Classifications of Poisson processes and spatial Poisson point processes

Mamadou DIOUF

Internship thesis
Laboratoire d'Analyse et de Mathématiques Appliquées (LAMA UMR8050)
Under the supervision of Miguel MARTINEZ
Mathematics and Computer science M2 Track

September 4, 2025

Introduction

- A Poisson process is a mathematical model used to describe random events that occur independently over time or space at a constant average rate. It is characterized by a single parameter, λ (lambda), which represents the average rate of events per unit of time (or space).
- In machine learning, classification is a supervised learning task where the goal is to predict the categorical class label for new, unseen data points based on patterns learned from a labeled dataset. Essentially, we're teaching a computer to categorize things.
- Our goal is to categorize the Poisson processes by there parameters λ, the labels, find the better λ for new Poisson process in our labels.

Homogeneous Poisson Process

- A counting process is a stochastic process with a series of random variables that represent the number of events in a given time interval, and can be denoted as

$$\{N(t); t \geq 0\}$$

where $N(t)$: Number of events in the time interval $(0, t]$.
It is assumed that $N(0) = 0$, i.e. at $t = 0$, no event occurs.
Note that

$N(t) - N(s)$: Number of events in the time interval $(s, t]$ for $s < t$

Homogeneous Poisson Process

- Stationary increments occur if the distribution of the number of events depends only on the length of the time interval. As an example, in case of stationary increments, $N(t+s)-N(s)$ will have the same distribution with $N(t)$ for all $s \geq 0,\ t > 0$, since the length of the time interval is t for both increments.
- Independent increments occur if the number of events in disjoint time intervals is independently distributed. As an example, in case of independent increments, $N(t)-N(s)$ and $N(s)$ will be independently distributed for all $s \geq 0,\ t > 0$, since $(0, s]$ and $(s, t]$ are disjoint intervals .

Homogeneous Poisson Process

- A counting process $\{N(t); t \geq 0\}$ is said to be a Poisson process with rate λ, $\lambda > 0$, if

$$N(t) = \sum_{n \geq 1} \mathbb{1}_{(T_n \leq t)}$$

for a sequence $(T_n)_{n \geq 1}$ with $Gamma(n, \lambda)$ having i.i.d. increments inter-arrival times $(I_n)_{n \geq 1}$ with $Exponential(\lambda)$ distribution, $I_n = T_n - T_{n-1}$.

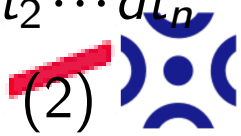
Proposition

Homogeneous Poisson Process

Proposition 1: Let $A \in \mathbb{R}^n$, $T_0 = 0$ and $T_n = \sum_{i=1}^{n} I_i$, $n \in \mathbb{N}^*$ then

$\forall t \geq 0$, $N(t) = \sum_{n \geq 1} \mathbb{1}_{(T_n \leq t)}$, we have:

1. $(I_1, ..., I_n) \sim \lambda^n e^{-\lambda \sum_{k=1}^{n} i_k}$
2. $$\mathbb{P}((I_1, ..., I_n) \in A, N(t) = n) = \int_A e^{-\lambda t} \lambda^n \mathbb{1}_{i_1 + i_2 + \cdots + i_n \leq t} di_1 di_2 \cdots di_n \tag{1}$$
3. $(T_1, ..., T_n) \sim \lambda^n e^{-\lambda t_n} \mathbb{1}_{0 < t_1 < \cdots < t_n}$.
4. $$\mathbb{P}((T_1, ..., T_n) \in A, N(t) = n) = \int_A e^{-\lambda t} \lambda^n \mathbb{1}_{0 < t_1 < t_2 < \cdots < t_n \leq t} dt_1 dt_2 \cdots dt_n \tag{2}$$



Proof

1. $(I_k), k \in [\![1; n]\!]$ are i.i.d .
2. We have $N(t) = n \iff (T_n \leq t < T_{n+1})$, $N(t)$ a Poisson process with rate λ and $t_j = i_1 + \cdots + i_j$.

