NScluster: An R Package for Simulation and Estimation of the Neyman-Scott type spatial cluster models *

The Institute of Statistical Mathematics
and
Ushio Tanaka

September 5, 2012

1 Introduction

The Neyman-Scott spatial clustering point processes play an important role to model clustering point patterns in stochastic geometry and the theory of point processes. The areas of their applications are manifold. In fact, they include astronomy, (development) economics, ecology, forestry, geography, image analysis, spatial epidemiology and statistical seismology, and so forth.

In order to be advantageous to users who are unfamiliar with the `FORTRAN` language, we provide the present software package **NScluster** in terms of the `R` language for simulation and estimation of the Neyman-Scott type spatial cluster models. The corresponding source files are published in [1]. We develop a shared library or DLL from the original `FORTRAN` source programs.

Thus the package enables one to facilitate [1] providing the programs for the simulation methods of various Neyman-Scott spatial clustering point processes and their superpositions and for computing their maximum Palm-likelihood estimate, MPLE for abbreviation.

For more details of the simulation methods and the maximum Palm-likelihood procedure for their parameter estimation, see [1] and [2] and the references therein for theoretical information on stochastic geometry and the theory of point processes.

2 `R` functions in the NScluster package

Classes of tasks performed by this package are listed below:

*This research was partly supported by "Function and Induction Research Project" held by the Transdisciplinary Research Integration Center at the Research Organization of Information and Sciences, Japan.

2.1 Simulation of the Neyman-Scott spatial cluster models and their extensions

SimulateThomas, SimulateIP, SimulateTypeA, SimulateTypeB, SimulateTypeC

2.2 Parameter estimation by the simplex method

SimplexThomas, SimplexIP, SimplexTypeA, SimplexTypeB, SimplexTypeC

2.3 Non-parametric Palm intensity functions

PalmThomas, PalmIP, PalmTypeA, PalmTypeB, PalmTypeC

3 Overview of the models

The Neyman-Scott model is defined as the set of all cluster points (offspring points) described as follows:

1. Each cluster centre (parent point) formed to be a homogeneous Poisson point process with intensity μ produces a random number M of cluster points that are realized independently and identically according to the probabilities $\Pr\{M = m\}, m = 0, 1, \ldots$.

2. Locations of cluster points are distributed independently and identically according to a density function $q_\tau(x, y)$, where (x, y) is coordinates of cluster points relative to their cluster centre, and the parameter τ indicates a set of coefficients of the density function.

Hereafter, we consider parameters $\theta = (\mu, \nu, \tau)$, where $\nu = \mathbb{E}[M] = \Sigma m \Pr\{M = m\}$, as the parameters that characterize the various models of the Neyman-Scott type. In particular, throughout this article, we assume that M has the Poisson distribution with mean ν.

Furthermore, we consider multi-type cluster models, which are formed as follows [2]: the models allow two types of clusters with the different cluster sizes and distance scales. The intensities of cluster centres are μ_1 and μ_2, respectively. The corresponding cluster points of a representative cluster distribute according to Poisson random variables with means ν_1 and ν_2, respectively.

While such point patterns can be considered in spaces of arbitrary dimension, we here restrict ourself to planar point patterns in observed rectangular region $W = [0, T_x] \times [0, T_y]$, assuming the periodic boundary condition. Furthermore, we restrict ourself to the standardized case where $T_x = 1$ and $T_y \geq 1$ without loss of generality except for the scaling. All the models of point patterns are assumed to be homogeneous (stationary) and isotropic. Therefore the density function $q_\tau(x, y)$ is denoted by polar coordinate $q_\tau(r)$ of distance r from its cluster centre.

3.1 Thomas model

The *Thomas model* has the density function that is given by the bivariate Gaussian distribution with zero mean and the covariance matrix $\sigma^2 I$ where I is 2×2 identity matrix as follows (e.g., [2]). Due to isotropy this is described in polar coordinate as a function

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

of distance r from the corresponding parent's location.

3.2 Inverse-power model

The *Inverse-power type* density function of the distance is given as follows (see [1], [2]):

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

where the parameters p and c represent a decay order and scaling factor with respect to the distance, respectively.

3.3 Generalized Thomas model of type A

The *generalized Thomas model of type A* is given by the mixed Gaussian density function with σ_1 and σ_2 as follows (see [1], [2]):

$$q_\tau(r) = \left\{ \frac{ar}{{\sigma_1}^2} \exp\left(-\frac{r^2}{2{\sigma_1}^2}\right) + \frac{(1-a)r}{{\sigma_2}^2} \exp\left(-\frac{r^2}{2{\sigma_2}^2}\right) \right\}, \quad \tau = (a, \sigma_1, \sigma_2),$$

where the parameter a indicates the mixture ratio of the density function.

3.4 Generalized Thomas model of type B

We assume two types of clusters with the different sizes and distributions of the Thomas type. The intensities of the respective centres are μ_1 and μ_2, respectively. The respective clusters have the Poisson means of the cluster points ν_1 and ν_2. The density functions of the cluster point relative to respective cluster centre are given by

$$q_{\sigma_1}(r) = \frac{r}{{\sigma_1}^2} \exp\left(-\frac{r^2}{2{\sigma_1}^2}\right), \quad q_{\sigma_2}(r) = \frac{r}{{\sigma_2}^2} \exp\left(-\frac{r^2}{2{\sigma_2}^2}\right), \tag{1}$$

where σ_1 and σ_2 indicate their scales. If the mean sizes of the two types of clusters are the same, then it is referred to as the *generalized Thomas model of type B* (see [1], [2]).

3.5 Generalized Thomas model of type C

If $\nu_1 \neq \nu_2$ and each density function q_{σ_i}, $i = 1, 2$ have the same forms (1), then, in particular, for σ_1 and σ_2 we may constrain ν_1 and ν_2 to be

$$\nu_2 = \nu_1 \left(\frac{\sigma_2}{\sigma_1}\right),$$

and it is referred to as the *generalized Thomas model of type C* (see [1], [2]).

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.

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