Full nonparametric estimate of space-time ETAS model

Giada Adelfio and Marcello Chiodi
{giada.adelfio, marcello.chiodi}@unipa.it
Dipartimento di Scienze Economiche Aziendali e Statistiche
Università di Palermo, Italy

Postdam (June, 15th)
1 Background and Aim: NP estimation of the CIF in branching space-time PP

2 Simultaneous estimation of parametric and nonparametric components

3 WORK in PROGRESS: MORE nonparametric

4 Remarks and future developments
Outline

1. Background and Aim: NP estimation of the CIF in branching space-time PP

2. Simultaneous estimation of parametric and nonparametric components

3. WORK in PROGRESS: MORE nonparametric

4. Remarks and future developments
Outline

1. Background and Aim: NP estimation of the CIF in branching space-time PP

2. Simultaneous estimation of parametric and nonparametric components

3. WORK in PROGRESS: MORE nonparametric

4. Remarks and future developments
Outline

1. Background and Aim: NP estimation of the CIF in branching space-time PP

2. Simultaneous estimation of parametric and nonparametric components

3. WORK in PROGRESS: MORE nonparametric

4. Remarks and future developments
Outline

1. Background and Aim: NP estimation of the CIF in branching space-time PP

2. Simultaneous estimation of parametric and nonparametric components

3. WORK in PROGRESS: MORE nonparametric

4. Remarks and future developments
In this paper, we propose a nonparametric estimation approach of the space-time intensity of a branching/Epidemic-type point process (ETAS).

Main aspects:

- simultaneous estimation of both components
- Use the forward predictive likelihood estimation approach for semi-parametric models (Chiodi and Adelfio, 2011).
Theoretical issue

To describe and interpret the features of realizations of space-time point processes (e.g. seismic data, fires data, diseases data) a reliable estimation of the conditional intensity function is necessary.

- In exploratory contexts or to assess the adequacy of a specific parametric model, some kind of nonparametric estimation (kernel) procedure could be useful.
- In some fields (e.g. seismological one) predictive properties of the estimated intensity function are pursued.
Theoretical issue

To describe and interpret the features of realizations of space-time point processes (e.g. seismic data, fires data, diseases data) a reliable estimation of the conditional intensity function is necessary.

- In exploratory contexts or to assess the adequacy of a specific parametric model, some kind of nonparametric estimation (kernel) procedure could be useful.
- In some fields (e.g. seismological one) predictive properties of the estimated intensity function are pursued.
Background and Aim: NP estimation of the CIF in branching space-time PP

Conditional Intensity Function in Point Processes

- Point process is a random collection of points, each one representing the time and space coordinates of a single event.
- Any analytic space-time point process is uniquely characterized by its associated conditional intensity function (CIF) (Daley and Vere-Jones, 2003):

$$\lambda(z) = \lambda(t, s|H_t) = \lim_{\Delta t, \Delta s \to 0} \frac{\mathbb{E}[N([t, t + \Delta t], [s, s + \Delta s]|H_t)]}{\Delta t\Delta s}$$

where $H_t$ is the space-time occurrence history of the process up to time $t$, $\Delta t, \Delta s$ are time and space increments, $\mathbb{E}[N([t, t + \Delta t], [s, s + \Delta s]|H_t)]$ is the history-dependent expected number of events occurring in the volume $\{[t, t + \Delta t) \times [s, s + \Delta s]\}$.
- Generally, intensities $\lambda(z)$ depend on some unknown parameter $\psi$. 
Background and Aim: NP estimation of the CIF in branching space-time PP

In a non-parametric context

\( \psi = \{ h \} \) is a vector of smoothing parameters.

- Let \( \hat{\lambda}_{\psi; H_{tk}}(\cdot) \) be a generic estimator of \( \lambda(\cdot) \) based on observations until \( t_k \).

Likelihood

Given the \( k \) observed values \( z_i \) the log-likelihood is:

\[
\log L(\hat{\lambda}_{\psi; H_{tk}}(z); H_{tk}) = \sum_{i=1}^{k} \log \hat{\lambda}_{\psi; H_{tk}}(z_i) - \int_{T_0}^{T_{\text{max}}} \int_{\Omega_s} \hat{\lambda}_{\psi; H_{tk}}(z) \, ds \, dt
\]  (1)
In a non-parametric context

\( \psi = \{h\} \) is a vector of smoothing parameters.

- Let \( \hat{\lambda}_{\psi;H_{tk}}(\cdot) \) be a generic estimator of \( \lambda(\cdot) \) based on observations until \( t_k \).

