Journal of Environmental Statistics
September 2018, Volume 8, Issue 3. http://www.jenvstat.org

Likelihood-based detection of cluster centers for Neyman–Scott point processes

Jiancang Zhuang
Institute of Statistical Mathematics, Research Organization of Information and Systems
Department of Statistical Science, Graduate University for Advanced Studies
London Mathematical Laboratory

Abstract

This study deals with the problem of estimating the unobservable cluster centers for a special type of Neyman–Scott point processes, in which the cluster sizes (numbers of members in each cluster) are distributed according to the Poisson distribution. The key point of the solution is the conversion among different forms of conditional intensities, $\lambda(t \mid \cdot)\,\mathrm{d}t = P\{N[t, t+\mathrm{d}t) = 1 \mid \cdot\} = \mathbf{E}[N[t, t+\mathrm{d}t) \mid \cdot]$, where $\cdot$ represents a σ-algebra generated by some information from the process N. Some recursive formulae associated with the filtering gain (information gain represented by the ratio of the likelihood of the point process when we know more information to the likelihood when we know less) are derived. These recursive equations can be solved numerically by using Monte Carlo integration. The proposed method is illustrated by two simulation experiments, a purely temporal and a multi-type spatiotemporal case.

Keywords: point process; conditional intensity; filtering; filtering gain; Neyman–Scott model.

1. Introduction

The Neyman–Scott process was used to describe the distribution of larvae by Neyman in 1939 in his paper *On a new class of "contagious" distributions, applicable in entomology and bacteriology*. Most people refer this model to Neyman and Scott (1953) or (1958). Later on, this important model and its modified versions have been widely used in many fields, for example, in astronomy for the distribution of galaxies (e.g., Neyman and Scott, 1958; Snethlage et al., 2002; Peebles, 2001), in ecology for the distributions of stands of trees in natural forests (e.g., Matern, 1971, 1986; Penttinen et al., 1992) and for the distribution of animals (e.g., Brown and Cowling, 1998), in seismology for modeling earthquake clusters

(see Lomnitz and Hax, 1966; Vere-Jones, 1970), in environmental studies (e.g., Barnett, 2005) especially for modeling the rainfall process (e.g., Guttorp, 1996; Cowpertwait, 1991; Cowpertwait and O'Connell, 1997), and in radiation physics (e.g., Lowen and Teich, 1991).

The statistical inferences in these applications are mostly associated with moment generating functionals (see, e.g., Diggle, 1983; Kerscher, 2001), moment estimates (see, e.g., Teich 1981; Lowen and Teich 1991; Cowpertwait, 1991; Brown and Cowling, 1998), K-functions (Dixon, 2002; Waagepetersen, 2007) or nearest-neighboring distances (Stoyan, 1992; Tanaka et al., 2008), and spectrums (Marrorquin et al., 1995). There have also been many other methods that are difficult to classify into the above categories. For example, Snethlage et al. (2002) used random shift to study the two-point correlation of a simple case of Neyman–Scott processes. Caears and Rombouts (2000) estimated the parameters for a Neyman–Scott model by minimizing a χ-squared type objective function related to the sample histogram of observations. Like most statistical analyses, likelihood-based inference methods are also always concerned. Lieshout and Baddeley (2001) obtained the likelihood for given observations of a Neyman-Scott process through Monte Carlo simulations. Castelloe and Zimmerman (2002) computed the likelihood directly by taking the expectation of all the possible configurations of the parent process and all the possible linkages between the parents and children, which is computationally intensive. Different from previous works, we introduce a new filtering technique for Neyman–Scott processes, with which we can obtain a simpler form of their likelihood functions, together with estimation of the intensity of cluster centers.

In this paper, we first provide an overview of some concepts associated with the Neyman–Scott processes or general point processes in Section 2. Section 3 gives the standard form of the likelihood for a point process in the form of the natural (internal) conditional intensity, which is the theoretical basis of this study. Section 4 derives the intensities of the Neyman–Scott point process and cluster-center process, conditioning on the observations of the process of the cluster members for one-dimensional cases. Sections 5 and 6 extend the methodology to higher-dimensional cases and cases of multi-type cluster centers, respectively. The method is illustrated by two numerical examples in Section 7.

2. Definitions and assumptions of Neyman–Scott processes

A point process on the Euclidean space $\mathbb{R}^d$ is usually defined as a random counting measure, i.e., a measurable mapping from the integer-valued measures on the $\mathbb{R}^d$ to the abstract probability space $(\Omega, \mathcal{F}, P)$. In this section, we consider the following type of temporal (one-dimensional) Neyman–Scott processes: the invisible cluster centers are assumed from a stationary Poisson process N^c with a rate $\lambda^c(t) = \mu$ on the real line, and each cluster has a size of a Poisson random variable S_i of mean A, with the members in the cluster independently distributed according to a density $g(\cdot - u)$, where u denotes the cluster center location. Such a Neyman–Scott process N is the superposition of all the in all the clusters. Let $\mathcal{G} \subset \mathcal{F}$ be any sub-σ-algebra on the abstract probability space. Define the $\mathcal{G}$-conditional intensity of N, $\lambda(t \mid \mathcal{G})$, by

$$\lambda(t \mid \mathcal{G})\,\mathrm{d}t = \mathbf{E}\,[N(\mathrm{d}t) \mid \mathcal{G}] = \mathrm{d}\mathbf{E}[N(t) \mid \mathcal{G}], \tag{1}$$

where $N(t) = N[0,t)$ for $t > 0$ and $N(t) = -N[-t,0)$ for $t \le 0$ and $N(\mathrm{d}t) = N[t, t+\mathrm{d}t)$. In this paper, we always use the notation $N(\mathrm{d}s) = N[s, s+\mathrm{d}s)$ to represent the number of events in N that fall in the infinitesimal interval $[s, s+\mathrm{d}s)$. As the process N is simple, $\lambda(t \mid \mathcal{G})$ can

also be defined by

$$\lambda(t \mid \mathcal{G})\,\mathrm{d}t = P\left\{N([t,\, t+\mathrm{d}t) = 1 \mid \mathcal{G}\right\}. \tag{2}$$

From the assumption that the occurrence of an event in the process N only depends on the locations of the parent events (cluster centers) from N^c and does not depend on the occurrences of any other events in N, for any σ-algebra $\mathcal{K}$, if $\mathcal{G} \subseteq \mathcal{K} \subseteq \sigma(\mathcal{G} \cup \sigma(N \setminus \{t\}))$, then

$$\lambda(t \mid \mathcal{K}) = \lambda(t \mid \mathcal{G}) = A \sum_{i:\, u_i \in N^c} g(t - u_i) = A \int_R g(t-u)\, N^c(\mathrm{d}u), \tag{3}$$

where A is the average number of cluster members produced in an arbitrary cluster center, and $g(\cdot - u)$ is the probability density for the occurrence times of cluster members that are generated by a cluster center event at time u.

