

Time-frequency analysis of locally stationary Hawkes processes

CIRM, Luminy

Rainer von Sachs

Institut de statistique, biostatistique et sciences actuarielles, UCL

joint work with

F. Roueff (*Télécom ParisTech*) and L. Sansonnet (*Agro ParisTech*).

February 18, 2016

Contents

- 1 Introduction
- 2 Approximation results
- 3 Time-frequency analysis of Hawkes processes
- 4 Numerical experiments
- 5 Estimation theory

Introduction

- Diverse application fields of Hawkes processes (Hawkes, 1971) = self-exciting point processes:
 - **seismology** (Ogata, 1988), **genomics** (Reynaud-Bouret and Schbath, 2010), **neuroscience** (Reynaud-Bouret *et al.*, 2013), **finance**: microstructure dynamics (Bacry *et al.*, 2012), order arrival rate modelling and high-frequency data (Bowsher, 2007), , ...
- Stationary linear Hawkes process N :
 - ▶ its **conditional intensity**:

$$\lambda(t) = \lambda_c + \int_{-\infty}^{t^-} p(t-s) N(ds) = \lambda_c + \sum_{T_i < t} p(t - T_i)$$

- ▶ alternatively: description by **clusters of point processes**: immigrants (follow a Poisson process and define the centers of clusters) and offsprings.

Aim of this work: Model and capture *time-varying* dynamics, i.e. define a model of point process that can be locally interpreted as a stationary Hawkes process (such as for "classical" time series by R. Dahlhaus).

Introduction - cont'd

Existing approaches of time-varying dynamics for Hawkes processes

Essentially only $\lambda_c = \lambda_c(t)$ (e.g. Chen and Hall, 2013) and no "locally stationary" approach via rescaling in time (taking $u = t/T$)

Some notation for this talk

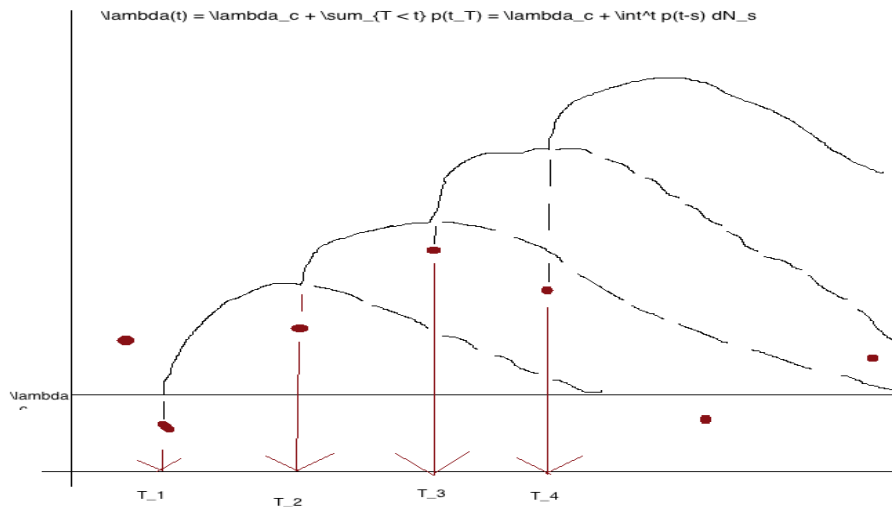
- We look at (time-continuous) processes with values in \mathbb{R}^ℓ (but for today mostly $\ell = 1$)
- A point process is identified with a **random measure** with discrete support: $N = \sum_k \delta_{T_k}$, where δ_t is the Dirac measure at point t and $\{T_k\}$ is the corresponding (countable) random set of points.
- For a test function g , $N(g) = \sum_k g(T_k)$.
- (Some) Functional norms:
 - ▶ L^q -norms $|\cdot|_q$ for $q \in [1; \infty]$, and
 - ▶ β -norm: $|h|_{(\beta)} := |h \times |\cdot|^\beta|_1 = \int |h(s)| |s|^\beta ds$, for a given $\beta > 0$.

