Advanced Probabilistic Machine Learning and Applications

Pablo Sánchez and Isabel Valera

Tutorial 5: Introduction to Temporal Point Process (TPP)

In this tutorial we will work with Temporal Point Processes (TPPs) used to model discrete events in continuous time. TTPs characterize the event times using the conditional intensity function $\lambda^*_{\boldsymbol{\theta}}(t)$. We will learn (and implement in Python) how to sample and how to find the ML estimate of the parameters $\boldsymbol{\theta}$ for three different forms of $\lambda^*_{\boldsymbol{\theta}}(t)$:

- Homogeneous Poisson process
- Inhomogeneous Poisson process
- Hawkes process

This document contains an explanation of both operations (i.e., sampling and estimation) for each of the intensity functions. On the other hand, there is a jupyter notebook available in the Github repository of the course with the actual implementation in Python[1].

Introduction

Notation: Through this document we will use the following notation:

- $H(t) = \{t_i | t_i < t\}$: history of all events up to, but not including, time t.
- t_i: occurrence time of the i-th event.
- $g^*(t)$: A function g with an asterisk * indicates that it depens on $H(t)$.
- $f^*(t)$: probability density function of the time t that a new event is generated.
- $S^*(t)$: survival function that accounts for the probability that the next event will not occur before time t.
- $\lambda^*_\theta(t)$: conditional intensity function parameterized by θ, i.e., rate of events.

The probability density function of a TPP takes the form

$$f^*(t) = \lambda^*(t)\, S^*(t) = \lambda^*(t)\ \exp\left(-\int_{t_n}^{t} \lambda^*(\tau) d\tau\right)$$

where t_n refers to the last event in $H(t)$. Thus, given a contitional intensity function $\lambda^*_{\boldsymbol{\theta}}(t)$ with a set of parameters $\boldsymbol{\theta}$ we can compute the log-likelihood that $\lambda^*_{\boldsymbol{\theta}}(t)$ generates a specific history of events $H(T) = \{t_1, \dots, t_n\}$ in an interval $[0, T)$ as

[1] We use the Python package CVXPY to solve the convex optimization problems.

$$\mathcal{L}(H(T);\boldsymbol{\theta}) = \log \prod_{i=1}^{n} f^*(t_i) = \sum_{i=1}^{n} \log \lambda^*_{\boldsymbol{\theta}}(t_i) - \sum_{i=1}^{n} \int_{t_{i-1}}^{t_i} \lambda^*_{\boldsymbol{\theta}}(\tau) d\tau = \sum_{i=1}^{n} \log \lambda^*_{\boldsymbol{\theta}}(t_i) - \int_0^T \lambda^*_{\boldsymbol{\theta}}(\tau) d\tau$$

where we assume $t_0 = 0$. Then, we can use the above expression to find the parameters $\boldsymbol{\theta}$ that maximize the likelihood as

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \mathcal{L}(H(T);\boldsymbol{\theta})$$

Submission: This tutorial is a guided demonstration and do not need any submission. It is intended to serve as the basis for the next tutorial session.

1. Homogeneous Poisson process

The conditional intensity function of a homogeneous Poisson process is a constant

$$\lambda^*(t) = \mu$$

which means it is independent of the history. We can compute the log-likelihood of a sequence of n events as

$$\begin{aligned} \mathcal{L}(H(T),\mu) &= \sum_{i=1}^{n} \log \mu - \int_0^T \mu d\tau \\ &= n \log \mu - \mu T \end{aligned}$$

Then, we compute the ML estimate of μ taking into consideration it must be larger or equal than zero

$$\hat{\mu} = \arg\max_{\mu} \mathcal{L}(H(T),\mu) \quad \text{subject to} \quad \mu \geq 0$$

taking the derivative we obtain

$$\frac{\partial \mathcal{L}}{\partial \mu} = \frac{n}{\mu} - T \quad \rightarrow \quad \hat{\mu} = \frac{n}{T}$$

If we plug the ML estimate in the probabily density function, we observe it has the form of a exponential distribution shifted to the time of the last event t_n. Therefore, we can sample from it using inversion sampling

$$\begin{aligned} f^*(t) &= \hat{\mu} \exp\left(-\int_{t_n}^{t} \hat{\mu} d\tau\right) \\ &= \frac{n}{T} \exp\left(-\frac{n}{T}(t - t_n)\right) \end{aligned}$$