3. Likelihoods

In the evaluation of the likelihood of a point process, the conditional intensity $\lambda(t \mid \mathcal{H}_t) = \mathbf{E}\left[\mathrm{d}N(t) \mid \mathcal{H}_t\right]/\mathrm{d}t$ of a point process N plays an important role, where $\mathcal{H}_t$ is the σ-algebra generated by the events in N that occur before t but not including t (e.g., Daley and Vere-Jones, 2003, Chap. 7). In this article, we call $\lambda(t \mid \mathcal{H}_t)$ the *natural conditional intensity* or the *internal conditional intensity* of N. Suppose that $\{t_i : i = \ldots, 0, 1, \ldots\}$ is the configuration of N. The likelihood of N in a given interval $[S, T]$ can be expressed through the natural conditional intensity by

$$\begin{aligned} L(N, [S,T], \mathcal{H}_t) &= \exp\left[\int_{[S,T]} \log\lambda(t \mid \mathcal{H}_t)\, N(\mathrm{d}t) - \int_{[S,T]} \lambda(t \mid \mathcal{H}_t)\,\mathrm{d}t\right]. \\ &= \exp\left[-\int_{[S,T]} \lambda(t \mid \mathcal{H}_t)\,\mathrm{d}t\right] \prod_{i:\, t_i \in N[S,T]} \lambda(t_i \mid \mathcal{H}_{t_i}). \end{aligned} \tag{4}$$

The above likelihood is indeed a likelihood conditional on a fixed history $\mathcal{H}_S$, i.e., $L(\mathcal{H}_{[S,T]} \mid \mathcal{H}_S)$, when $\mathcal{H}_{[S,T]}$ is the σ-algebra that consists of the information generated by N in the time interval $[S, T]$. Unfortunately, we cannot observe the history of the point process to the infinite past, but only starting from a certain time point. Assume that this starting time of an observation is 0. Denote by $\mathcal{R}_t$ the σ-algebra generated by the observed events. It is clear that $\mathcal{R}_t \subset \mathcal{H}_t$. Assume that $S > 0$, integrating over all the possibilities of the history before 0 on both sides in (4), the expected likelihood becomes

$$\begin{aligned} L(N, [S,T], \mathcal{R}_t) &= \exp\left[\int_{[S,T]} \log\lambda(t \mid \mathcal{R}_t)\, N(\mathrm{d}t) - \int_{[S,T]} \lambda(t \mid \mathcal{R}_t)\,\mathrm{d}t\right]. \\ &= \exp\left[-\int_{[S,T]} \lambda(t \mid \mathcal{R}_t)\,\mathrm{d}t\right] \prod_{i:\, t_i \in N[S,T]} \lambda(t_i \mid \mathcal{R}_{t_i}). \end{aligned} \tag{5}$$

If we set $S = 0$, the above likelihood is a complete likelihood.

4. Filtering formulae

To obtain the conditional intensity of a Neyman–Scott process N conditional on the information from observed events, $\lambda(t \mid \mathcal{R}_t)$, take the expectation of $\lambda(t \mid \sigma(\mathcal{G} \cup \mathcal{R}_t))$ conditioning on $\mathcal{R}_t$, i.e.,

$$\begin{aligned}
\lambda(t \mid \mathcal{R}_t) &= \mathbf{E}\left[\lambda(t \mid \sigma(\mathcal{G} \cup \mathcal{R}_t)) \mid \mathcal{R}_t\right] \\
&= \mathbf{E}\left[\lambda(t \mid \mathcal{G}) \mid \mathcal{R}_t\right] \qquad \text{(from (3))} \\
&= \mathbf{E}\left[\int_R A\, g(t-s)\, N^c(\mathrm{d}s) \,\middle|\, \mathcal{R}_t\right] \\
&= A \int_R g(t-s)\ \mathbf{E}\left[N^c(\mathrm{d}s) \mid \mathcal{R}_t\right] \\
&= A \int_R g(t-s)\, P\left\{N^c(s,\, s+\mathrm{d}s) = 1 \mid \mathcal{R}_t\right\} \\
&= A \int_R g(t-s)\, \lambda^c(s \mid \mathcal{R}_t)\, \mathrm{d}s. \qquad (6)
\end{aligned}$$

In the above equation, $\lambda^c(s \mid R_t)\,\mathrm{d}s$ is the probability that a cluster center falls in $[s, s+\mathrm{d}s)$, given the observations of N on $[0,\, t)$. Based on the Bayesian formula, we can derive formally

$$\begin{aligned}
\lambda^c(s \mid \mathcal{R}_t)\,\mathrm{d}s &= P\{N^c(\mathrm{d}s) = 1 \mid \mathcal{R}_t\} \\
&= \frac{P\{N^c(\mathrm{d}s) = 1 \wedge \mathcal{R}_t\}}{P\{\mathcal{R}_t\}} \\
&= \frac{P\{\mathcal{R}_t \mid N^c(\mathrm{d}s) = 1\} P\{N^c(\mathrm{d}s) = 1\}}{P\{\mathcal{R}_t\}} \\
&= \mu\,\mathrm{d}s\, \frac{P\{\mathcal{R}_t \mid N^c(\mathrm{d}s) = 1\}}{P\{\mathcal{R}_t\}} \quad (\text{by } \lambda^c(s) = \mu). \qquad (7)
\end{aligned}$$

In the above, the quantity

$$G(s \mid \mathcal{R}_t) = \frac{P\{\mathcal{R}_t \mid N^c(\mathrm{d}s) = 1\}}{P\{\mathcal{R}_t\}} \qquad (8)$$

is called the *filtering gain*. Note that $P\{\mathcal{R}_t\}$ is another notation of the likelihood of the observations in $[0, t)$. The filtering gain is the ratio of the likelihood of the observations before t given a cluster center at s to the corresponding likelihood of the observation but not conditional on the occurrence of a point in N^c at s. That is,

$$\begin{aligned}
G(s \mid \mathcal{R}_t) &= \frac{L(N, [0,t], \mathcal{R}_u \wedge N^c(\mathrm{d}s) = 1)}{L(N, [0,t], \mathcal{R}_u)} \\
&= \frac{\mathrm{e}^{-\int_0^t \lambda(u \mid N^c(\mathrm{d}s)=1, \mathcal{R}_u)\,\mathrm{d}u} \prod_{i:\, t_i \in [0,\,t]} \lambda(t_i \mid N^c(\mathrm{d}s) = 1, \mathcal{R}_{t_i})}{\mathrm{e}^{-\int_0^t \lambda(u \mid \mathcal{R}_u)\,\mathrm{d}u} \prod_{i:\, t_i \in [0,\,t]} \lambda(t_i \mid \mathcal{R}_{t_i})} \\
&= \mathrm{e}^{-\int_0^t [\lambda(u \mid N^c(\mathrm{d}s)=1, \mathcal{R}_u) - \lambda(u \mid \mathcal{R}_u)]\,\mathrm{d}u} \prod_{i:\, t_i \in [0,\,t]} \frac{\lambda(t_i \mid N^c(\mathrm{d}s) = 1, \mathcal{R}_{t_i})}{\lambda(t_i \mid \mathcal{R}_{t_i})}. \qquad (9)
\end{aligned}$$