Hawkes processes as cluster processes

- **Cluster processes** = point processes constructed via conditioning on the realization of a so-called *center process*, usually a PPP in the sequel (Daley and Vere-Jones, 2003).
- Starting point: N_c a PPP with intensity measure μ_c and density λ_c that represents the *immigrants* which appear spontaneously.
- At each *center point* t of N_c , a (independent branching) PPP $N(\cdot|t)$ with finite intensity $\mu(\cdot|t)$, generating the *offsprings* at t .
- Cluster process N = all the immigrants + all the offspring of each immigrant.
- Standard Hawkes processes are made stationary by assuming that N_c is a homogeneous PPP on the whole space \mathbb{R}^ℓ and $t \mapsto \mu(\cdot|t)$ is shift invariant: $\mu(\cdot|t) = \mu \circ S^t$, where μ is fixed (i.e. $\mu = \mu(\cdot|0)$).

Dynamics of Hawkes processes

$$\lambda(t) = \lambda_c + \int_{-\infty}^{t^-} p(t-s) N(ds) = \lambda_c + \sum_{T_i < t} p(t - T_i)$$



Our work on non- and locally stationary Hawkes processes

Roadmap of our approach on *approximation* (SPA 2015):

- First, we show existence of *non-stationary* Hawkes processes
- Then, we define *locally stationary* Hawkes processes using a series of regularity conditions on immigrant intensity $\lambda_c(t/T)$ and fertility function $p(\cdot; t/T)$
- The key tool for existence of an approximating stationary process: local approximation of the log Laplace functional
- This allows for approximation of the first and second order moment structure of the non-stationary by the (locally) stationary Hawkes process
- In particular: existence of a local (i.e. time-varying) Bartlett spectrum (time-frequency analysis)

Existence of *non-stationary* Hawkes process

Give up shift invariance of $t \mapsto \mu_c(t)$ and $t \mapsto \mu(\cdot|t)$, i.e. their densities $\lambda_c(t)$ (immigrant intensity) and $p(\cdot; t)$ (fertility function) depend on t .

We can define a *non stationary* Hawkes process via these t-v densities if

$$\zeta_{\mathbf{1}} := \sup_{t \in \mathbb{R}} \mu(\mathbb{R}^\ell | t) = \sup_{t \in \mathbb{R}^\ell} \int p(s; t) ds < 1 \quad \text{and} \quad |\lambda_c|_\infty < \infty .$$

Theorem 1: The component process $N(\cdot|t)$ has finite moment measure $\mathbb{E}[N(\mathbb{R}^\ell | t)] \leq \frac{1}{1-\zeta_{\mathbf{1}}}$ and its density is uniformly bounded by $\frac{|\lambda_c|_\infty}{1-\zeta_{\mathbf{1}}}$.

Proof via the cluster construction: each component $N(\cdot|t)$ can be constructed as the superposition of point processes defined iteratively.

Non-stationary Hawkes processes (under the density assumption) with λ_c and $p(\cdot; \cdot)$ still quite general: can still **evolve quite arbitrarily** in the space.

Locally stationary Hawkes processes

- **Aim:** Define a locally stationary approximation - similarly to autoregressive processes. In the following, to simplify notation, consider univariate case only, i.e. $\ell = 1$ w.l.o.g.

Definition of a locally stationary Hawkes process

A locally stationary Hawkes process with *local immigrant intensity* $\lambda_c^{<LS>}$ and *local fertility function* $p^{<LS>}(\cdot; \cdot)$ is

- a collection $(N_T)_{T>0}$ of non-stationary Hawkes processes with
- immigrant intensity $\lambda_{cT}(t) = \lambda_c^{<LS>}(t/T)$ and
- fertility function $p_T(\cdot; t) = p^{<LS>}(\cdot; t/T)$.

