests**: tests based on quadrats counting, based on the void probability,...
Homogeneous case

- We consider here the problem of estimating the parameter $\rho$ of a homogeneous Poisson point process defined on $S$ and observed on a window $W \subseteq S$.

- Since $N(W) \sim \mathcal{P}(\rho | W)$, the natural estimator of $\rho$ is

$$\hat{\rho} = \frac{N(W)}{|W|}$$

Properties
Homogeneous case

• We consider here the problem of estimating the parameter \( \rho \) of a homogeneous Poisson point process defined on \( S \) and observed on a window \( W \subseteq S \).

• Since \( N(W) \sim \mathcal{P}(\rho|W|) \), the natural estimator of \( \rho \) is

\[
\hat{\rho} = \frac{N(W)}{|W|}
\]

Properties

• (i) \( \hat{\rho} \) corresponds to the maximum likelihood estimate.
Homogeneous case

• We consider here the problem of estimating the parameter $\rho$ of a homogeneous Poisson point process defined on $S$ and observed on a window $W \subseteq S$.

• Since $N(W) \sim \mathcal{P}(\rho|W|)$, the natural estimator of $\rho$ is

$$\hat{\rho} = \frac{N(W)}{|W|}$$

Properties

• (i) $\hat{\rho}$ corresponds to the maximum likelihood estimate.

• (ii) $\hat{\rho}$ is unbiased.
Homogeneous case

- We consider here the problem of estimating the parameter $\rho$ of a homogeneous Poisson point process defined on $S$ and observed on a window $W \subseteq S$.
- Since $N(W) \sim \mathcal{P}(\rho | W|)$, the natural estimator of $\rho$ is
  \[
  \hat{\rho} = \frac{N(W)}{|W|}
  \]

Properties

- (i) $\hat{\rho}$ corresponds to the maximum likelihood estimate.
- (ii) $\hat{\rho}$ is unbiased.
- (iii) $\text{Var} \hat{\rho} = \frac{\rho}{|W|}$.

Proof: (i) follows from the definition of the density (ii-iii) can be checked using the Campbell formulae.
Homogeneous case (2)

Asymptotic results
Homogeneous case (2)

Asymptotic results

- For large $N(W)$, $\hat{\rho}|W| \simeq \mathcal{N}(\rho|W|, \rho|W|)$ and so

$$|W|^{1/2}(\hat{\rho} - \rho) \simeq \mathcal{N}(0, \rho).$$

(the approximation is actually a convergence as $W \rightarrow \mathbb{R}^d$)
Homogeneous case (2)

Asymptotic results

- For large $N(W)$, $\hat{\rho}|W| \approx \mathcal{N}(\rho|W|, \rho|W|)$ and so

$$|W|^{1/2}(\hat{\rho} - \rho) \approx \mathcal{N}(0, \rho).$$

(the approximation is actually a convergence as $W \rightarrow \mathbb{R}^d$)

- Variance stabilizing transform:

$$2|W|^{1/2}(\sqrt{\hat{\rho}} - \sqrt{\rho}) \approx \mathcal{N}(0, 1)$$
Homogeneous case (2)

Asymptotic results

- For large $N(W)$, $\hat{\rho}|W| \approx N(\rho|W|, \rho|W|)$ and so

  $$|W|^{1/2}(\hat{\rho} - \rho) \approx N(0, \rho).$$

  (the approximation is actually a convergence as $W \to \mathbb{R}^d$)

- Variance stabilizing transform:

  $$2|W|^{1/2}(\sqrt{\hat{\rho}} - \sqrt{\rho}) \approx N(0, 1)$$

- We deduce a $1 - \alpha$ ($\alpha \in (0, 1)$) confidence interval for $\rho$

  $$\text{IC}_{1-\alpha}(\rho) = \left(\sqrt{\hat{\rho}} \pm \frac{z_{\alpha/2}}{2|W|^{1/2}}\right)^2.$$
We generated $m = 10000$ replications of homogeneous Poisson point processes with intensity $\rho = 100$ on $[0, 1]^2$ (black plots) and on $[0, 2]^2$ (red plots).
A simulation example

We generated $m = 10000$ replications of homogeneous Poisson point processes with intensity $\rho = 100$ on $[0, 1]^2$ (black plots) and on $[0, 2]^2$ (red plots).

<table>
<thead>
<tr>
<th></th>
<th>$W = [0, 1]^2$</th>
<th>$W = [0, 2]^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Emp. Mean of $\hat{\rho}$</td>
<td>100.17</td>
<td>100.07</td>
</tr>
<tr>
<td>Emp. Var. of $\hat{\rho}$</td>
<td>98.57</td>
<td>25.69</td>
</tr>
<tr>
<td>Emp. Coverage rate of 95% confidence intervals</td>
<td>95.31%</td>
<td>94.78%</td>
</tr>
</tbody>
</table>


Quadrat counting based test

Performed as follows:

1. divide $W$ into quadrats $B_1, \ldots, B_m$.