Similar to (6),

$$\begin{aligned}\lambda(t \mid N^c(\mathrm{d}s) = 1, \mathcal{R}_t) &= \mathbf{E}\left[\int_R A\, g(t-u)\, \mathrm{d}N^c(u) \,\middle|\, N^c(\mathrm{d}s) = 1, \mathcal{R}_t\right] \\ &= A \int_{R\setminus \mathrm{d}s} g(t-u)\ \mathbf{E}\left[\mathrm{d}N^c(u) \mid N^c(\mathrm{d}s) = 1, \mathcal{R}_t\right] + A\, g(t-s) \\ &= A \int_R g(t-u)\, \lambda^c(u \mid N^c(\mathrm{d}s) = 1, \mathcal{R}_t)\, \mathrm{d}u + A\, g(t-s). \end{aligned} \tag{10}$$

In addition, similar to (7),

$$\begin{aligned} & \lambda^c(s_1 \mid \mathcal{R}_t,\, N^c(\mathrm{d}s) = 1)\, \mathrm{d}s \\ = \;& P\{N^c(\mathrm{d}s_1) = 1 \mid \mathcal{R}_t,\, N^c(\mathrm{d}s) = 1\} \\ = \;& \frac{P\{N^c(\mathrm{d}s_1) = 1 \wedge \mathcal{R}_t \mid N^c(\mathrm{d}s) = 1\}}{P\{\mathcal{R}_t \mid N^c(\mathrm{d}s) = 1\}} \\ = \;& \mu\, \mathrm{d}s\, \frac{P\{\mathcal{R}_t \mid N^c(\mathrm{d}s_1) = 1, N^c(\mathrm{d}s) = 1\}}{P\{\mathcal{R}_t \mid N^c(\mathrm{d}s_1) = 1\}}. \quad (\text{by } \lambda^c(s) = \mu) \\ = \;& \mu\, \mathrm{d}s\, \exp\left\{-\int_0^t [\lambda(u \mid N^c(\mathrm{d}s_1) = 1, N^c(\mathrm{d}s) = 1, \mathcal{R}_u) - \lambda(u \mid N^c(\mathrm{d}s) = 1, \mathcal{R}_u)]\, \mathrm{d}u\right\} \\ & \quad \times \prod_{i:\, t_i \in [0,\, t]} \frac{\lambda(t_i \mid N^c(\mathrm{d}s_1) = 1\}, N^c(\mathrm{d}s) = 1, \mathcal{R}_{t_i})}{\lambda(t_i \mid N^c(\mathrm{d}s) = 1, \mathcal{R}_{t_i})}. \end{aligned} \tag{11}$$

Similarly, we can obtain the following recursive formulae:

$$\begin{aligned} & \lambda(t \mid N^c(\mathrm{d}s_1) = \cdots = N^c(\mathrm{d}s_k) = 1, \mathcal{R}_t) \\ = \;& A \int_R g(t - s_{k+1})\, \lambda^c(s_{k+1} \mid N^c(\mathrm{d}s_1) = \cdots = N^c(\mathrm{d}s_k) = 1, \mathcal{R}_t)\, \mathrm{d}s_{k+1} + A \sum_{i=1}^{k} g(t - s_i); \end{aligned} \tag{12}$$

$$\begin{aligned} & \lambda^c(s_{k+1} \mid \mathcal{R}_t,\, N^c(\mathrm{d}s_1) = \cdots = N^c(\mathrm{d}s_k) = 1) \\ = \;& \mu\, \exp\left\{-\int_0^t \Big[\lambda(u \mid N^c(\mathrm{d}s_1) = \cdots = N^c(\mathrm{d}s_{k+1}) = 1, \mathcal{R}_u) \right. \\ & \qquad \left. - \lambda(u \mid N^c(\mathrm{d}s_1) = \cdots = N^c(\mathrm{d}s_k) = 1, \mathcal{R}_u)\Big]\, \mathrm{d}u\right\} \\ & \quad \times \prod_{i:\, t_i \in [0,\, t]} \frac{\lambda(t_i \mid N^c(\mathrm{d}s_1) = \cdots = N^c(\mathrm{d}s_{k+1}) = 1, \mathcal{R}_{t_i})}{\lambda(t_i \mid N^c(\mathrm{d}s_1) = \cdots = N^c(\mathrm{d}s_k) = 1, \mathcal{R}_{t_i})}. \end{aligned} \tag{13}$$

For numerical computation, such recursions can be stopped at a certain step, say k, by the approximation

$$\begin{aligned} & \lambda(t \mid N^c(\mathrm{d}s_1) = \cdots = N^c(\mathrm{d}s_{k+1}) = 1, \mathcal{R}_t) \\ \approx \;& A \int_R g(t-u)\, \lambda^c(u \mid N^c(\mathrm{d}s_1) = \cdots = N^c(\mathrm{d}s_k) = 1, \mathcal{R}_t)\, \mathrm{d}u + A \sum_{i=1}^{k+1} g(t - s_i). \end{aligned} \tag{14}$$

We can then substitute (14) sequentially to (13) and (12), and the recursion is solved.