For a given *real location* t , the scaled location t/T is typically called an *absolute location* in $[0; 1]$ and denoted by u .

Analogy with locally stationary AR-processes

Note the similarities/differences with an AR(p)-process ($p = 1$ w.l.o.g.):

Dahlhaus (SPA 1996) shows that ($t = 1, \dots, T$, $T \geq 1$)

$$X_{t,T} = a(t/T) X_{t-1,T} + \varepsilon_t, \quad \varepsilon_t \sim (0, 1),$$

has a representation

$$X_{t,T} = \int_{-\pi}^{\pi} A_{t,T}^{\circ}(\omega) \exp(i\omega t) d\xi(\omega)$$

with the existence of an $A(u, \omega) = (1 - a(u) \exp(-i\omega))^{-1}$

such that $\sup_{t,\omega} |A_{t,T}^{\circ}(\omega) - A(t/T, \omega)| = O(T^{-1})$.

Here, $A_{t,T}^{\circ}(\omega) = \sum_{\ell \geq 0} (-1)^{\ell} \prod_{j=0}^{\ell-1} a(\frac{t-j}{T}) \exp(-i\omega \ell)$.

Unfortunately, this is not possible for the more elaborated recursive dynamics of a Hawkes process.

Under assumption (LS-1):

$$\zeta_1^{<LS>} := \sup_{u \in [0,1]} \int p^{<LS>}(t; u) dt < 1 \quad \text{and} \quad |\lambda_c^{<LS>}|_\infty < \infty ,$$

for all $T > 0$, the non-stationary Hawkes process N_T admits a uniformly bounded intensity function.

- Under (LS-1), moreover, for each $u \in \mathbb{R}^\ell$, the function $t \mapsto p^{<LS>}(t; u)$ satisfies the required condition for the fertility function of a *stationary* Hawkes process.
- Notation: $N(\cdot; u)$ = a "stationary" Hawkes process with immigrant intensity $\lambda_c^{<LS>}(u)$ and fertility function $t \mapsto p^{<LS>}(t; u)$.
- More regularity assumptions:
 - ▶ (LS-2): smoothness conditions on $\lambda_c^{<LS>}(u)$
 - ▶ (LS-3): smoothness conditions on $p^{<LS>}(t, u)$ w.r.t. its second argument u
 - ▶ (LS-4): some uniform decay condition on $p^{<LS>}(\cdot, \cdot)$ w.r.t. its first argument

Local approximation of the log Laplace functional

The process $N_T \circ S^{-Tu}$, i.e. N_T shifted at the *real location* Tu , follows approximately the distribution of a stationary Hawkes process $N(\cdot; u)$.

Proof: via Convergence of the log-Laplace function of N_T (i.e. log of $\mathcal{L}_T(g) = \mathbb{E}[\exp N_T(g)]$):

Theorem 2

Under Ass. (LS-1:4), for $\beta \in (0; 1]$ and an appropriate test-function g

- the log-Laplace transform $\log \mathcal{L}_T(S^{-Tu}g)$ of $N_T \circ S^{-Tu}$, converges to the log-Laplace $\log \mathcal{L}(g; u)$ of $N(\cdot; u)$,
- with rate $T^{-\beta}$ and with **explicit bounds** for the constants depending only on the ("functional") β - and ℓ_1 -norms of the test function g .

Corollary 3

As application, for any $u \in [0, 1]$, as $T \rightarrow \infty$, the point process $N_T \circ S^{-Tu}$ converges in distribution to $N(\cdot; u)$.

Local approximation of the cumulants

Apply Theorem 2 to $\mathcal{L}_T(S^{-Tu}g) : z \mapsto \mathbb{E}[\exp N_T(g(\cdot - Tu, z))]$.