$$\mathbb{P}((I_1, \cdots, I_n) \in A, N(t) = n) = \mathbb{P}((I_1, \cdots, I_n) \in A, T_n \leq t < T_{n+1})$$

$$= \iint_{A \times \mathbb{R}_+^*} \mathbb{1}_{i_1 + i_2 + \cdots + i_n \leq t < t_{n+1}} \lambda^{n+1} e^{-\lambda t_{n+1}} di_1 \cdots di_n di_{n+1}$$

$$= \int_A \mathbb{1}_{t_n \leq t} \lambda^n e^{-\lambda t_n} di_1 \ldots di_n \left(\int_{t - t_n}^{+\infty} \lambda e^{-\lambda i_{n+1}} di_{n+1} \right)$$

$$= \int_A e^{-\lambda t} \lambda^n \mathbb{1}_{t_n \leq t} di_1 \cdots di_n.$$

Proof

3. we have $N(t) = n \iff (T_n \leq t < T_{n+1})$, and $N(t)$ a Poisson process with rate λ. Let h be a Borel measurable function positive definite in $\mathbb{R}^n$. Since the r.v $I_1, \ldots, I_n$ are independents, we have

$$E(h(T_1, \ldots, T_n)) = \int_D h(i_1, i_1 + i_2, \ldots, i_1 + \cdots + i_n) \lambda^n e^{-\lambda i_n} di_1 \ldots di_n,$$

where $D \subset (\mathbb{R}_+^*)^n$. By making the change of variable $t_j = i_1 + \cdots + i_j$, $j = 1, \ldots, n$, we obtain

$$E(h(T_1, \ldots, T_n)) = \int_{\mathbb{R}^n} h(t_1, \ldots, t_n) \lambda^n e^{-\lambda t_n} 1_{0<t_1<\cdots<t_n} dt_1 \ldots dt_n.$$

Hence the density of $(T_1, \ldots, T_n)$ is : $\lambda^n e^{-\lambda t_n} 1_{0<t_1<\cdots<t_n}$.

Proof

4. We have $P(N(t)=n)=e^{-\lambda t}\frac{(\lambda t)^n}{n!}$ and $N(t)=n \iff T_n \le t < T_{n+1}$. By the result of 1 we obtain,

$$\mathbb{P}((T_1,\dots,T_n)\in A, N(t)=n)=\mathbb{P}((T_1,\dots,T_n)\in A, T_n\le t<T_{n+1})$$

$$=\iint_{\mathbb{A}\times\mathbb{R}} 1_{0<t_1<\cdots<t_n\le t<t_{n+1}}\lambda^{n+1}e^{-\lambda t_{n+1}}dt_1\dots dt_n dt_{n+1}$$

$$=\int_A 1_{0<t_1<\cdots<t_n\le t}\lambda^n dt_1\dots dt_n\left(\int_t^{+\infty}\lambda e^{-\lambda t_{n+1}}dt_{n+1}\right)$$

$$=\int_A e^{-\lambda t}\lambda^n 1_{0<t_1<\cdots<t_n\le t}dt_1\dots dt_n.$$ Hence

$$\mathbb{P}((T_1,\dots,T_n)\in A, N(t)=n)=\int_A e^{-\lambda t}\lambda^n 1_{0<t_1<\cdots<t_n\le t}dt_1\dots dt_n.$$

[3] □

Simulation of Poisson process

Simulation of Poisson process

Algorithm 1 Poisson Process Simulation Algorithm

Require: λ (lambda): The average rate of events (events per unit of time).
Require: $total\_time$: The total duration of the simulation.
Ensure: $arrival\_times$: A list of the times at which events occur.
Ensure: $time\_event\_counts$: A tuple containing two lists:
$times$: A list of time points.
$event\_counts$: A list of the number of events that have occurred up to each time point in $times$.

$arrival\_times \leftarrow []$
$current\_time \leftarrow 0$
while $current\_time < total\_time$ **do**
    Generate $inter\_arrival\_time$ from an exponential distribution with mean $1/\lambda$.
    $current\_time \leftarrow current\_time + inter\_arrival\_time$
    if $current\_time < total\_time$ **then**
        Append $current\_time$ to $arrival\_times$
    end if
end while
Generate a list of time points, $times$, from 0 to $total\_time$.
Create an empty list $event\_counts$.
for each $time$ in $times$ **do**
    $count \leftarrow$ number of $arrival\_times$ that are $\leq time$
    Append $count$ to $event\_counts$
end for
return $arrival\_times$ and $(times, event\_counts)$

Point Poisson processes

- Poisson point process (also known as: Poisson random measure, Poisson random point field and Poisson point field) with rate λ is a type of mathematical object that consists of points randomly located on a mathematical space with the essential feature that the points occur independently of one another.
- If the rate λ is constant, it is the average density of the points in the Poisson process located in some region of space. The resulting point process is called a homogeneous or stationary Poisson point process. Otherwise the point process is called an inhomogeneous or nonhomogeneous Poisson point process, and the average density of points depend on the location of the underlying space of the Poisson point process.