Likelihood

Given the \( k \) observed values \( z_i \) the log-likelihood is:

\[
\log L(\hat{\lambda}_{\psi;H_{tk}}(z); H_{tk}) = \sum_{i=1}^{k} \log \hat{\lambda}_{\psi;H_{tk}}(z_i) - \int_{T_0}^{T_{\max}} \int_{\Omega_s} \hat{\lambda}_{\psi;H_{tk}}(z) \, ds \, dt \quad (1)
\]
In a non-parametric context

\( \psi = \{ h \} \) is a vector of smoothing parameters.

- Let \( \hat{\lambda}_\psi; H_{t_k}(\cdot) \) be a generic estimator of \( \lambda(\cdot) \) based on observations until \( t_k \).

**Likelihood**

Given the \( k \) observed values \( z_i \) the log-likelihood is:

\[
\log L(\hat{\lambda}_\psi; H_{t_k}(z); H_{t_k}) = \sum_{i=1}^{k} \log \hat{\lambda}_\psi; H_{t_k}(z_i) - \int_{T_0}^{T_{\max}} \int_{\Omega_s} \hat{\lambda}_\psi; H_{t_k}(z) \, ds \, dt
\]  

(1)
A branching point process

- In probability theory, a branching process is a Markov process in which each individual in generation $n$ produces some random number of individuals in generation $n+1$, according to a probability distribution that does not vary from individual to individual.
- Branching processes are used to model reproduction phenomena.
- The CIF of the branching model is defined as the sum of:

$$\lambda_\theta(t, s|\mathcal{H}_t) = \mu f(s) + \tau_\phi(t, s)$$

(long-scale variation term + short-scale variation term)

with $\theta = (\phi, \mu)'$, the vector of parameters of the induced intensity ($\phi$) together with the parameter of the background general intensity ($\mu$) and

$$\tau_\phi(t, s) = \sum_{t_j < t} g(t - t_j; \phi)r(x - x_j, y - y_j|\phi).$$
A branching point process

- In probability theory, a branching process is a Markov process in which each individual in generation \( n \) produces some random number of individuals in generation \( n + 1 \), according to a probability distribution that does not vary from individual to individual.

- Branching processes are used to model reproduction phenomena.

- The CIF of the branching model is defined as the sum of:

\[
\lambda_\theta(t, s | \mathcal{H}_t) = \mu f(s) + \tau_\phi(t, s) \quad (2)
\]

where \( \theta = (\phi, \mu)' \), the vector of parameters of the induced intensity \( \phi \) together with the parameter of the background general intensity \( \mu \) and

\[
\tau_\phi(t, s) = \sum_{t_j < t} g(t - t_j; \phi) r(x - x_j, y - y_j | \phi).
\]
A branching point process

- In probability theory, a branching process is a Markov process in which each individual in generation $n$ produces some random number of individuals in generation $n+1$, according to a probability distribution that does not vary from individual to individual.
- Branching processes are used to model reproduction phenomena.
- The CIF of the branching model is defined as the sum of:

$$
\lambda_{\theta}(t, s|\mathcal{H}_t) = \mu f(s) + \tau_\phi(t, s)
$$

long-scale variation term + short-scale variation term

with $\theta = (\phi, \mu)'$, the vector of parameters of the induced intensity ($\phi$) together with the parameter of the background general intensity ($\mu$) and

$$
\tau_\phi(t, s) = \sum_{t_j < t} g(t - t_j; \phi) r(x - x_j, y - y_j|\phi).
$$
A branching point process

- In probability theory, a branching process is a Markov process in which each individual in generation $n$ produces some random number of individuals in generation $n+1$, according to a probability distribution that does not vary from individual to individual.
- Branching processes are used to model reproduction phenomena.
- The CIF of the branching model is defined as the sum of:

$$\lambda_\theta(t, s|\mathcal{H}_t) = \mu f(s) + \tau_\phi(t, s)$$

(long-scale variation term + short-scale variation term)

with $\theta = (\phi, \mu)'$, the vector of parameters of the induced intensity ($\phi$) together with the parameter of the background general intensity ($\mu$) and

$$\tau_\phi(t, s) = \sum_{t_j < t} g(t - t_j; \phi)r(x - x_j, y - y_j|\phi).$$
A branching process for earthquake description: the ETAS model

- The Epidemic Type Aftershocks-Sequences (ETAS) model is a self-exciting point processes, widely used in seismological context (Ogata, 1988).
- In the ETAS model, the background activity is assumed stationary in time.
- The occurrence rate of aftershocks at time $t$, following the event of time $t_j$ and magnitude $m_j$, is described by:

$$
g(t - t_j | m_j) = \frac{\kappa \ e^{(\alpha - \gamma) (m_j - m_0)}}{(t - t_j + c)^p}, \quad \text{with} \quad t > t_j
$$

where $p$ is useful for indicating the decay rate of the induced activity in time.