Key result for treatment of "all" **moments** (used in estimation theory) via

$$\text{Cum}(N(g_1), \dots, N(g_m)) = \partial^{1_m} \Big|_{z=0_m} \log \mathcal{L}(z_1 g_1 + \dots + z_m g_m),$$

Theorem 4

Let $\beta \in (0, 1]$. Assume (LS-1:4).

Let for any $m \geq 1$, g_1, \dots, g_m be real valued bounded integrable.

Then for any T and any $u \in [0, 1]$, we have

$$\begin{aligned} & \left| \text{Cum}(N_T(S^{-Tu}g_1), \dots, N_T(S^{-Tu}g_m)) - \text{Cum}(N(g_1; u), \dots, N(g_m; u)) \right| \\ & \leq \frac{2^{m-1} C_1 T^{-\beta}}{(-\log \zeta_1^{<LS>})^{m-1}} \left\{ \sum_{j=1, \dots, m} (|g_j|_{(\beta)} + C_2 |g_j|_1) \right\} \left\{ \sum_{j=1, \dots, m} (|g_j|_\infty + C_3 |g_j|_1) \right\}^{m-1}. \end{aligned}$$

First Application: Convergence of local mean density

Theorem 4 with $m = 1$ implies:

- For any T , N_T admits a uniformly bounded density function $m_{1T}(t)$.
- The intensity measure of the (approximating) stationary Hawkes process admits a time-constant mean density:

$$m_1^{<LS>}(u) = \frac{\lambda_c^{<LS>}(u)}{1 - \int p^{<LS>}(\cdot; u)}$$

→ *local mean density* at absolute location u .

- Convergence of $m_{1T}(t)$ to $m_1^{<LS>}(\cdot; u)$ with rate $T^{-\beta}$ in a local neighbourhood of Tu
(and explicit bounds on the constants depending only on the β - and ℓ_1 -norms of the test function g).

Second application: Local Bartlett spectrum

- Locally stationary time series provide a nonparametric statistical framework for time-frequency analysis of time series (Dahlhaus, 2009).
- Such ideas can now be applied to locally stationary Hawkes processes.
- Stationary case
 - ▶ The Bartlett spectrum Γ of a second order point process N on \mathbb{R} is defined as the (unique) nonnegative measure on \mathbb{R} s.t., for any bounded and compactly supported function f on \mathbb{R} ,

$$\text{Var}(N(f)) = \Gamma(|\hat{f}|^2) = \int \left| \hat{f}(\omega) \right|^2 \gamma(\omega) \, d\omega ,$$

where $\hat{f}(\omega) = \int f(t) e^{-it\omega} \, dt$.

- ▶ For Hawkes processes, the density of the Bartlett spectrum is given by

$$\gamma(\omega) = \frac{\lambda_c}{2\pi(1 - \int \rho)} |1 - \hat{\rho}(\omega)|^{-2} .$$

Under (LS-1), applying this result to the stationary Hawkes process $N(\cdot; u)$, we have, for any bounded and compactly supported function f ,

$$\text{Var}(N(f; u)) = \Gamma^{\langle \text{LS} \rangle}(|\hat{f}|^2; u) = \int |\hat{f}(\omega)|^2 \gamma^{\langle \text{LS} \rangle}(\omega; u) d\omega,$$

where

- $\gamma^{\langle \text{LS} \rangle}(\omega; u) = \frac{m_1^{\langle \text{LS} \rangle}(u)}{2\pi} |1 - \hat{p}^{\langle \text{LS} \rangle}(\omega; u)|^{-2}$, with
- $\hat{p}^{\langle \text{LS} \rangle}(\omega; u) = \int p^{\langle \text{LS} \rangle}(t; u) e^{-it\omega} dt$.
- $\Gamma^{\langle \text{LS} \rangle}(\cdot; u)$ and $\gamma^{\langle \text{LS} \rangle}(\omega; u)$ are called the *local Bartlett spectrum* and the *local spectral density*, respectively, at absolute location u .
- **Corollary 6:** Under (LS-1:4), for $\beta \in (0, 1]$ and bounded functions f supported inside $[-b, b]$, for some $b > 0$:

Convergence of $\text{Var}(N_T(S^{-T}u f))$ to $\Gamma^{\langle \text{LS} \rangle}(|\hat{f}|^2; u)$ with rate of order $T^{-\beta}$ (and again, explicit bounds on constants ..)

