TD 3 - Temporal point processes

Statistics and Hawkes process

1 Basic exercises

Exercise 1 (MLE - Point process vs standard framework)**.** Let $\{f_\theta, \theta \in \mathbb{R}^d\}$ be a parametric family of probability density functions on $\mathbb{R}_+$ and denote $\lambda_{\theta;C}(t) = C f_\theta(t)$ for all $t \geq 0$. Assume that the observation window is $[0,+\infty)$ and denote $0 < t_1 < \cdots < t_n$ the observed times. In the following, this sample can be considered in two ways:

1. as a sample of a Poisson process with intensity $\lambda_{\theta;C}$ (in particular, the size of the sample, namely n, is random),
2. as a n-sample (in the classical sense) of the probability density function f_θ.

Let us remark that there are two parameters to estimate in the first case (θ and C) whereas there is only one in the second case (θ).

Show that the two cases are equivalent in terms of maximum likelihood estimation of θ.

Exercise 2 (Goodness-of-fit tests)**.** Use the simulation algorithm of Exercise 4 of TD1 to simulate several trajectories of a homogeneous Poisson process with rate $\lambda = 1$ on $[0,T]$. Perform one of the 3 tests presented during the course (Uniform test, Exponential test, Donsker test) on these trajectories.

1. Compute the empirical type I error of these procedures while T is varying from 10 to 1000.
2. Compute the empirical type II error of these procedures with $\lambda \neq 1$. How does it vary with respect to T ?

Use the simulation algorithm of Exercise 6 of TD1 to simulate several trajectories of the non homogeneous Poisson process described there with $T = \alpha = \beta = 1$. After using the time change Theorem, perform one of the 3 tests presented during the course (Uniform test, Exponential test, Donsker test) on these trajectories.

1. Compute the empirical type I error of these procedures while C is varying from 20 to 1000.
2. Compute the empirical type II error of these procedures when the values of α and β are mis-specified. How does it vary with respect to C ?

2 Intermediate exercises

Exercise 3 (MLE - Hawkes process)**.** Consider the Hawkes process model given by intensity:

$$\lambda_t = \mu + \int_0^t \alpha e^{-\beta(t-t')} N(dt'),$$

where $\theta = (\mu, \alpha, \beta)$ are parameters with positive values. Let $T > 0$ be a finite time horizon for the observations and denote $0 < t_1 < \cdots < t_n \leq T$ the observed times. Let $A(1) = 0$ and for all $i \geq 2$,

$$A(i) = \sum_{j=1}^{i-1} e^{-\beta(t_i - t_j)}$$

1. Show that the log-likelihood is given by

$$\ell(\theta) = -\mu T + \frac{\alpha}{\beta} \sum_{i=1}^{n} (e^{-\beta(T-t_i)} - 1) + \sum_{i=1}^{n} \log\{\mu + \alpha A(i)\}.$$

Let $B(1) = 0$, $C(1) = 0$ and for all $i \geq 2$,

$$B(i) = \sum_{j=1}^{i-1} (t_i - t_j) e^{-\beta(t_i - t_j)} \quad \text{and} \quad C(i) = \sum_{j=1}^{i-1} (t_i - t_j)^2 e^{-\beta(t_i - t_j)}$$

2. Compute the gradient of the log-likelihood and check that the Hessian is given by[1]:

$$\frac{\partial^2 \ell}{\partial \alpha^2} = -\sum_{i=1}^{n} \left[\frac{A(i)}{\mu + \alpha A(i)}\right]^2,$$

$$\begin{aligned} \frac{\partial^2 \ell}{\partial \alpha \partial \beta} &= -\sum_{i=1}^{n} \beta^{-1}(t_n - t_i) e^{-\beta(t_n - t_i)} + \beta^{-2}(e^{-\beta(t_n - t_i)} - 1) \\ &\quad + \sum_{i=1}^{n} \frac{-B(i)}{\mu + \alpha A(i)} + \frac{\alpha A(i) B(i)}{(\mu + \alpha A(i))^2}, \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 \ell}{\partial \beta^2} &= \alpha \sum_{i=1}^{n} \beta^{-1}(t_n - t_i)^2 e^{-\beta(t_n - t_i)} + 2\beta^{-2}(t_n - t_i) e^{-\beta(t_n - t_i)} + 2\beta^{-3}(e^{-\beta(t_n - t_i)} - 1) \\ &\quad + \sum_{i=1}^{n} \frac{\alpha C(i)}{\mu + \alpha A(i)} - \left(\frac{\alpha B(i)}{\mu + \alpha A(i)}\right)^2, \end{aligned}$$

$$\frac{\partial^2 \ell}{\partial \mu^2} = -\sum_{i=1}^{n} \frac{1}{(\mu + \alpha A(i))^2}, \quad \frac{\partial^2 \ell}{\partial \mu \partial \alpha} = -\sum_{i=1}^{n} \frac{A(i)}{(\mu + \alpha A(i))^2}, \quad \frac{\partial^2 \ell}{\partial \mu \partial \beta} = \sum_{i=1}^{n} \frac{\alpha B(i)}{(\mu + \alpha A(i))^2}.$$

3. Use Newton-Raphson method to numerically approximate the maximum likelihood estimator of θ.

Exercise 4 (Population dynamics)**.** Let X_t, $t \geq 0$ denote the number of individuals in some population (for instance a natural baboon troop). We consider here a 4 dimensional point process:

- N^{b} is the set of birth type events,

[1] Can be found in (Ozaki, 1979).

- N^{d} is the set of death type events,
- N^{e} is the set of emigration type events,
- N^{i} is the set of immigration type events.

We assume that the rate of birth, death or emigration type events are proportional to the number of individuals whereas the rate of immigration type events is constant. Hence, we come up with the following intensities:

$$\lambda_t^{\mathrm{b}} = \theta_{\mathrm{b}} X_{t-}, \quad \lambda_t^{\mathrm{d}} = \theta_{\mathrm{d}} X_{t-}, \quad \lambda_t^{\mathrm{e}} = \theta_{\mathrm{e}} X_{t-}, \quad \lambda_t^{\mathrm{i}} = \theta_{\mathrm{i}}.$$

To be complete, the number of individuals X_t is related to the number of birth, death, emigration or immigration events. For instance a birth event increases by 1 the number of individuals. More precisely, we have for all $t \geq 0$,

$$X_t = X_0 + N_t^{\mathrm{b}} - N_t^{\mathrm{d}} - N_t^{\mathrm{e}} + N_t^{\mathrm{i}}.$$

Suppose that these four point processes (as well as X) are observed during the interval $[0,T]$. What is the maximum likelihood estimator of $\Theta := (\theta_{\mathrm{b}}, \theta_{\mathrm{d}}, \theta_{\mathrm{e}}, \theta_{\mathrm{i}})$?