A Guide to NScluster: R Package for Maximum Palm Likelihood Estimation for Cluster Point Process Models using OpenMP

Ushio Tanaka
Osaka Prefecture University

Masami Saga
Indigo Corporation

Junji Nakano
The Institute of Statistical Mathematics

March 19, 2018

1 Preliminaries

A *point process* is a stochastic model governing the location of events in a given set. We consider the point process in a subset of Euclidean space. A *point pattern* is considered a realization of the point process. To analyze the point pattern, we first plot it as observed in the subset, which is considered an *observation window* denoted W. Following the preceding study, for simplicity, we restrict our discussion to W of a two-dimensional Euclidean space $\mathbb{R}^2$ to be standardized, i.e., a unit square ($W = [0,1] \times [0,1]$). In the following, we assume W to be a satisfied periodic boundary condition, i.e., W is considered a torus.

Throughout the package NScluster, we refer readers to Tanaka et al. [1] and [2] for details.

2 Overview of models

First, we consider a homogeneous Poisson point process with intensity μ to set a cluster center, where each point is referred to as a *parent point*. Each parent point generates a random number M of descendant points, which are realized independently and identically. Let ν be the expectation of M. The descendant points are distributed isotropically around each parent point, and the distances between each parent point and its descendant points are distributed independently and identically according to a *dispersal kernel* q_τ, where τ indicates the set of parameters of the dispersal kernel. The *Neyman-Scott cluster point process* is a union of all descendant points, with the exception of all parent points. In other words, the cluster process is unobservable for each cluster center. The Neyman-Scott cluster point process is also homogeneous, and its intensity λ equals $\mu\nu$.

We describe five cluster point process models, i.e., the Thomas and Inverse-power type models, and the generalized Thomas models of type A, B, and C.

2.1 Neyman-Scott cluster point process model

2.1.1 Thomas model

The *Thomas model* is the most popular Neyman-Scott cluster point process model. In this model, descendant points are scattered according to bivariate Gaussian distribution with zero mean and covariance matrix $\sigma^2 I$, $\sigma > 0$, where I is a 2×2 identity matrix. The corresponding dispersal kernel with $\tau = \sigma$ is given by

$$q_\sigma(r) = \frac{r}{\sigma^2} \exp\left(-\frac{r^2}{2\sigma^2}\right), \quad r \geq 0.$$

2.1.2 Inverse-power type model

The *Inverse-power type model* originated from the frequency of aftershocks per unit time interval (one day, one month, etc.), which has been referred to as the "modified Omori formula". The corresponding dispersal kernel with $\tau = (p, c)$ is given by

$$q_{(p,c)}(r) = \frac{c^{p-1}(p-1)}{(r+c)^p}, \quad r \geq 0,$$

where $p > 1$ and $c > 0$ imply the decay order and scaling with respect to the distance between each parent point and its descendant points, respectively. It can be inferred that the Inverse-power type model is more suitable for vaguely clustering point pattern data than the Thomas model.

2.1.3 Type A model

The *generalized Thomas model of type A* (*Type A model* for short) is a Neyman-Scott cluster point process model where the dispersal kernel is mixed by that of the two Thomas models with variable cluster sizes as follows:

$$q_{(a,\sigma_1,\sigma_2)}(r) = a q_{\sigma_1}(r) + (1-a) q_{\sigma_2}(r), \quad r \geq 0, \tag{1}$$

where a is a mixture ratio parameter with $0 < a < 1$. From Equation (1), it can be inferred that the Type A model is suitable for densely and vaguely clustering point pattern data to be fitted by mixing the Thomas model with the mixture ratio a.

2.2 Superposed Neyman-Scott cluster point process model

The extended Neyman-Scott cluster point process discussed in this article implies the superposition of the Neyman-Scott cluster point processes, i.e., the *superposed Neyman-Scott cluster point process*. Note that the extended cluster point process is not the Neyman-Scott cluster point process.

Here, we focus on the superposed Thomas model. The parameters to estimate are given by those of the two Thomas model: (μ_i, ν_i, σ_i), where $i = 1, 2$. Note that the intensity λ of the superposed Neyman-Scott cluster point process is given by $\mu_1\nu_1 + \mu_2\nu_2$.

2.2.1 Type B model

The *generalized Thomas model of type B* (*Type B model* for short) is a superposed Thomas model with $\nu_1 = \nu_2$.

2.2.2 Type C model

The *generalized Thomas model of type C* (*Type C model* for short) is a superposed Thomas model with $\nu_1 \neq \nu_2$.

3 Overview of functions

The package NScluster comprises three tasks, i.e., simulation, MPLE, and non-parametric and parametric Palm intensity comparison.