Kernel estimation of the local Bartlett spectrum

- f a test function and m a moment function (such as $m(x) = x$, $m(x) = x^2, \dots$).
- b_1 a given time bandwidth and u_0 a fixed time in $[0; 1]$ (namely, $u_0 = t_0/T$ with $t_0 \in [0; T]$).
- We build an estimator of $\mathbb{E}[m(N(f; u_0))]$ based on the empirical observations of N_T and defined by

$$\widehat{E}[m \circ N_T(S^{-Tu_0} f); W_{b_1}] := \frac{1}{T} \int m \circ N_T(f(\cdot - t - Tu_0)) W_{b_1}(t/T) dt,$$

for some fixed kernel function W .

- In practice, f should be compactly supported, so that this integral can be computed from a finite set of observations in $[0, T]$.

- Localisation (smoothing) in frequency by a real valued kernel K , compactly supported, with Fourier transform \hat{K} such that $\int |\hat{K}(\omega)|^2 d\omega = 1$.
- b_2 a given frequency bandwidth and ω_0 a fixed frequency.
- We wish to estimate the quantity

$$\gamma_{b_2}(\omega_0; u_0) := \int \frac{1}{b_2} |\hat{K}((\omega - \omega_0)/b_2)|^2 \Gamma^{\langle \text{LS} \rangle}(d\omega; u_0),$$

which in turn is an approximation of the local spectral density $\gamma^{\langle \text{LS} \rangle}(\omega_0; u_0)$ of $\Gamma^{\langle \text{LS} \rangle}(\cdot; u_0)$ at ω_0 .

- $f = K_{b_2, \omega_0}$ is the kernel having Fourier transform $\omega \mapsto b_2^{-1/2} \hat{K}((\omega - \omega_0)/b_2)$.
- Consequently, we get that $K_{b_2, \omega_0}(t) = b_2^{1/2} e^{i\omega_0 t} K(b_2 t)$.
- Finally, take $m(x) = x^2$ and $m(x) = x$ to define: $\hat{\gamma}_{b_2, b_1}(\omega_0; u_0) =$

$$= \hat{E} (|N_T(S^{-T u_0} K_{b_2, \omega_0})|^2; W_{b_1}) - \left| \hat{E} (N_T(S^{-T u_0} K_{b_2, \omega_0}); W_{b_1}) \right|^2.$$

Numerical experiments

Specific class of **Gamma-shaped local fertility functions** $p^{<LS>}(\cdot; u)$ and **time-constant immigrant intensity** $\lambda_c^{<LS>}(u) = 1/2$.

- Example 1 [Exponential without delay]:

$$p^{<LS>}(s; u) = \zeta(u) \frac{\theta(u)e^{-\theta(u)s}}{\Gamma(1)} \mathbb{1}_{s>0}, \quad \zeta(u), \theta(u) \text{ of cosine form.}$$

- Example 2 [Gamma with varying delay $\delta(u)$ and constant $\zeta(u)$]:

$$p^{<LS>}(s; u) = \frac{1}{2}(s - \delta(u)) \frac{e^{-(s-\delta(u))}}{\Gamma(2)} \mathbb{1}_{s>\delta(u)}$$

with $\delta(u) = (6 - 10u) \times \mathbb{1}_{[0;1/2]}(u) + (10u - 4) \times \mathbb{1}_{(1/2;1]}(u)$

inducing a **periodic phenomenon in the self-excitation**.

