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 $\lambda_{\theta;C}(t) = Cf_\theta(t)$ for all $t \geq 0$. Assume that the observation window is $[0,+\infty)$ and denote $0 < t_1 < \cdots < t_n$ the observed times.

1. Suppose that $0 < t_1 < \cdots < t_n$ is a sample of a Poisson process N with intensity $\lambda_{\theta;C}(t)$. By definition, its log-likelihood is
$$\ell(\theta,C) := \int_0^\infty \ln \lambda_{\theta;C}(t)N(dt) - \int_0^\infty \lambda_{\theta;C}(t)dt.$$
On the one side,
$$\int_0^\infty \ln \lambda_{\theta;C}(t)N(dt) = \sum_{i=1}^n \ln \lambda_{\theta;C}(t_i) = n\ln C + \sum_{i=1}^n \ln f_\theta(t_i).$$
On the other side,
$$\int_0^\infty \lambda_{\theta;C}(t)dt = C\int_0^\infty f_\theta(t)dt = C,$$
since f_θ is a probability density. All in all, we have
$$\ell(\theta,C) = n\ln C - C + \sum_{i=1}^n f_\theta(t_i). \tag{1}$$
The likelihood is separable with respect to the two parameters C and θ. Hence, the optimization can be done separately. In particular, for any C, the maximum of ℓ is reached for all $\theta^* \in \arg\max_\theta \sum_{i=1}^n f_\theta(t_i)$.

2. Suppose that $(t_1,\dots,t_n)$ is a n-sample (in the classical sense) of the probability density function f_θ. By definition, its log-likelihood is
$$\ell(\theta) := \sum_{i=1}^n f_\theta(t_i).$$
Hence, it is maximal for exactly the same values $\theta^* \in \arg\max_\theta \sum_{i=1}^n f_\theta(t_i)$.

Remark: Even under this very general framework, one could give the MLE of C in the Poisson framework. Indeed, in terms of the variable C, we can differentiate the log-likelihood (1) as
$$\frac{\partial \ell}{\partial C}(\theta,C) = n\frac{1}{C} - 1.$$
In particular, we have $\frac{\partial \ell}{\partial C}(\theta,C) \geq 0$ iff $C \leq n$ which in turn implies that $\ell(\theta,C)$ is maximal at $C^* = n$.

Exercise 2 (Goodness-of-fit tests)**.** See the Julia notebook.

2 Intermediate exercises

Exercise 3 (MLE - Hawkes process)**.** Consider the Hawkes process model N 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. By definition, the log-likelihood is

$$\ell(\theta) = \int_0^T \ln \lambda_t N(dt) - \int_0^T \lambda_t dt.$$

On the one side,

$$\int_0^T \ln \lambda_t N(dt) = \sum_{i=1}^n \ln \left(\mu + \int_0^{t_i} \alpha e^{-\beta(t_i - t')} N(dt') \right) = \sum_{i=1}^n \ln \left(\mu + \alpha A(i) \right),$$

since

$$\int_0^{t_i} \alpha e^{-\beta(t_i - t')} N(dt') = \alpha \sum_{j=1}^{i-1} e^{-\beta(t_i - t_j)} = \alpha A(i).$$

On the other side,

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

By Fubini, we have

$$\int_0^T \int_0^t e^{-\beta(t-t')} N(dt') dt = \int_0^T \int_{t'}^T e^{-\beta(t-t')} dt N(dt') = \int_0^T \frac{1}{\beta} (e^{-\beta(T-t')} - 1) N(dt').$$

Finally, writing the integral with respect to the points measure N as a sum, we have

$$\int_0^T \int_0^t e^{-\beta(t-t')} N(dt') dt = \frac{1}{\beta} \sum_{i=1}^n (e^{-\beta(T-t_i)} - 1).$$

All in all, it gives the log-likelihood,

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

2. The computations can be found in the paper : Ozaki, T. Maximum likelihood estimation of Hawkes' self-exciting point processes. Ann Inst Stat Math 31, 145–155 (1979). `https://doi.org/10.1007/BF02480272`

3. See the Julia notebook.

Exercise 4 (Population dynamics)**.** Let us denote $A := \{\mathrm{b}, \mathrm{d}, \mathrm{e}, \mathrm{i}\}$ the set of event types. For all $\alpha \in A$, let us denote $0 < t_1^\alpha < \cdots < t_{n_\alpha}^\alpha < T$ the observed times of events of type α during the time interval $[0, T]$. The log-likelihood of observing these four sets of times is the sum of each one. More precisely,

$$\ell(\Theta) := \sum_{\alpha \in A} \left(\int_0^T \ln \lambda_t^\alpha N^\alpha(dt) - \int_0^T \lambda_t^\alpha dt \right).$$

On the one side, we have, for $\alpha = \mathrm{i}$,

$$\int_0^T \ln \lambda_t^{\mathrm{i}} N^{\mathrm{i}}(dt) = \sum_{j=1}^{n_{\mathrm{i}}} \ln \theta_{\mathrm{i}} = n_{\mathrm{i}} \ln \theta_{\mathrm{i}},$$

and, for $\alpha \neq \mathrm{i}$,

$$\int_0^T \ln \lambda_t^\alpha N^\alpha(dt) = \sum_{j=1}^{n_\alpha} \ln \left(\theta_\alpha X_{t_j^\alpha -} \right) = n_\alpha \ln \theta_\alpha + \sum_{j=1}^{n_\alpha} \ln X_{t_j^\alpha -}.$$

On the other side, we have, for $\alpha = \mathrm{i}$,

$$\int_0^T \lambda_t^{\mathrm{i}} dt = \int_0^T \theta_{\mathrm{i}} dt = \theta_{\mathrm{i}} T,$$

and, for $\alpha \neq \mathrm{i}$,

$$\int_0^T \lambda_t^\alpha dt = \int_0^T \theta_\alpha X_{t-} dt = \theta_\alpha R(T),$$

where $R(T) := \int_0^T X_{t-} dt$ is the weighted total time at risk of the population.

Then, it is clear that the four parameters can be optimized separately, and the MLE of Θ is:

$$\hat{\Theta} = \left(\frac{n_{\mathrm{b}}}{R(T)}, \frac{n_{\mathrm{d}}}{R(T)}, \frac{n_{\mathrm{e}}}{R(T)}, \frac{n_{\mathrm{i}}}{T} \right).$$