2. Inhomogeneous Poisson Process

The conditional intensity function of an inhomogeneous Poisson process is a time-varying function. In this tutorial, we choose it to be a mixture of N Radial Basis Function (RBF) kernels

$$\lambda^*(t) = g_{\boldsymbol{\theta}}(t) = \sum_{j=1}^{N} \theta_j \exp\left(-\beta \left(t - \tau_j\right)^2\right)$$

where $\{\tau_j\}$ and β are given constants. Then, we can compute the log-likelihood of a sequence of n events as

$$\begin{aligned}
\mathscr{L}(H(T); \{\theta_j\}) &= \sum_{i=1}^{n} \log(g_\theta(t_i)) - \int_0^T g_\theta(\tau) d\tau \\
&= \sum_{i=1}^{n} \log\left(\sum_{j=1}^{N} \theta_j \exp\left(-\beta \left(t_i - \tau_j\right)^2\right)\right) - \sum_{j=1}^{N} \int_0^T \theta_j \exp\left(-\beta \left(\tau - \tau_j\right)^2\right) d\tau \\
&= \sum_{i=1}^{n} \log\left(\sum_{j=1}^{N} \theta_j \exp\left(-\beta \left(t_i - \tau_j\right)^2\right)\right) - \sum_{j=1}^{N} \theta_j \int_0^T \exp\left(-\beta \left(\tau - \tau_j\right)^2\right) d\tau \\
&= \sum_{i=1}^{n} \log\left(\sum_{j=1}^{N} \theta_j C_{ij}\right) - \sum_{j=1}^{N} \theta_j K_j
\end{aligned}$$

where we have defined K_j and C_{ij} as

$$K_j = \int_0^T \exp\left(-\beta \left(\tau - \tau_j\right)^2\right) d\tau = \frac{\sqrt{\pi}\left(\operatorname{erf}(\tau_j\sqrt{\beta}) + \operatorname{erf}(\sqrt{\beta}(T - \tau_j))\right)}{2\sqrt{\beta}}$$

$$C_{ij} = \exp\left(-\beta \left(t_i - \tau_j\right)^2\right)$$

Now, we compute the ML estimate of $\{\theta_j\}$ taking into consideration they must be larger or equal than zero

$$\hat{\theta}_j = \underset{\theta_j}{\arg\max}\, \mathscr{L}(H(T); \{\theta_j\}) \quad \text{subject to} \quad \theta_j \geq 0$$

taking the partial derivative we obtain

$$\frac{\partial \mathscr{L}}{\partial \theta_j} = \sum_{i=1}^{n} \frac{C_{ij}}{\sum_{j=1}^{N} \theta_j C_{ij}} - K_j \quad \rightarrow \quad \text{Convex optimization}$$

We can not obtain a closed form solution. However, we know our objective function is convex on the parameters so we can solve it using convex optimization techniques.

3. Hawkes Process

The conditional intensity function of a Hawkes is history dependent. It considers a self-excitatory property in the events, or in other words, it assumes that each event increases the intensity according to some triggering kernel $k_w(\cdot)$. The intensity has the following form

$$\lambda^*(t) = \mu + \alpha \sum_{t_i \in H(t)} k_w(t - t_i)$$

where μ (known as the baseline intensity) and α are the parameters. In this tutorial, we choose an exponential triggering kernel $k_w(t) = \exp(-wt)\mathbb{I}[t \geq 0]$. Then, we can write the log-likelihood as

$$
\begin{aligned}
\mathscr{L}(H(T);\mu,\alpha) &= \sum_{i=1}^{n} \log\left(\mu + \alpha \sum_{t_j \in H(t)} k_w\left(t_i - t_j\right)\right) - \int_0^T \left(\mu + \alpha \sum_{t_i \in H(t)} k_w(\tau - t_i)\right) d\tau \\
&= \sum_{i=1}^{n} \log\left(\mu + \alpha \sum_{t_j \in H(t)} k_w\left(t_i - t_j\right)\right) - T\mu - \alpha \sum_{t_i \in H(t)} \int_0^T k_w(\tau - t_i)\, d\tau \\
&= \sum_{i=1}^{n} \log(\mu + \alpha C_i) - T\mu - \alpha K
\end{aligned}
$$

where K and C_i are constants defined as

$$
\begin{aligned}
K &= \sum_{t_i \in H(t)} \int_0^T k_w(\tau - t_i)\, d\tau \\
&= \sum_{t_i \in H(t)} \int_{t_i}^T \exp(-w(\tau - t_i))\, d\tau \\
&= \sum_{t_i \in H(t)} \frac{1 - \exp(-w(T - t_i))}{w}
\end{aligned}
$$

$$
C_i = \sum_{t_j \in H(t)} k_w\left(t_i - t_j\right) = \exp\left(-w\left(t_i - t_j\right)\right)\mathbb{I}[t_i \geq t_j]
$$

Now, we compute the ML estimate of α and μ taking into consideration they must be larger or equal than zero

$$
\hat{\alpha}, \hat{\mu} = \underset{\alpha,\mu}{\arg\max}\ \mathscr{L}(H(T);\mu,\alpha) \quad \text{subject to} \quad \mu \geq 0, \alpha \geq 0
$$

taking the partial derivative we obtain

$$
\frac{\partial \mathscr{L}}{\partial \mu} = \sum_{i=1}^{n} \frac{1}{\mu + \alpha C_i} - T \quad \rightarrow \quad \text{Convex optimization}
$$

$$
\frac{\partial \mathscr{L}}{\partial \alpha} = \sum_{i=1}^{n} \frac{C_i}{\mu + \alpha C_i} - K \quad \rightarrow \quad \text{Convex optimization}
$$

We can not obtain a closed form solution. However, we know our objective function is convex on the parameters so we can solve it using convex optimization techniques.