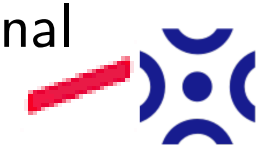
Homogeneous Poisson point process

- A spatial Poisson process is a Poisson point process defined in $\mathbb{R}^d$. For some bounded region S of Euclidean space $\mathbb{R}^d$, if the points form a homogeneous Poisson process with parameter $\lambda > 0$, then the probability of n points existing in $S \subset \mathbb{R}^d$ is given by:

$$\mathbb{P}\big(N(S) = n\big) = \frac{(\lambda |S|)^n}{n!} e^{-\lambda |S|}$$

  where $|S|$ now denotes the d-dimensional volume of S.

- Furthermore, for a collection of disjoint, bounded Borel sets $S_1, \ldots, S_k \subset \mathbb{R}^d$, let $N(S_i)$ denote the number of points of N existing in S_i. Then the corresponding homogeneous Poisson point process with parameter $\lambda > 0$ has the finite-dimensional distribution:



Homogeneous Poisson point process

$$\mathbb{P}\big(N(S_i) = n_i, i = 1, \ldots, k\big) = \prod_{i=1}^{k} \frac{(\lambda |S_i|)^{n_i}}{n_i!} e^{-\lambda |S_i|}.$$

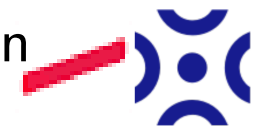
- **Positioning of points**
  Homogeneous Poisson point processes do not depend on the position of the underlying space through its parameter λ, which implies it is both a stationary process (invariant to translation) and an isotropic (invariant to rotation) stochastic process.

Inhomogeneous Poisson point process

- The inhomogeneous or nonhomogeneous Poisson point process is a Poisson point process with a Poisson parameter set as some location-dependent function in the underlying space on which the Poisson process is defined. For Euclidean space $\mathbb{R}^d$, this is achieved by introducing a locally integrable positive function $\lambda : \mathbb{R}^d \to [0, \infty)$, such that for every bounded region S the (d-dimensional) volume integral of $\lambda(x)$ over region S is finite. In other words, if this integral, denoted by $\Lambda(S)$, is:

$$\Lambda(S) = \int_S \lambda(x)\,\mathrm{d}x < \infty,$$

- where $\mathrm{d}x$ is a (d-dimensional) volume element,then for every collection of disjoint bounded Borel measurable sets $S_1, \ldots, S_k$, an inhomogeneous Poisson process with (intensity) function $\lambda(x)$ has the finite-dimensional distribution:



Spatial Poisson point process

$$\mathbb{P}\big(N(S_i) = n_i, i = 1, \ldots, k\big) = \prod_{i=1}^{k} \frac{(\Lambda(S_i))^{n_i}}{n_i!} e^{-\Lambda(S_i)}.$$

- Furthermore, $\Lambda(S)$ has the interpretation of being the expected number of points of the Poisson process located in the bounded region S, namely $\Lambda(S) = \mathrm{E}[N(S)]$.
- An inhomogeneous Poisson process defined in the plane $\mathbb{R}^2$ is called a spatial Poisson process It is defined with intensity function and its intensity measure is obtained performing a surface integral of its intensity function over some region. For example, its intensity function (as a function of Cartesian coordinates x and y) can be

Spatial Poisson point process

$$\lambda(x,y) = e^{-(x^2+y^2)},$$

so the corresponding intensity measure is given by the surface integral

$$\Lambda(S) = \int_S e^{-(x^2+y^2)}\,\mathrm{d}x\,\mathrm{d}y,$$

where S is some bounded region in the plane $\mathbb{R}^2$.

In higher dimensions, in the plane,$\Lambda(S)$ corresponds to a surface integral while in $\mathbb{R}^d$ the integral becomes a (d-dimensional) volume integral.

Spatial Poisson point process

- A couple of different methods can be used depending on the nature of the intensity function $\lambda(x)$. If the intensity function is sufficiently simple, then independent and random non-uniform (Cartesian or other) coordinates of the points can be generated. For example, simulating a Poisson point process on a circular window can be done for an isotropic intensity function (in polar coordinates r and θ), implying it is rotationally variant or independent of θ but dependent on r, by a change of variable in r if the intensity function is sufficiently simple.
- For more complicated intensity functions, one can use an acceptance-rejection method, which consists of using (or 'accepting') only certain random points and not using (or 'rejecting') the other points, based on the ratio:

$$\frac{\lambda(x_i)}{\Lambda(S)} = \frac{\lambda(x_i)}{\int_S \lambda(x)\,\mathrm{d}x.}$$

Spatial Poisson point process

where x_i is the point under consideration for acceptance or rejection.