If we stop the recursion at $k = 0$, i.e., starting from (10) and (6), reasonable (first-order) precision can be provided by

$$\lambda(t \mid N^c(\mathrm{d}s) = 1, \mathcal{R}_t) \approx A \int_R g(t-u)\, \lambda^c(u \mid \mathcal{R}_t)\, \mathrm{d}u + A\, g(t-s) = \lambda(t \mid \mathcal{R}_t) + A\, g(t-s).. \quad (15)$$

Substituting (15) into (9),

$$G(s \mid \mathcal{R}_t) \approx \mathrm{e}^{-A \int_0^t g(u-s)\, \mathrm{d}u} \prod_{i:\, t_i \in N[0,\, t)} \left[1 + \frac{A\, g(t_i - s)}{\lambda(t_i \mid \mathcal{R}_{t_i})}\right]. \quad (16)$$

Considering (16), (8), and (11),

$$\lambda^c(s \mid \mathcal{R}_t) \approx \mu\, \mathrm{e}^{-A \int_0^t g(u-s)\, \mathrm{d}u} \prod_{i:\, t_i \in N[0,\, t)} \left[1 + \frac{A\, g(t_i - s)}{\lambda(t_i \mid \mathcal{R}_{t_i})}\right]. \quad (17)$$

Substituting the above equation into (6), we have

$$\lambda(t \mid \mathcal{R}_t) \approx A\, \mu \int_R g(t-s)\, \mathrm{e}^{-A \int_0^t g(u-s)\, \mathrm{d}u} \prod_{i:\, t_i \in N[0,\, t)} \left[1 + \frac{A\, g(t_i - s)}{\lambda(t_i \mid \mathcal{R}_{t_i})}\right] \mathrm{d}s, \quad (18)$$

which can be used to obtain the conditional intensity of the process recursively. The initial conditions for the above equations are

$$\lambda^c(s \mid \mathcal{R}_t) = \lambda^c(s) = \mu, \qquad t \le 0 \quad (19)$$

and

$$\lambda(t \mid \mathcal{R}_t) = \mu\, A, \qquad t \le 0, \quad (20)$$

because $\mathcal{R}_t$ contains null information for $t \le 0$.

In summary, (17) and (18) give, respectively, the occurrence rates of clustering members and cluster centers in a Neyman–Scott process conditional on the observations of the clustering members in the process.

5. Multidimensional cases

The above results can be easily extended to multidimensional cases such as spatiotemporal and marked Neyman–Scott processes. Consider a Neyman–Scott process N in $\mathbb{R} \times \mathbb{R}^d$ ($d \ge 1$). We suppose that its cluster-center process N^c has an intensity

$$\lambda^c(t, x \mid \mathcal{K}) = \mu(t, x), \quad (21)$$

where x represents that the spatial or the mark component, and the intensity function of N conditional on the σ-algebra $\mathcal{G}$ generated by the configuration of N^c is

$$\lambda(t, x \mid \mathcal{G}) = A \int_{\mathbb{R} \times \mathbb{R}^d} g(t-s, x-y)\, N^c(\mathrm{d}s \times \mathrm{d}y). \quad (22)$$

Denote the σ-algebra generated by the observation of N before t but not including t by $\mathcal{R}_t$. In practice, the observation of N can only be started from a certain time, which is assumed

to be 0 here, and for each time the observation is limited within a bounded region $S(t)$. For simplification, we use $D(t) = \bigcup_{u\in[0,t)}(\{u\} \times S(u))$, i.e., the spatiotemporal range that the observation covers. With the above notation, the conditional intensity of N given $\mathcal{R}_t$ is

$$\begin{aligned}
\lambda(t, x \mid \mathcal{R}_t) &= \mathbf{E}[\lambda(t, x \mid \mathcal{G}, \mathcal{R}_t) \mid \mathcal{R}_t] \\
&= \mathbf{E}\left[A \int_{\mathbb{R}\times\mathbb{R}^d} g(t-s, x-y)\, N^c(\mathrm{d}s \times \mathrm{d}y) \,\middle|\, \mathcal{R}_t \right] \\
&= A \int_{\mathbb{R}\times\mathbb{R}^d} g(t-s, x-y)\ \mathbf{E}\left[N^c(\mathrm{d}s \times \mathrm{d}y) \mid \mathcal{R}_t\right] \\
&= A \int_{\mathbb{R}\times\mathbb{R}^d} g(t-s, x-y)\, \lambda^c(s, y \mid \mathcal{R}_t)\, \ell(\mathrm{d}s \times \mathrm{d}y). \qquad (23)
\end{aligned}$$

With similar calculations as for the one-dimensional case,

$$\lambda^c(s, y \mid \mathcal{R}_t) = \mu(s, y)\, G(s, y \mid \mathcal{R}_t), \qquad (24)$$

where the filtering gain

$$G(s, y \mid \mathcal{R}_t) = \frac{P\{\mathcal{R}_t \mid N^c([s, s+\mathrm{d}s) \times [y, y+\mathrm{d}y)) = 1\}}{P\{\mathcal{R}_t\}}, \qquad (25)$$

i.e., the ratio of the likelihood for the observation before t given a cluster center occurring at (s, y) and the likelihood for the same observation without such a condition. Thus, the filtering gain can be written in full as follows:

$$\begin{aligned}
G(s, y \mid \mathcal{R}_t) &= \exp\left\{ -\int_{D(t)} [\lambda(u, z \mid N^c(\mathrm{d}s \times \mathrm{d}y) = 1, \mathcal{R}_u) - \lambda(u, z \mid \mathcal{R}_u)]\, \ell(\mathrm{d}u \times \mathrm{d}z) \right\} \\
&\quad \times \prod_{i:\,(t_i,x_i)\in N(D(t))} \frac{\lambda(t_i, x_i \mid N^c(\mathrm{d}s \times \mathrm{d}y) = 1, \mathcal{R}_{t_i})}{\lambda(t_i, x_i \mid \mathcal{R}_{t_i})} \qquad (26) \\
&\approx \exp\left\{ -A \int_{D(t)} g(u-s, z-y)\, \ell(\mathrm{d}u \times \mathrm{d}z) \right\} \\
&\quad \times \prod_{i:\,(t_i,x_i)\in N(D(t))} \left[1 + \frac{A g(t_i - s, x_i - y)}{\lambda(t_i, x_i \mid \mathcal{R}_{t_i})} \right]. \qquad (27)
\end{aligned}$$

The initial conditions for the above equations are

$$\lambda^c(u, y \mid \mathcal{R}_t) = \lambda^c(u, y) = \mu(t, y), \qquad t \le 0 \qquad (28)$$

and

$$\lambda(t, x \mid \mathcal{R}_t) = A \int_{\mathbb{R}\times\mathbb{R}^d} g(t-s, x-y)\, \mu(s, y)\, \ell(\mathrm{d}s \times \mathrm{d}y), \qquad t \le 0, \qquad (29)$$

because $\mathcal{R}_t$ contains null information for $t \le 0$.

6. Multi-type Neyman–Scott processes

The problem of parameter estimation of multi-type Neyman–Scott processes was introduced by Tanaka et al. (2007). In a multi-type Neyman–Scott process, the cluster centers are

classified into two or more classes, and different types of centers have different average numbers and different location distributions of cluster members. For simplification, we only give the solution for the one-dimensional two-type Neyman–Scott processes, and higher-dimensional processes with more than two types of cluster centers can be solved in the same way.