2. count $n_j = n(x_{B_j})$ for $j = 1, \ldots, m$. Under the null hypothesis, then $n_j$ are realizations of $P(\rho|B_j|)$.

3. Construct the $\chi^2$ statistic

$$T = \sum_j \frac{(n_j - e_j)^2}{e_j} = \sum_j \frac{(n_j - \bar{\rho}|B_j|)^2}{\bar{\rho}|B_j|}$$

where $\bar{\rho} = n/|W|$

4. Under the null hypothesis if $T \sim \chi^2(m - 1)$

Example for the swedishpines dataset:

```
8 6 7
8 11 9
5 6 11
0.005 0.007 0.009
8 6 7
8 11 9
5 6 11
7.9 7.9 7.9
7.9 7.9 7.9
7.9 7.9 7.9
0.04 -0.67 -0.32
0.04 1.1 0.4
-1 -0.67 1.1
```
Applications in R

Pines datasets

- Consider the three unmarked datasets: japanesepines, swedishpines, finpines.
- Plot the data, estimate the intensity parameter.
- Construct a confidence interval for each of them. Which one is significantly the most abundant?
- Judge the assumption of the Poisson model using a GoF test based on quadrats.

bei dataset

Show that the bei (tropical forest dataset), the qk (earthquakes), the lightning strikes (for each summer) datasets cannot be modelled by a homogeneous Poisson point process.
Inhomogeneous case: parametric estimation

• Assume that $\rho$ is parametrized by a vector $\theta \in \mathbb{R}^p$ ($p \geq 1$). The most well-known model is the log-linear one:

$$\rho(u) = \rho(u; \theta) = \exp(\theta^\top z(u))$$

where $z(u) = (z_1(u), z_2(u), \ldots, z_p(u))$ corresponds to known spatial functions or spatial covariates.

• $\theta$ can be estimated by maximizing the log-likelihood on $W$

$$l_W(\mathbf{X}, \theta) = \sum_{u \in \mathbf{X}_W} \log \rho(u; \theta) + \int_W (1 - \rho(u; \theta)) du$$

$$= |W| + \sum_{u \in \mathbf{X}_W} \theta^\top z(u) - \int_W \exp(\theta^\top z(u)) du .$$

In other words

$$\hat{\theta} = \text{Argmax}_\theta \ell_W(\mathbf{X}, \theta) = \text{Argmin}_\theta \left\{ \frac{d}{d\theta} \ell_W(\mathbf{X}; \theta) \right\} .$$
A very short background on estimating equations

- \( \hat{\theta} = \text{Argmin} \, U_n(\theta), \, \theta_0 : \text{true parameter.} \)
- \( U_n(\theta) \) is said to define an unbiased estimating equation if \( \mathbb{E} \, U_n(\theta_0) = 0. \)
- We usually define
  (a) \( \Sigma = \text{Var}(U_n(\theta_0)) \) the variance-covariance matrix of \( U_n(\theta_0). \)
  (b) \( S = S(\theta_0) \) where \( S(\theta) = -\mathbb{E}\left( \frac{d}{d\theta} U_n(\theta) \right) \) the sensitivity matrix.
- If \( U_n(\theta_0) \overset{\text{approx.}}{\sim} \mathcal{N}(0, \Sigma), \) then under some technical and regularity assumptions

\[
\hat{\theta} - \theta_0 \overset{\text{approx.}}{\sim} \mathcal{N}(0, S^{-1}\Sigma S^{-1})
\]

Sketch of the proof: there exists \( \tilde{\theta} \) s.t. \( \|\tilde{\theta} - \theta_0\| \leq \|\hat{\theta} - \theta_0\| \)

\[
U_n(\theta_0) - U_n(\hat{\theta}) = -U_n(\theta_0) = (\hat{\theta} - \theta_0)S(\tilde{\theta}) \simeq (\hat{\theta} - \theta_0)S
\]

so \( \hat{\theta} - \theta_0 \approx S^{-1}U_n(\theta_0). \)
Inhomogeneous case: parametric estimation (2)

- The score function $s_W(X, \theta) = \frac{d}{d\theta} \ell_W(X, \theta)$ is an unbiased estimating equation! Indeed,
Inhomogeneous case: parametric estimation (2)

- The score function $s_W(X, \theta) = \frac{d}{d\theta} \ell_W(X, \theta)$ is an unbiased estimating equation! Indeed,

$$s_W(X, \theta) = \sum_{u \in X_W} z(u) - \int_W z(u) \exp(\theta^T z(u)) \, du = \rho(u)$$
Inhomogeneous case: parametric estimation (2)