- For the spatial distribution (conditioned to magnitude of the generating event):

$$
r(x - x_j, y - y_j | m_j) = \left\{ \frac{(x - x_j)^2 + (y - y_j)^2}{e^{\gamma (m_j - m_0)}} + d \right\}^{-q}
$$

- $(\alpha, \gamma, c, p, q, d, \kappa)$ are unknown parameters which characterize the induced seismicity.
A branching process for earthquake description: the ETAS model

- The Epidemic Type Aftershocks-Sequences (ETAS) model is a self-exciting point processes, widely used in seismological context (Ogata, 1988).
- In the ETAS model, the background activity is assumed stationary in time.
- The occurrence rate of aftershocks at time $t$, following the event of time $t_j$ and magnitude $m_j$, is described by:

$$g(t - t_j | m_j) = \frac{\kappa e^{(\alpha - \gamma) (m_j - m_0)}}{(t - t_j + c)^p}, \quad \text{with} \quad t > t_j$$

where $p$ is useful for indicating the decay rate of the induced activity in time.

- For the spatial distribution (conditioned to magnitude of the generating event):

$$r(x - x_j, y - y_j | m_j) = \left\{ \frac{(x - x_j)^2 + (y - y_j)^2}{e^{\gamma (m_j - m_0)}} + d \right\}^{-q}$$

- $(\alpha, \gamma, c, p, q, d, \kappa)$ are unknown parameters which characterize the induced seismicity.
Outline

1 Background and Aim: NP estimation of the CIF in branching space-time PP

2 Simultaneous estimation of parametric and nonparametric components

3 WORK in PROGRESS: MORE nonparametric

4 Remarks and future developments
A crucial statistical issue

estimate *at the same time* the background intensity and the triggered intensity components of branching model.

- While the first component $f(x, y)$ is usually estimated by nonparametric techniques, $\theta$ is estimated by ML approach.
Simultaneous estimation of parametric and nonparametric components

Alternating estimation of components

A crucial statistical issue

estimate at the same time the background intensity and the triggered intensity components of branching model.

- While the first component $f(x, y)$ is usually estimated by nonparametric techniques, $\theta$ is estimated by ML approach.
Given $n$ observed events $z_1, z_2 \ldots, z_n$ in a $d$-dimensional closed region, the kernel estimator of the unknown density $f$ (Silverman, 1986) in $\mathbb{R}^d$ is:

$$\hat{f}_\Sigma(z) = \frac{\sum_{i=1}^{n} K(z - z_i, \Sigma) \cdot w_i}{\sum_i w_i} \quad (3)$$

where $K(\cdot, \cdot)$ is a multivariate kernel function centered at observed points and $\Sigma$ is a matrix of smoothing constants and $w_i$ weights.

- bivariate normal kernel for the spatial distribution of background seismicity, with two smoothing constants, $h = (h_x, h_y)$ estimated by the FLP approach.
FLP (Chiodi and Adelfio, 2011; Adelfio and Chiodi, 2015)

A nonparametric estimation procedure based on the subsequent increments of log-likelihood obtained adding an observation one at a time, to account for the information of the observations until $t_k$ on the next one.

- Let $\log L(\hat{\lambda}_{\psi;H_{tk}}(z); H_{tk+1})$ be the log-likelihood computed on the first $k + 1$ observations, but using the estimator $\hat{\lambda}_{\psi;H_{tk}}(z)$ on the previous $k$ observations.
- We measure the *predictive information* of the first $k$ observations on the $k + 1$-th as:

$$\delta_{k,k+1}(\psi) = \log L(\hat{\lambda}_{\psi;H_{tk}}(z); H_{tk+1}) - \log L(\hat{\lambda}_{\psi;H_{tk}}(z); H_{tk})$$
FLP (Chiodi and Adelfio, 2011; Adelfio and Chiodi, 2015)

A nonparametric estimation procedure based on the subsequent increments of log-likelihood obtained adding an observation one at a time, to account for the information of the observations until $t_k$ on the next one.

Let $\log L(\hat{\lambda}_\psi; H_{t_k}(z); H_{t_{k+1}})$ be the log-likelihood computed on the first $k + 1$ observations, but using the estimator $\hat{\lambda}_\psi; H_{t_k}(z)$ on the previous $k$ observations.