Suppose that the cluster centers are classified as either Type I or Type II. The process N_1^c of Type I cluster centers is a Poisson process with a rate $\mu_1(t)$ and the process N_2^c of Type II cluster centers is also a Poisson process with a rate $\mu_2(t)$. The intensity of the children process N, given all the locations of events in N_1^c and N_2^c, is

$$\lambda(t \mid \mathcal{G}_1, \mathcal{G}_2) = A_1 \int_R g_1(t-s)\, N_1^c(\mathrm{d}s) + A_2 \int_R g_2(t-s)\, N_2^c(\mathrm{d}s), \tag{30}$$

where $\mathcal{G}_1$ and $\mathcal{G}_2$ are the σ-algebra generated by N_1^c and N_2^c, respectively. Similar to the one-dimensional case in Section 4, the conditional intensity of N given the observation history R_t, i.e., the σ-algebra generated by N up to time t but not including t, is

$$\begin{aligned}
\lambda(t \mid \mathcal{R}_t) &= \mathbf{E}[\lambda(t \mid \mathcal{G}_1, \mathcal{G}_2) \mid \mathcal{R}_t] \\
&= A_1 \int_R g_1(t-s)\ \mathbf{E}[N_1^c(\mathrm{d}s) \mid \mathcal{R}_t] + A_2 \int_R g_2(t-s)\ \mathbf{E}[N_2^c(\mathrm{d}s) \mid \mathcal{R}_t] \\
&= A_1 \int_R g_1(t-s)\, \lambda_1^c(s \mid \mathcal{R}_t)\, \mathrm{d}s + A_2 \int_R g_2(t-s)\, \lambda_2^c(s \mid \mathcal{R}_t)\, \mathrm{d}s.
\end{aligned} \tag{31}$$

Again, we have

$$\lambda_1^c(s \mid \mathcal{R}_t) = \mu_1(s)\, G_1(s \mid \mathcal{R}_t) \quad \text{and} \quad \lambda_2^c(s \mid \mathcal{R}_t) = \mu_2(s)\, G_2(s \mid \mathcal{R}_t), \tag{32}$$

where G_1 and G_2 are the filtering gains for N_1^c and N_2^c at s given the observation history up to t, respectively. It is easy to derive

$$\begin{aligned}
G_1(s \mid \mathcal{R}_t) &= \frac{P\{\mathcal{R}_t \mid N_1^c(\mathrm{d}s) = 1\}}{P\{\mathcal{R}_t\}} \\
&= \exp\left\{ - \int_0^t [\lambda(u \mid \mathcal{R}_u, N_1^c(\mathrm{d}s) = 1) - \lambda(u \mid \mathcal{R}_u)]\, \mathrm{d}t \right\} \\
&\qquad \times \prod_{i:\, t_i \in N[0,t)} \frac{\lambda(t_i \mid \mathcal{R}_{t_i}, N_1^c(\mathrm{d}s) = 1)}{\lambda(t_i \mid \mathcal{R}_{t_i})}.
\end{aligned} \tag{33}$$

Using the first-order approximations $\lambda_1^c(u \mid \mathcal{R}_t, N_k^c(\mathrm{d}s) = 1) \approx \lambda_1^c(u \mid \mathcal{R}_t)$ and $\lambda_2^c(u \mid \mathcal{R}_t, N_k^c(\mathrm{d}s) = 1) \approx \lambda_2^c(u \mid \mathcal{R}_t)$, where k is either 1 or 2, we can obtain

$$\begin{aligned}
\lambda(t \mid \mathcal{R}_t, N_1^c(\mathrm{d}s) = 1) \approx\ & A_1 \int_R g_1(t-u)\, \lambda_1^c(u \mid \mathcal{R}_t)\, \mathrm{d}u + A_2 \int_R g_2(t-u)\, \lambda_2^c(u \mid \mathcal{R}_t)\, \mathrm{d}u \\
& + A_1\, g_1(t-s)
\end{aligned} \tag{34}$$

and

$$\begin{aligned}
\lambda(t \mid \mathcal{R}_t, N_2^c(\mathrm{d}s) = 1) \approx\ & A_1 \int_R g_1(t-u)\, \lambda_1^c(u \mid \mathcal{R}_t)\, \mathrm{d}u + A_2 \int_R g_2(t-u)\, \lambda_2^c(u \mid \mathcal{R}_t)\, \mathrm{d}u \\
& + A_2\, g_2(t-s).
\end{aligned} \tag{35}$$

Substitute the above two equations into (33),

$$G_1(s \mid \mathcal{R}_t) \approx \exp\left\{-A_1 \int_0^t g_1(u-s)\,\mathrm{d}u\right\} \prod_{i:\, t_i \in N[0,t)} \left[1 + \frac{A_1\, g_1(t_i - s)}{\lambda(t_i \mid \mathcal{R}_{t_i})}\right]. \tag{36}$$

Similarly

$$G_2(s \mid \mathcal{R}_t) \approx \exp\left\{-A_2 \int_0^t g_2(u-s)\,\mathrm{d}u\right\} \prod_{i:\, t_i \in N[0,t)} \left[1 + \frac{A_2\, g_2(t_i - s)}{\lambda(t_i \mid \mathcal{R}_{t_i})}\right]. \tag{37}$$

We can use the above two equations and (32) to compute the conditional intensities of the processes of the cluster centers from each type.

7. Examples

7.1. One-dimensional cases

To illustrate the above procedures, we simulate a Neyman–Scott process with the cluster-center process unobservable. We first simulate a Poisson process with a rate of 0.5 on the interval $[-1000, 1000]$ as the cluster-center process. For each event in the cluster-center process, the total number of its children is a Poisson random variable of mean 3, and their locations are distributed according to a normal density centered at the cluster center with a standard deviation of 0.5. We assume the process is observed in the interval [0, 100].

The conditional intensity is calculated by using (18) and Monte Carlo integration. That is, for each t, generate $K = 1{,}000{,}000$ samples, $s_1, s_2, \ldots, s_K$, of s according to the density $g(t-s)$, then

$$\lambda(t \mid \mathcal{R}_t) \approx \frac{A\,\mu}{K} \sum_{j=1}^{K} G(t_i' \mid \mathcal{R}_t), \tag{38}$$

where $G(s \mid \mathcal{R}_t)$ is calculated based on (16). Once the conditional intensity $\lambda(t_i \mid \mathcal{R}_{t_i})$ at the occurrence times of each observed event is calculated, $\lambda^c(s \mid \mathcal{R}_t)$ can be easily obtained by using (17).

The outputs are shown in Figure 1. We can make the following observations.

1. The conditional intensity $\lambda(t \mid \mathcal{R}_t)$ of the process is similar in appearance to the internal conditional intensity of the epidemic type aftershock sequence (ETAS) model (see Ogata, 1988) or Hawkes' process (Hawkes, 1977). This is because, in (18), $\lambda(t \mid \mathcal{R})$ is continuous between the occurrence times of two adjacent events and takes a jump at the occurrence of each new event.