Here, $\zeta(u), \theta(u)$ and $\delta(u)$ are Lipschitz $\beta = 1$, and one gets explicit formulas for the local mean density

$$m_1^{<LS>}(u) = m_1^{<LS>}(\zeta(u)) = \frac{\lambda_c}{1 - \zeta(u)}.$$

and local Bartlett spectrum $\Gamma^{<LS>}(d\omega; u) = \Gamma^{<LS>}(d\omega; \delta(u), \zeta(u), \theta(u))$.

Simulation of locally stationary Hawkes processes

Use time-varying conditional intensity

$$\lambda_{\mathcal{T}}(t) := \lambda_c^{\langle \text{LS} \rangle}(t/T) + \sum_{T_i < t} p^{\langle \text{LS} \rangle}(t - T_i; t/T),$$

where $(T_i)_{i \in \mathbb{Z}}$ denote the points of $N_{\mathcal{T}}$.

Use **Ogata's modified thinning algorithm** (Ogata, 1981) to simulate the non-stationary Hawkes process $N_{\mathcal{T}}$ on the interval $[0, T]$.

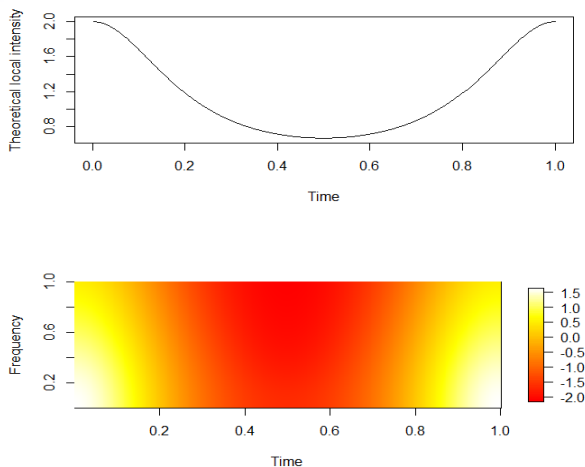


Figure 1 : *Theoretical local mean density (top) and Bartlett spectrum (bottom) for Example 1.*

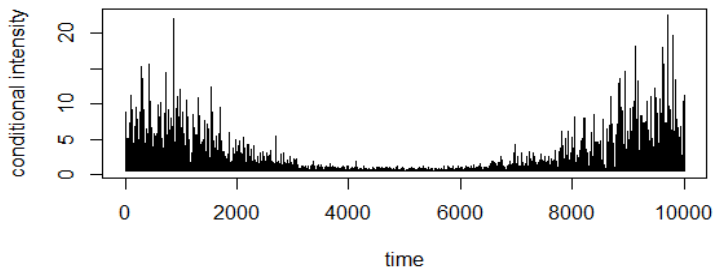


Figure 2 : *Conditional intensity function of a simulated Hawkes process following Example 1, with $T = 10000$.*

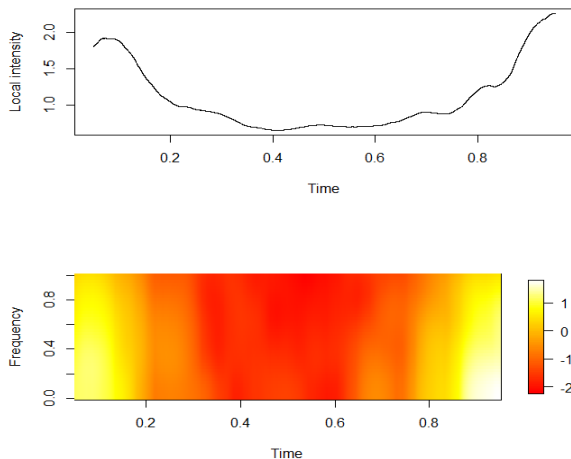


Figure 3 : *Estimation of the local mean density (top) and of the local Bartlett spectrum (bottom) for Example 1.*