- That is, a location is uniformly randomly selected for consideration, then to determine whether to place a sample at that location a uniformly randomly drawn number in $[0,1]$ is compared to the probability density function $\frac{\lambda(x)}{\Lambda(S)}$, accepting if it is smaller than the probability density function, and repeating until the previously chosen number of samples have been drawn.

Inhomogeneous Poisson process in a separable and bounded metric space

Let (S,ρ) be a separable and bounded metric space, endowed with a Borel measure ν, let us denote by $\mathscr{B}(S)$ the Borel σ-algebra on S and by S^{∞} the set of elements (subsets) x of S whose cardinality, $\#x$, is finite. That is,

$$S^{\infty} = \{x \subset S : \#x < \infty\}.$$

Let $\lambda : S \to \mathbb{R}^{+}$ be an integrable function. Given a probability space $(\Omega, \mathscr{A}, \mathbb{P})$, we will say that a function $X : \Omega \to S^{\infty}$ is a *Poisson process on S with intensity λ* (we will denote $X \sim \mathscr{P}(S,\lambda)$) if:

Spatial Poisson point process

- the functions $N_A : \Omega \to \{0, \ldots, \infty\}$ defined by $N_A(\omega) = \#\{\omega : X(\omega) \cap A\}$ are random variables for all $A \in \mathscr{B}(S)$;
- given n disjoint Borel subsets $A_1, \ldots, A_n$ of S, the random variables $N_{A_1}, \ldots, N_{A_n}$ are independent;
- N_A follows a Poisson process with mean $\mu(A)$ (we will write $N_A \sim \mathscr{P}(\mu(A))$), with

$$\mu(A) = \int_A \lambda(\xi) d\nu(\xi).$$

Let $S^\infty = 2^S$ be the σ-algebra of part of S^∞. If X is a Poisson process, the distribution $\mathbb{P}_X$ of X on S^∞ is defined by $\mathbb{P}_X(B) = \mathbb{P}(X \in B)$ for $B \in S^\infty$.
for all $X \sim \mathscr{P}(S, \lambda)$, its density is

$$f_X(x) = \exp\{\nu(S) - \mu(S)\} \prod_{\xi \in x} \lambda(\xi)$$

where $\mu(S) = \int_S \lambda d\nu$.

Maximum Likelihood Estimation (MLE)

The maximum likelihood estimation (MLE) is a method of estimating the parameters of an assumed probability distribution, given some observed data. This is a maximum of the likelihood function under the assumed statistical model, where the observed data is most probable.
Let $\mathbb{X}$ be our observed data and λ our parameter in the space $\mathscr{Y}$. if $\mathbb{X}|\lambda$ has a:

- Discrete distribution the likelihood function is $L(\lambda|\mathbb{X}) = \mathbb{P}(\mathbb{X}|\lambda)$ and its MLE is $\arg\max_{\lambda\in\mathscr{Y}} L(\lambda|\mathbb{X})$.
- Continuous distribution the likelihood function is $L(\lambda|\mathbb{X}) = f(\mathbb{X}|\lambda)$ and its MLE is $\arg\max_{\lambda\in\mathscr{Y}} L(\lambda|\mathbb{X})$.

Maiximum A Posteriori (MAP) estimation

Let $\mathbb{X}$ be our observed data, λ our parameter in the space $\mathscr{Y}$ with $\pi(\lambda)$ its probability distribution. if $\pi(\lambda)$ is a:

- Discrete distribution the posterior probability is :
  $$\pi(\lambda|\mathbb{X}) = \frac{L(\lambda|\mathbb{X})\pi(\lambda)}{\sum_{\lambda' \in \mathscr{Y}} L(\lambda'|\mathbb{X})\pi(\lambda')}$$
  and its MAP estimation is $\arg\max_{\lambda \in \mathscr{Y}} \pi(\lambda|\mathbb{X})$.
- Continuous distribution the posterior probability is :
  $$\pi(\lambda|\mathbb{X}) = \frac{L(\lambda|\mathbb{X})\pi(\lambda)}{\int_{\lambda' \in \mathscr{Y}} L(\lambda'|\mathbb{X})\pi(\lambda')d\lambda'}$$
  and its MAP estimation is $\arg\max_{\lambda \in \mathscr{Y}} \pi(\lambda|\mathbb{X})$.