2. We can imagine that the expected rate of the cluster center process conditioning on the previously observed events, $\lambda^c(t \mid \mathcal{R}_t)$, takes a similar shape by replacing s in (17) with t. Unlike the internal conditional intensity of the ETAS model or Hawkes' process, which is additive, $\lambda^c(t \mid \mathcal{R}_t)$ is log-additive.

3. The conditional intensity $\lambda^c(t \mid \mathcal{R}_{100})$ in Figure 1(d) gives an estimate of the occurrence of a cluster center at t, when the observation of the Neyman–Scott process in $[0, 100]$ is

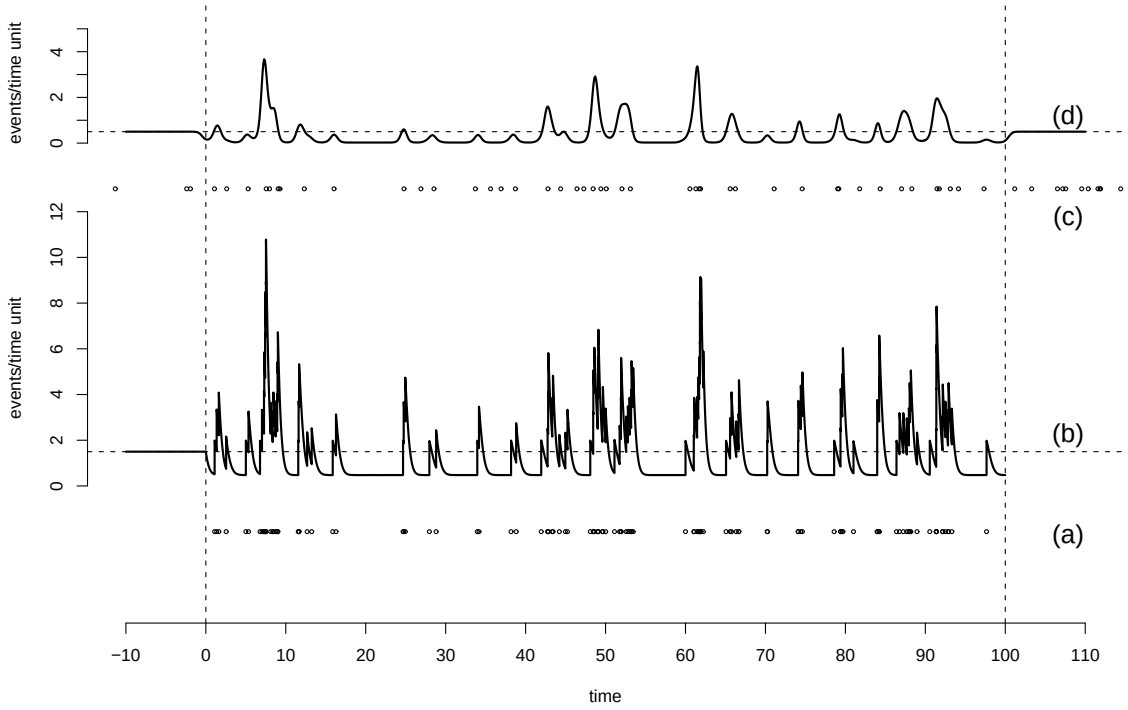

Figure 1: Channel (a) gives the locations of points in the observation period [0, 100]. Channel (b) gives the conditional intensity $\lambda(t \mid \mathcal{R}_t)$ calculated by using (18). Channel (c) gives the locations of unobserved clusters. Channel (d) gives the intensity of the process of cluster centers conditional on all the observation data of N in the period [0, 100], i.e., $\lambda^c(t \mid \mathcal{R}_{100})$.

known. When $t < 0$ or $t > 100$, its value is almost the same as μ because no observation data are available near such times. In $[0, 100]$, it grows high near the occurrences times of cluster centers and drops to low levels when cluster centers are absent. These results imply that the proposed filtering method provides good estimates of the unobserved centers.

7.2. Multi-type two-dimensional cases

Let us consider a more complicated example. Assume that there are two types of cluster centers from Poisson point processes, N_1^c and N_2^c, with rates of 0.03 and 0.01, respectively. The observation time–space window is a rectangle $(0, 200) \times (0, 100)$. We set the parameters $A_1 = 10$ and $A_2 = 30$. The probability densities for the occurrence times and locations of cluster members are

$$g_1(t, x) = \frac{1}{8\pi\sqrt{2}} \exp\left[-\frac{t^2}{8} - \frac{x^2}{32}\right] \tag{39}$$

and

$$g_2(t, x) = \frac{12}{\pi} \frac{1}{16 + t^2} \frac{1}{9 + x^2}. \tag{40}$$

The simulation of N, N_1^c, and N_2^c are illustrated in Figure 2(a).

For each (t, x), generate K_1 samples, (s'_j, x'_j), $j = 1, \ldots, K_1$, according to the density of

$g_1(s'-t, x'-x)$ and K_2 samples (t'_j, y'_j), $j = 1, \ldots, K_2$, according to $g_2(t'-t, y'-x)$, then

$$\lambda(t, x \mid \mathcal{R}_t) = \frac{A_1\,\mu_1}{K_1} \sum_{j=1}^{K_1} G_1(s'_i, x'_i \mid \mathcal{R}_t) + \frac{A_2\,\mu_2}{K_2} \sum_{j=1}^{K_2} G_2(t'_i, y'_i \mid \mathcal{R}_t), \tag{41}$$

where

$$\begin{aligned} G_1(s', x' \mid \mathcal{R}_t) \;&\approx\; \exp\left\{-A_1 \int_0^{100} \int_0^t g_1(u-s', y-x')\,\mathrm{d}u\,\mathrm{d}y\right\} \\ &\quad \times \prod_{i:\, t_i \in N[0,t)} \left[1 + \frac{A_1\, g_1(t_i - s', x_i - x')}{\lambda(t_i \mid \mathcal{R}_{t_i})}\right] \end{aligned}$$

and

$$\begin{aligned} G_2(t', y' \mid \mathcal{R}_t) \;&\approx\; \exp\left\{-A_2 \int_0^{100} \int_0^t g_2(u-t', y-y')\,\mathrm{d}u\,\mathrm{d}y\right\} \\ &\quad \times \prod_{i:\, t_i \in N[0,t)} \left[1 + \frac{A_2\, g_2(t_i - t', x_i - y')}{\lambda(t_i \mid \mathcal{R}_{t_i})}\right]. \end{aligned}$$