We measure the *predictive information* of the first $k$ observations on the $k + 1$-th as:

$$\delta_{k,k+1}(\psi) = \log L(\hat{\lambda}_\psi; H_{t_k}(z); H_{t_{k+1}}) - \log L(\hat{\lambda}_\psi; H_{t_k}(z); H_{t_k})$$
This leads to a technique similar to cross-validation, but applied only on next observations.

Therefore, we choose \( \tilde{\psi}(H_{tk}) \) which maximizes:

\[
FLP_{k_1,k_2}(\hat{\psi}) = \sum_{k=k_1}^{k_2} \delta_{k,k+1},
\]

in our applications we used \( k_1 = \left\lceil \frac{n}{2} \right\rceil \) and \( k_2 = n - 1 \).

FLP provides better kernel estimates (in terms of MISE) of space-time intensity functions than classical methods (Chiodi and Adelfio, 2011).
This leads to a technique similar to cross-validation, but applied only on next observations.

Therefore, we choose \( \tilde{\psi}(H_{tk}) \) which maximizes:

\[
FLP_{k_1,k_2}(\hat{\psi}) \equiv \sum_{k=k_1}^{k_2} \delta_{k,k+1},
\]

in our applications we used \( k_1 = \left\lceil \frac{n}{2} \right\rceil \) and \( k_2 = n - 1 \).

FLP provides better kernel estimates (in terms of MISE) of space-time intensity functions than classical methods (Chiodi and Adelfio, 2011).
Simultaneous estimation of nonparametric and parametric components of a branching-type model:

alternate the standard parametric likelihood method (to estimate the parameters of the offsprings component) with the FLP approach (to estimate the background intensity).

- Given the lack of specific open-source tools, the R package etasFLP (Chiodi and Adelfio, 2014) provides tools to implement this mixed approach for a wide class of ETAS models for the description of seismic events, developed in a R environment.
- The package has various options which allow to deal with different models of the ETAS family and also different estimation strategies.
- current version 1.3.0
Simultaneous estimation of nonparametric and parametric components of a branching-type model:

**alternate the standard parametric likelihood method (to estimate the parameters of the offsprings component) with the FLP approach (to estimate the background intensity).**

- Given the lack of specific open-source tools, the R package etasFLP (Chiodi and Adelfio, 2014) provides tools to implement this mixed approach for a wide class of ETAS models for the description of seismic events, developed in a R environment.
- The package has various options which allow to deal with different models of the ETAS family and also different estimation strategies.
- current version 1.3.0
Outline

1. Background and Aim: NP estimation of the CIF in branching space-time PP
2. Simultaneous estimation of parametric and nonparametric components
3. WORK in PROGRESS: MORE nonparametric
4. Remarks and future developments
More recently, nonparametric methods have been introduced for self-exciting point process estimation (Zhuang 2006; Marsan and Lengliné 2008).

In this paper we consider a Kernel estimator approach for both the components of the branching process, where the second one is computed for the inter-point distances, to measure the triggering effect.

For the selection of inter-point distances we follow the Monte Carlo-based iterative procedure, with some differences, of Moheler et al. (2011).
Consider space-time point data $\{(t_k, x_k, y_k)\}_{k=1}^{N}$ and a general self-exciting point process model as reported above (2)

Assuming model correctness, the Kernel estimation step, based on data $\{(t_k, x_k, y_k)\}_{k=1}^{N}$ and $\{(t_i - t_j, x_i - x_j, y_i - y_j)\}_{t_j < t_i}$, is computationally expensive, since the number of data points is $O(N^2)$ and are both three-dimensional.

Then the kernel computational effort is $O(N^4)$. 
Weights computation

The probability that event $j$ is a background event, $p_{jj}$, is given by

$$p_{jj} = \frac{\mu f(x_j, y_j)}{\lambda(t_j, x_j, y_j)}$$

and the probability that event $i$ is triggered by $j$, $p_{ij}$, is

$$p_{ij} = \frac{g(t_i - t_j)r(x_i - x_j, y_i - y_j)}{\lambda(t_i, x_i, y_i)}$$

The NP Algorithm

Given a set of \( n \) events occurred in a fixed space-time region, and set \( v = 1 \), let \( \hat{f}_{\Sigma(0)}(x, y) \) be a starting estimation of the background intensity, obtained by the ETAS/FLP. The \( v \) th iteration of the simultaneous estimation of the nonparametric, including a Monte-Carlo selection of distances, is:

- Sample \( N_0^{(v)} \) offspring/parent inter-point distances \( \{(t_i - t_j, x_i - x_j, y_i - y_j)\}_{t_j < t_i} \) with prob. \( p_{\Delta ij}^{(v)} = p_{ij}^{(v)} p_{jj}^{(v)} \) and \( N_0^{(v)} = \sum p_{\Delta ij}^{(v)} \).
- Estimate \( f(\cdot) \) background density by using the ETAS/FLP bandwidth with weighted Kernel Density Estimation and weights \( p_{jj}^{(v)} \).
- Estimate \( g(\cdot)r(\cdot) \) triggered intensity by Kernel Density Estimation, for sampled pairs according to weights \( p_{\Delta ij}^{(v)} \).
- Update \( P^{(v)} \).
- Update \( v \) and start a new iteration, until some convergence rule is reached.
As an application, consider the California catalog of 3137 events from 1968 to 2012 with a threshold magnitude of 3.8. in the space window $(-127.5, -112.1) \times (31.7, 45.6)$. 
Figure: Observed time density (left side) and space-magnitude-time-time distribution (right side) of the California seismicity with magnitude threshold 3.8.
Figure: Estimated whole model (left side) and background (right side) intensities of the California seismicity with magnitude threshold 3.8 by the FLP approach.
Figure: Rescaled times for California ETAS/FLP model (8 parameters), such that
\[ \tau_i = \Lambda(t_i) = \int_0^{t_i} \lambda(t) dt \]
Spatial residuals: differences between the theoretical and empirical frequencies in each of the $k \times k$ cells of the observed space area

Figure: Spatial residual analysis for the whole model (left side) and background (right side) intensities of the California seismicity with magnitude threshold 3.8 by ETAS/FLP.
Some comments

- From these plots, it seems that the background model, estimated by the ETAS/FLP approach, is quite satisfying mostly in space.
- It seems that the FLP improves the fitting of the background that in general appears very appropriate for data, both in time and space, while some lack are still evident for the offspring component, mostly in space domain.
- This suggested the possibility of make more flexible the induced component of the ETAS model.
NP approach: some estimation results

Figure: (log)-Productivity estimated by the NP approach and computed as \[ \sum_{i=1}^{n} \lambda_j(t_i, x_i, y_i \mid m_j) \] with \( t_j < t_i \), with \( a = -15.3665 \) and \( b = 0.1426 \)
Figure: Triggered intensity $\sum_{t_j < t} g(t - t_j) r(x - x_j, y - y_j)$ estimated by the NP approach versus $\Delta(s)$ (on the left) and $\Delta(t) < 250$ days (on the right), plus the average and standard errors.
Figure: Triggered intensity $\sum_{t_j < t} g(t - t_j) r(x - x_j, y - y_j)$ estimated by the NP approach versus time: vertical lines corresponds to big events ($\text{magn} \geq 7$ in red, $\text{magn} \geq 6.5$ in green)
Global Diagnostics: use of K-function for the thinned process, by the inverse of the CIF (Møller and Schoenberg (2010))

**Figure**: Estimated K-function (red line) together with the theoretical Poisson case (green line) and confidence bands (black lines) for the thinned process with the NP-model (on the left) and the ETAS/FLP (on the right)
Triggered Diagnostics: compute the weighted spectrum (Second-Order Residuals, Adelfio and Schoenberg, 2011), weighting the observed points by the inverse of the triggered CIF.

Figure: Estimated spectrum and weighted spectrum versions (by the inverse of the triggered intensity function) for the NP-model (on the left) and for the ETAS/FLP (on the right).
Remarks and future developments

Outline

1. Background and Aim: NP estimation of the CIF in branching space-time PP
2. Simultaneous estimation of parametric and nonparametric components
3. WORK in PROGRESS: MORE nonparametric
4. Remarks and future developments

G. Adelfio and M. Chiodi
Nonparametric estimate of space-time ETAS model
STATSEI9 - 2015
• The proposed nonparametric estimation of the components of a branching process is very flexible.
• It accounts for predictive properties of the estimated intensity including also FLP.
• Let data to speak for themselves: could be a starting point for further fitting of any parametric model (like the ETAS).
Remarks and future developments

Future

- Consider asymmetric Kernels for time intensity (Marcon et al 2011).
- OR... Categorize pairs data by histograms instead of kernels.
- Anisotropic kernel with variable smoothing parameters to take into account more realistic situations (please, see the Poster Session).
A nonparametric model with variable bandwidth values is in progress, to study variations of observed intensity in space and time.
References


References