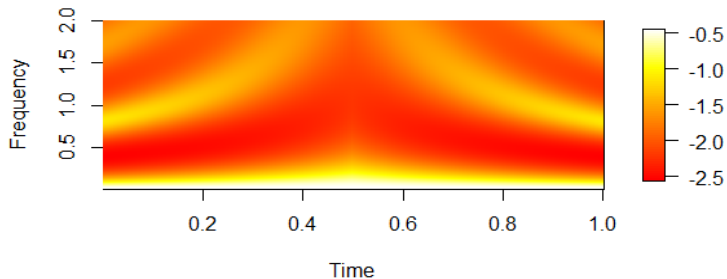


Figure 4 : *Theoretical local Bartlett spectrum for Example 2 (local mean density being constant over time).*

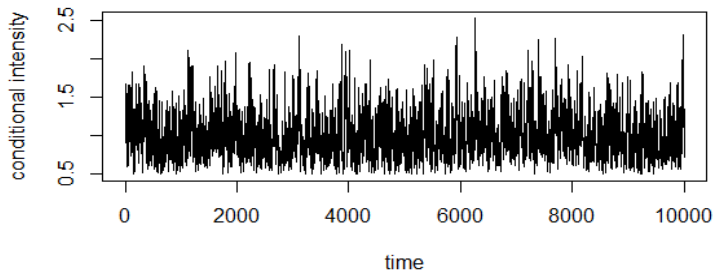


Figure 5 : *Conditional intensity function of a simulated Hawkes process following Example 2, with $T = 10000$.*

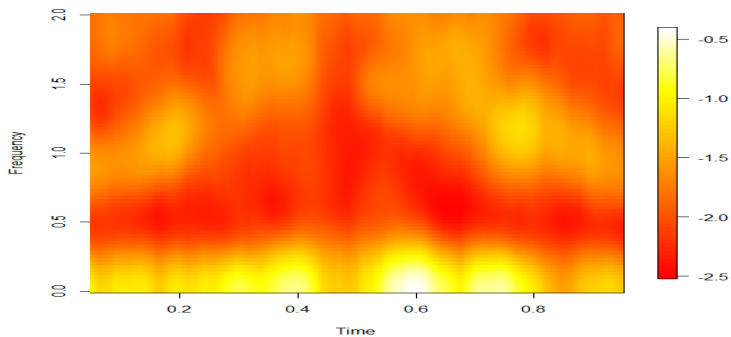


Figure 6 : *Estimation of the local Bartlett spectrum for Example 2.*

Some asymptotic estimation theory (work in progress)

Estimation of the local mean intensity

A straightforward estimator of

$$m_1^{<LS>}(u_0) = \frac{\lambda_c^{<LS>}(u_0)}{1 - \int p^{<LS>}(\cdot; u_0)}.$$

is

$$\widehat{m}_1(u_0) := N_T(S^{-Tu_0} W_{b_1 T}) = \int W_{b_1 T}(s - Tu_0) N_T(ds),$$

with time-kernel $W_{b_1 T}(\cdot) = \frac{1}{b_1 T} W\left(\frac{\cdot}{b_1 T}\right)$.

Under suitable conditions (of Theorem 4, integrability of W), with $\beta \in (0, 1]$, we have

- **Bias:** $O\left(b_1^\beta + T^{-\beta}\right)$
- **Variance:** $O\left((Tb_1)^{-1} + b_1^\beta + T^{-\beta}\right)$

This gives an MSE-rate of convergence of order $T^{-\frac{\beta}{\beta+1}}$.