We implemented the computation in R. Figure 2(b) gives the contour image of $\lambda(t, x \mid \mathcal{R}_t)$ on a logarithmic scale. We can see that before the observation starts, i.e., when $t < 0$, $\lambda(t \mid \mathcal{R}_t)$ takes its unconditional expectation $\mu_1 A_1 + \mu_2 A_2$, because there is no observational information that can be used in predicting the occurrence of future events of N. Once an event is observed, $\lambda(t \mid \mathcal{R}_t)$ is updated by an increment meaning that the possibility (risk) that one or more events are following is increased. Such an "exciting" effect decays with distance and time. Figure 2(c) shows the estimated intensity of the unobserved cluster-center process N_1^c, given that the locations and times of events from N in the space–time window. The locations and times of N_1^c events and neighborhoods are marked by high intensities of $\lambda_1^c(t, x \mid \mathcal{R}_{200})$. It is remarkable that $\lambda_1^c(t, x \mid \mathcal{R}_{200})$ also has high values around the times and locations of N_2^c events. This is not surprising, as a cluster produced by an N_2^c event has a heavier tail decaying in space–time and a larger number of members, which is similar to the overlapping of clusters produced by several N_1^c events if they are close to each other. Such a phenomenon does not appear in Figure 2(d). High values of λ_2^c exist only around the locations and times of unobserved N_2^c events but not N_1^c events. In summary, $\lambda(t, x \mid \mathcal{R}_t)$ gives the predicted risk of an event occurring in N in the very near future and nearby, and $\lambda_1^c(t, x \mid \mathcal{R}_{200})$ and $\lambda_2^c(t, x \mid \mathcal{R}_{200})$ show their good abilities in estimating the locations of corresponding types of unobserved cluster centers.

8. Concluding remarks

Compared with the existing inference methods for Neyman–Scott processes, this paper has introduced a more natural approach of inference to different cases of this kind of processes, including higher-dimensional and multi-type cases. All the conditional intensities in this paper can be written in the form

$$\lambda(t \mid \cdot)\,\mathrm{d} = P\{N[t, t+\mathrm{d}t) = 1 \mid \cdot\} = \mathbf{E}[N[t, t+\mathrm{d}t) \mid \cdot], \tag{42}$$

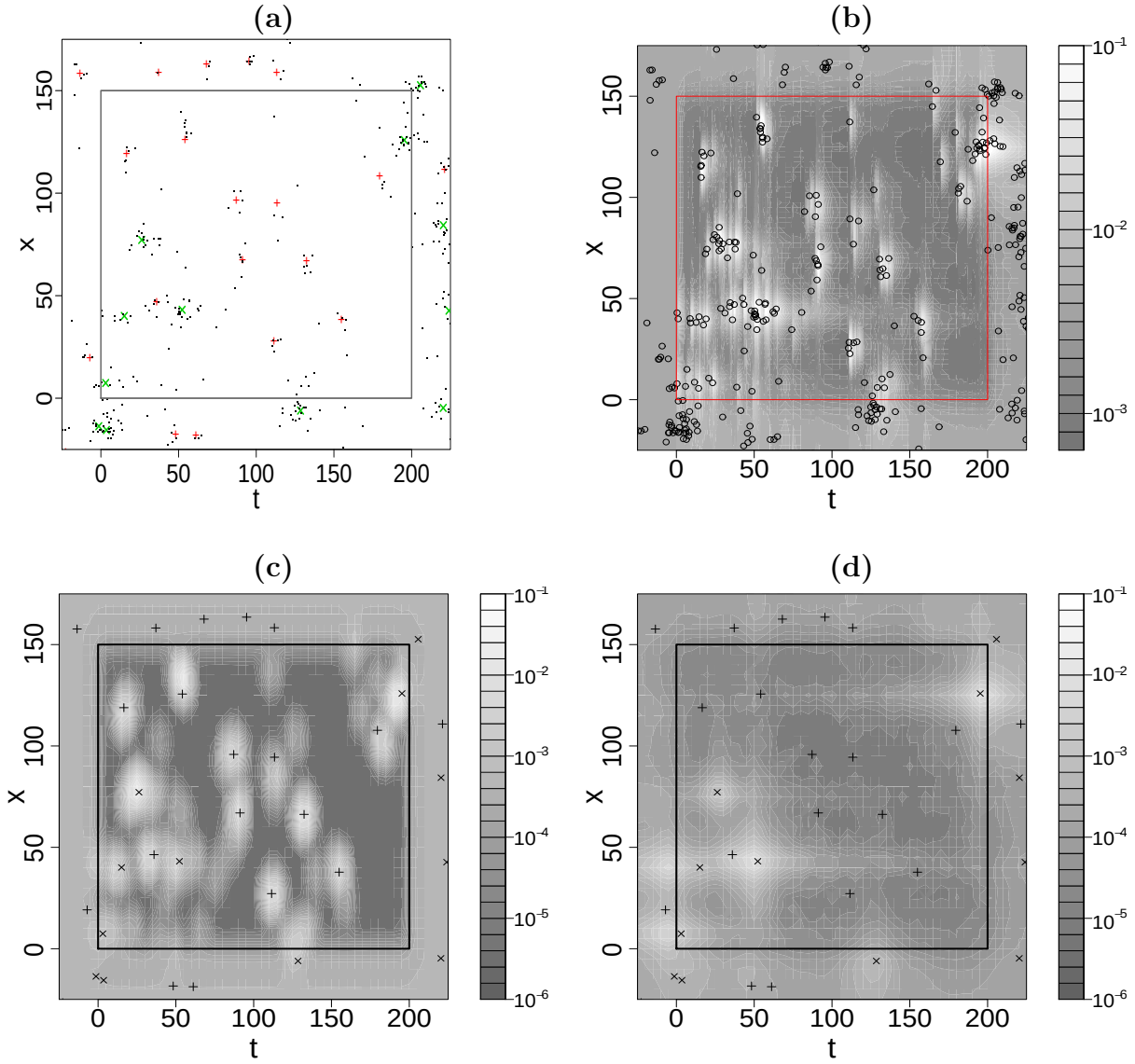

Figure 2: (a) Locations of simulated points (dots), Type I cluster centers (+'s) and Type II cluster centers (×). (b) Left-conditional intensity $\lambda(t, x \mid \mathcal{R}_t)$ with simulated points in N marked by small circles. (c) Estimated intensity for the Type I cluster process (+), $\lambda_1^c(t \mid \mathcal{R}_{200})$. (d) Estimated intensity for the Type II cluster process (×), $\lambda_2^c(t \mid \mathcal{R}_{200})$. The rectangle, $[0, 200] \times [0, 150]$, in each panel represents the window of observations.

where $\cdot$ represents a wildcard σ-algebra generated by some information associated with the process N. For example, $\lambda(t \mid \mathcal{R}_t)$ can be used for real-time prediction of future events and for calculating the likelihood, $\lambda(t, \cdot \mid \mathcal{R}_{t+s})$, $s > 0$, gives the re-evaluation of risk at time t when more information is known, and $\lambda^c(t, x \mid \mathcal{R})$ provides an estimate of the locations of cluster centers given the observational information contained in $\mathcal{R}$. The core of this method is the recursive formulae associated the filtering gain (information gain represented by the ratio of the likelihood of the point process when we know more information to the likelihood when we know less). In this study, we have discussed both the temporal and spatiotemporal/marked cases. For purely spatial Neyman–Scott processes, the solution can be obtained by simply regarding one of the axes as the time axis and then apply the same algorithm as in Section 5. Numerical solution of the above recursions can even be implemented in a computer language as slow as R with sufficient precisions and acceptable computational times (complexity).