3.1 Simulation

The first and most intuitive step to understanding the model characteristics is to observe the data generated by the model. This can be realized by various simulation functions: `SimulateThomas`, `SimulateIP`, `SimulateTypeA`, `SimulateTypeB`, and `SimulateTypeC` for respective models.

3.2 MPLE

We briefly review Palm intensity of uniform and isotropic point processes on W. Translating each point of the point process into the origin $\boldsymbol{o} \in \mathbb{R}^2$, we obtain a superposed point process at $\boldsymbol{o}$. The superposed point process is referred to as a *difference process*, which is symmetric with respect to $\boldsymbol{o}$ as well as isotropic and inhomogeneous. Its intensity function is given by $N(W)$ times Palm intensity of the original one. In other words, the Palm intensity focuses on the difference process induced from pairwise coordinates of the original process rather than the original given point process.

Let N denote a counting measure. The *Palm intensity* $\lambda_{\boldsymbol{o}}$ is defined as follows:

$$\lambda_{\boldsymbol{o}}(\boldsymbol{x}) := \frac{\Pr(\{\, N(\boldsymbol{dx}) = 1 \mid N(\{\boldsymbol{o}\}) = 1 \,\})}{\mathrm{Vol}(\boldsymbol{dx})}, \tag{2}$$

where $\boldsymbol{dx}$ represents an infinitesimal set containing an arbitrary given point $\boldsymbol{x} \in W$.

Here, we examine Equation (2). $\lambda_{\boldsymbol{o}}$ implies the occurrence rate at an arbitrary given point x provided that a point is at $\boldsymbol{o}$. In this setting, r is the distance from $\boldsymbol{o}$ to $\boldsymbol{x}$, and $\lambda_{\boldsymbol{o}}$ depends on only r. Thus, for the Palm intensity of uniform and isotropic point processes, we obtain its polar coordinate representation with respect to distance r as follows:

$$\lambda_{\boldsymbol{o}}(\boldsymbol{x}) = \lambda_{\boldsymbol{o}}(r, \theta) = \lambda_{\boldsymbol{o}}(r), \quad 0 \leq r < \infty, \quad 0 \leq \theta < 2\pi.$$

The maximum Palm likelihood estimation procedure is based on the assumption that the difference process is well approximated by an isotropic and inhomogeneous Poisson point process with intensity function $N(W)\lambda_{\boldsymbol{o}}(r)$, which is centered at $\boldsymbol{o}$.

We are positioned to state the log-*Palm likelihood function*. Let $\boldsymbol{\theta}$ denote the parameter set of the cluster point process models. The log-Palm likelihood function, denoted $\ln L$ based on the Palm intensity $\lambda_{\boldsymbol{o}}$ (including $\boldsymbol{\theta}$) is given as follows:

$$\ln L(\boldsymbol{\theta}) = \sum_{i,j;i<j,r_{ij}\leq 1/2} \ln\left(N(W)\lambda_{\boldsymbol{o}}(r_{ij})\right) - 2\pi N(W) \int_0^{1/2} \lambda_{\boldsymbol{o}}(r) r \, dr, \tag{3}$$

where the summation is taken over each pair (i, j) such that the distance r_{ij} between distinct points x_i and x_j of the cluster point processes with $r_{ij} \leq 1/2$. Note that "1/2" in Equation (3) is due to the periodic boundary condition for $W = [0, 1] \times [0, 1]$.

The *maximum Palm likelihood estimate*s (*MPLE*s for short) are those that maximize Equation (3). Note that maximizing $\ln L(\boldsymbol{\theta})$ in Equation (3) to obtain MPLEs, $N(W)$ assigning the non-parametric part of Equation (3) is removable.

The package NScluster provides users with various parameter estimation functions: `EstimateThomas`, `EstimateIP`, `EstimateTypeA`, `EstimateTypeB`, and `EstimateTypeC` for respective models. The computation of MPLEs is implemented by simplex maximization with parallel computation via OpenMP.

3.3 Display of Palm intensity

To see adequateness of MPLEs, NScluster provides users with a non-parametric estimation of the Palm intensity. We can depict the Palm intensity of respective models using `PalmThomas`, `PalmIP`, `PalmTypeA`, `PalmTypeB`, and `PalmTypeC`.

References

[1] U. Tanaka, Y. Ogata and K. Katsura, Simulation and estimation of the Neyman-Scott type spatial cluster models, *Computer Science Monographs* **No.34,** 2008, 1–44: URL: http://www.ism.ac.jp/editsec/csm/

[2] U. Tanaka, Y. Ogata and D. Stoyan, Parameter estimation and model selection for Neyman-Scott point processes, *Biom. J.* **50,** 2008, 43–57.