Principle of proof, using Theorem 4 with $g = W_{b_1 T}$

- **Bias** (Thm 4 with $m = 1$):

$$\begin{aligned} \mathbb{E} \widehat{m}_1(u_0) - m_1(u_0) &\lesssim T^{-\beta} \left(|W_{b_1 T}|_1 + |W_{b_1 T}|_{(\beta)} \right) \\ &\lesssim T^{-\beta} (1 + (b_1 T)^\beta) \lesssim b_1^\beta + T^{-\beta} . \end{aligned}$$

- **Variance** (Thm 4 with $m = 2$ and $g_1 = g_2 = g$):

$$\begin{aligned} \text{Var}(\widehat{m}_1(u_0)) &\lesssim \text{Var}((N(W_{b_1 T, u_0}))) + \\ &+ T^{-\beta} (|W_{b_1 T}|_1 + |W_{b_1 T}|_{(\beta)}) \times (|W_{b_1 T}|_\infty + |W_{b_1 T}|_1) \lesssim \\ &\lesssim (T b_1)^{-1} + T^{-\beta} (1 + (b_1 T)^\beta) \times ((b_1 T)^{-1} + 1) \lesssim (T b_1)^{-1} + b_1^\beta + T^{-\beta} . \end{aligned}$$

Asymptotic estimation theory for local Bartlett spectrum

Employed technique:

- Control of the expectation and the variance of

$$\widehat{E}[N_T(f); w] := \frac{1}{T} \int N_T(f(\cdot - t))w(t/T) dt$$

estimator of first moment of "stationary" Hawkes process $N(f; u_0)$.

- Control of the expectation and the variance of

$$\widehat{E}[N_T^2(f); w] := \frac{1}{T} \int N_T^2(f(\cdot - t))w(t/T) dt$$

estimator of the second moment of $N(f; u_0)$.

Here again with the time-kernel $w = W_{b_1, u_0}$ and now with a "test-function" $f = K_{b_2, \omega_0}$, i.e. the kernel in frequency direction.

Again use Theorem 4, similarly to the proof of consistency of the local mean density estimator (but MUCH more involved).

Possibility of a CLT (future work)....

Bias of the estimator $\hat{\gamma}_{b_2, b_1}(\omega_0; u_0)$

Recall: Estimation of $\gamma_{b_2}(\omega_0; u_0)$ via

$$\hat{\gamma}_{b_2, b_1}(\omega_0; u_0) = \hat{E}(|N_T(S^{-Tu_0}K_{b_2, \omega_0})|^2; W_{b_1}) - \left| \hat{E}(N_T(S^{-Tu_0}K_{b_2, \omega_0}); W_{b_1}) \right|^2.$$

Under suitable conditions (of Theorem 4, integrability of K and W), with $\beta \in (0, 1]$, and for a Bartlett spectrum sufficiently regular in frequency, we have for the *Bias* a rate of order

$$\frac{1}{Tb_1b_2} + b_2^{-1} \left((Tb_2)^{-\beta} + b_1^\beta \right) + b_2^2$$

(The behaviour of the *Variance* has still to be worked out.)

Convergence of the bias under conditions for the two bandwidths:

$$Tb_1b_2 \rightarrow \infty, Tb_2^{1+\frac{1}{\beta}} \rightarrow \infty \text{ and } \frac{b_1^\beta}{b_2} \rightarrow 0,$$

plus a traditional bias-condition $b_2 \rightarrow 0$ for the convergence of $\gamma_{b_2}(\omega_0; u_0)$ to the "point" spectral density $\gamma^{<LS>}(\cdot; u_0)$ in ω_0 (for a given u_0).

Conclusion

- **Self-exciting point processes** ("Hawkes" processes) show a far more evolved dynamics over time than, e.g., autoregressive processes. Consequently, **locally stationary approximations** are more difficult to derive.
- Unfortunately, unlike "classical" time series (under usual conditions) **no spectral representation** exists that would render the approach more "direct".
- The key to success: approximation theory for **local Laplace transforms** and its derivatives (control of cumulants):
- Existence of (local) mean density (no "explosive" behaviour) and **local Bartlett spectrum** via control of convergence of first and second moment structure
- Work in progress: **Asymptotic estimation theory** (control of convergence of first and second empirical moments); real data analysis