Acknowledgements

Encouragement from and helpful discussions with Frederick Schoenberg, Yosihiko Ogata, and David Vere-Jones are gratefully acknowledged.

References

Barnett V (2005). *Environmental Statistics: Methods and Applications.* John Wiley & Sons, Inc., Chichester.

Brown BM, Cowling A (1998). "Clustering and abundance estimation for Neyman–Scott models and line transect surveys." *Biometrika*, **85**, 427–438.

Caears J, Rombouts L (2000). "Geostatistical modelling of offshore diamond deposits". In W Kleingeld, D Krige (eds.), "Proceeding to the 6th International Geostatistics Congress," Cape Town, South Africa.

Castelloe JM, Zimmerman DL (2002). "On Maximum Likelihood Estimation of a Spatial Poisson Cluster Process." Technical report No 312, Department of Statistics and Actuarial Science, the University of Iowa.

Cowpertwait PSP (1991). "Further developments of the Neyman–Scott cluster point process for modelling rainfall." *Water Resource Research*, **27**, 1431–1438.

Cowpertwait PSP, O'Connell PE (1997). "A regionalised Neyman–Scott model of rainfall with convective and strationform cells." *Hydrology and Earth System Science*, **1**, 71–80.

Daley DD, Vere-Jones D (2003) *An Introduction to Theory of Point Processes – Volume 1: Elementary Theory and Methods (2nd Edition).* Springer, New York.

Diggle PJ (1983). *Statistical Analysis of Spatial Point Patterns.* Academic Press, London, UK.

Dixon PM (2002). "Ripley's K function". In El-Shaarawi AH, Walter WP (eds.), *Encyclopedia of Environmetrics.* John Wiley & Sons, Ltd., Chichester.

Guttorp, P (1996). "Stochastic Modeling of Rainfall". In Wheeler, MF. (eds.), *Environmental Studies*, 171-187. Springer, New York.

Kerscher M (2001). "Constructing, characterizing, and simulating Gaussian and higher-order point distributions". *Physical Review E*, 64, 05619.

van Lieshout MNM, Baddeley AJ (2001). "Extrapolating and interpolating spatial patterns." In Lawson AB, Denison DGT (eds.), *Spatial Cluster Modelling,* , 61–86, Chapman and Hall/CRC Press, Boca Raton.

Lomnitz C, Hax A (1966). "Clustering in aftershocks sequences". In Steinhart JS, Jefferson-Smith T (eds.), *The Earthquake Beneath the Continents, Geophysical Monograph Series, Vol. 10*, 502-508, American Geophysical Union, Washington.

Lowen SB, Teich MC (1991). "Doubly stochastic Poisson point process driven by fractal shot noise." *Physical Review A*, **43**, 4192–4215.

Marrorquin A, Garcia JA, Garrido J, Mateos VL (2002). "Neyman–Scott cluster model for daily rainfall processes in low extremadura (Spain): rainfall generating mechanisms." *Theory and Application of Climatology*, **52**, 183–193.

Matern B (1971). "Doubly stochastic Poisson processes in the plan." In Patil G (ed), *Statistical Ecology, Vol. 1*, Penn State Univ Press, University Park and London.

Matern B (1986). *Spatial Variation. Lecture Notes in Statistics*, **36**. Springer, Berlin Heidelberg, New York.

Neyman J (1939). "On a new class of 'contagious' distributions, applicable in entomology and bacteriology." *The Annals of Mathematical Statistics*, **10**, 35–57.

Neyman JE, Scott EL (1953). "Frequency of separation and interlocking of clusters of galaxies." *Proceedings of the National Academy of Sciences of the United States of America*, **39**, 737–743.

Neyman JE, Scott EL (1958). "A statistical approach to problems of cosmology." *Journal of the Royal Statistical Society, Series B (Methodological)*, **20**, 1–43.

Ogata Y (1988). "Statistical models for earthquake occurrences and residual analysis for point processes." *Journal of American Statistical Association, Application*, **83**, 9–27.

Peebles PJE (2001). "The galaxy and mass N-point correlation functions: a blast from the past." In Martinez VJ, Trimble V, Pons-Borderia MJ (eds.), *Historical Development of Modern Cosmology*, ArXiv: astro-ph/0103040. ISBN: 1-58381-092-7.

Penttinen AK, Stoyan D, Henttonen HM (1992). "Marked point processes in forest statistics." *Forest Science*, **38**, 806-824.

Snethlage M, Martínez VJ, Stoyan D, Saar E (2002). "Point field models for the galaxy point pattern – modelling the singularity of the two-point correlation function." *Astronomy and Astrophysics.* **388**, 758–765.

Stoyan D, Kendall WS, Mecke J (1977). *Stochastic Geometry and its Applications.* Wiley, Chichester.

Stoyan D (1992). "Statistical estimation of model parameters of planar Neyman–Scott cluster processes." *Metrika*, **39**, 67–74.

Tanaka U, Ogata Y, Stoyan D (2008). "Parameter estimation and model selection for Neyman–Scott point processes." *Biometrical Journal*, **50**, 43–57.

Teich MC (1981). "Role of the doubly stochastic Neyman Type A and Thomas counting distributions in photo detection." *Applied Optics*, **20**, 2457–2467.

Vere-Jones D (1970). "Stochastic models for earthquake occurrence." *Journal of the Royal Statistical Society, Series B (Methodological)* **32**, 1–62.

Waagepetersen RP (2007). "An Estimating Function Approach to Inference for Inhomogeneous Neyman–Scott Processes." *Biometrics*, **63**, 252–258.

Affiliation:

Jiancang Zhuang
Institute of Statistical Mathematics
Research Organization of Information and Systems
10-3 Midori-cho, Tachikawa, Tokyo 190-8562, Japan
E-mail: zhuangjc@ism.ac.jp
URL: http://bemlar.ism.ac.jp/zhuang