- The score function \( s_W(X, \theta) = \frac{d}{d\theta} \ell_W(X, \theta) \) is an unbiased estimating equation! Indeed,

\[
s_W(X, \theta) = \sum_{u \in X_W} z(u) - \int_W z(u) \exp(\theta^T z(u)) \, du = \rho(u)
\]

and from Campbell Theorem
Inhomogeneous case: parametric estimation (2)

• The score function $s_W(X, \theta) = \frac{d}{d\theta} \ell_W(X, \theta)$ is an unbiased estimating equation! Indeed,

$$s_W(X, \theta) = \sum_{u \in X_W} z(u) - \int_W z(u) \exp(\theta^T z(u)) \, du$$

and from Campbell Theorem

$$E s_W(X, \theta) = \int_W z(u) \left( \exp(\theta_0^T z(u)) - \exp(\theta^T z(u)) \right) \, du = 0$$

when $\theta = \theta_0$.

• Exercise: verify that $\Sigma = S = \int_W z(u)z(u)^T \exp(\theta^T z(u))du$.

• Hence, $\hat{\theta} - \theta_0 \approx \mathcal{N}(0, S^{-1})$, i.e. $S(\hat{\theta})^{1/2}(\hat{\theta} - \theta_0) \approx \mathcal{N}(0, I_p)$. 
Inhomogeneous case: parametric estimation (2)

- The score function $s_W(X, \theta) = \frac{d}{d\theta} \ell_W(X, \theta)$ is an unbiased estimating equation! Indeed,

$$s_W(X, \theta) = \sum_{u \in X_W} z(u) - \int_W z(u) \exp(\theta^T z(u))\, du$$

and from Campbell Theorem

$$E s_W(X, \theta) = \int_W z(u) \left( \exp(\theta_0^T z(u)) - \exp(\theta^T z(u)) \right)\, du = 0$$

when $\theta = \theta_0$.

- Exercise: verify that $\Sigma = S = \int_W z(u)z(u)^T \exp(\theta^T z(u))\, du$.

- Hence, $\hat{\theta} - \theta_0 \approx \mathcal{N}(0, S^{-1})$, i.e. $S(\hat{\theta})^{1/2}(\hat{\theta} - \theta_0) \approx \mathcal{N}(0, I_p)$.

- Asymptotic results formalized by Rathbun and Cressie (1994).
Data example: dataset $\text{bei}$

A point pattern giving the locations of 3605 trees in a tropical rainforest. Accompanied by covariate data giving the elevation (altitude) ($z_1$) and slope of elevation ($z_2$) in the study region.

Assume an inhomogeneous Poisson point process (which is not true, see the next chapter) with intensity

$$\log \rho(u) = \beta + \theta_1 z_1(u) + \theta_2 z_2(u).$$

Question: how can we prove that each covariate has a significant and positive influence?
Nonparametric estimation (see e.g. Diggle 2003)

- Idea is to mimic the kernel density estimation to define a nonparametric estimator of the spatial function $\rho$.

- Let $k : \mathbb{R}^d \to \mathbb{R}^+$ a symmetric kernel with intensity one. Examples of kernels
  - Gaussian kernel : $(2\pi)^{-d/2} \exp(-\|y\|^2/2)$.
  - Cylindric kernel : $\frac{1}{\pi} 1(\|y\| \leq 1)$.
  - Epanecnikov kernel : $\frac{3}{4} 1(|y| < 1)(1 - |y|^2)$.

- Let $h$ be a positive real number (which will play the role of a bandwidth window), then the nonparametric estimate (with border correction) at the location $v$ is defined as

$$\hat{\rho}_h(v) = K_h(v)^{-1} \sum_{u \in X_W} \frac{1}{h^d} k\left(\frac{\|v - u\|}{h}\right), \quad K_h(v) = h^{-d} \int k\left(\frac{\|v - u\|}{h}\right) du$$

- Analogy with heatmaps!
Intuitively, this works . . .

Indeed, using the Campbell formula and a change of variables we can obtain

\[
E \tilde{\rho}_h(v) = K_h(v)^{-1} E \sum_{u \in X_W} \frac{1}{h^d} k \left( \frac{\|v - u\|}{h} \right) \\
= K_h(v)^{-1} \int_W \frac{1}{h^d} k \left( \frac{\|v - u\|}{h} \right) \rho(u) du \\
= K_h(v)^{-1} \int_{\frac{W - v}{h}} k \left( \|\omega\| \right) \rho(\omega h + v) d\omega \\
\text{h small} \quad \approx K_h(v)^{-1} \int_{\frac{W - v}{h}} k \left( \|\omega\| \right) \rho(v) d\omega \\
\approx \rho(v).
\]

More theoretical justifications and properties and a discussion on the bandwidth parameter and edge corrections can be found in Diggle (2003).
Application to the $qk$ dataset (quakes)