Classification

- Classification is supervised learning problem, its goal is to learn a mapping from inputs x to outputs y, where $y \in \{\lambda_1, ..., \lambda_J\}$, with J being the number of classes. If $J = 2$, this is called binary classification (in which case we often assume $y \in \{0, 1\}$); if $J > 2$, this is called multiclass classification.
- One way to formalize the problem is as function approximation. We assume $y = g(x)$ for some unknown function f, and the goal of learning is to estimate the function f given a labeled training set, and then to make predictions using $\widehat{y} = \widehat{g}(x)$. Our main goal is to make predictions on novel inputs, meaning ones that we have not seen before .

Prediction algorithm and loss function

A prediction function is a (measurable) function $g : \mathscr{X} \to \mathscr{Y}$. Let $\mathscr{F}(\mathscr{X}, \mathscr{Y})$ be the set of all prediction functions, $Z_i = (X_i, Y_i) \sim P$. $\{Z_1, ..., Z_n\}$ is the training set.
An algorithm is any function that takes a training set to produce a prediction:

$$G : \cup_{n=1}^{\infty} \mathscr{Z}^n \to \mathscr{F}(\mathscr{X}, \mathscr{Y}); \qquad \mathscr{Z} = (\mathscr{X}, \mathscr{Y}).$$

Let $l(y, y')$ be the resulting loss when the true label is y and the predicted label y'. The function $l : \mathscr{Y} \times \mathscr{Y} \to \mathbb{R}$ is the cost function.

Bayes predictor

The quality of a prediction function $g : \mathscr{X} \to \mathscr{Y}$ is measured through its risk (or generalization error):

$$R_P(g) = \mathrm{E}[l(Y, g(X))] = \int_{\mathscr{X},\mathscr{Y}} l(y, g(x)) d\mathbb{P}(x, y)$$

The "best" prediction function is the (or more precisely, a) function of $\mathscr{F}(\mathscr{X},\mathscr{Y})$ minimizing the risk R_P:

$$g^* \in \arg \min_{g \in \mathscr{F}(\mathscr{X},\mathscr{Y})} R_P(g)$$

Such a function g^* might not exist but exits for usual loss functions. The "best" function is called *Bayes predictor*. It depends on the unknown probability P and is the unknown itself.

Bayes Classifier

Proposition 2: The Bayes classifier $\arg\max_{y \in \mathscr{Y}} \mathbb{P}\big(Y = y \big| X\big)$, the MAP estimation of y, is a Bayes predictor when the loss is (0-1 loss) $l(y, y') = \mathbb{1}(y \neq y')$.

Gaussian naive Bayes

- Let $\mathscr{D} = \{(\mathbb{X}_1, Y_1), \cdots, (\mathbb{X}_n, Y_n)\}$ be a given data set, with $\mathbb{X}_i \in \mathbb{R}^d$ and $Y_i \in \mathscr{Y} = \{\lambda_1, \cdots, \lambda_K\}$ classes set for $i \in [\![1, n]\!]$. Let $\mathbb{X}_i = (X_{i1}, \cdots, X_{id})$ with X_{ij} i.i.d for a fixed i and $j \in [\![1, d]\!]$.
- We segment our data by class and compute the mean and variance of variable of $\mathbb{X}_i$. Let μ_{kj} and σ^2_{kj} be respectively the mean and the variance of $\mathbb{X}_i$ associated with class $\lambda_k, k \in [\![1, K]\!]$.
- For a observation $\mathbb{X} = (X_1, \cdots, X_d)$, X_i i.i.d assume that $X_j = x_j | Y = \lambda_k \sim \mathscr{N}(\mu_{kj}, \sigma_{kj})$ which is

$$f(x_j | \mu_{kj}, \sigma_{kj}) = \frac{1}{\sqrt{2\pi\sigma^2_{kj}}} e^{-\frac{(x_j - \mu_{kj})^2}{2\sigma^2_{kj}}}$$

Classification

- Thus the likelihood is $L(\lambda_k|\mathbb{X}) = \prod_{j=1}^{d} f(x_j|\mu_{kj}, \sigma_{kj})$ and the posterior probability is
$$\pi(\lambda_k|\mathbb{X}) = \frac{L(\lambda_k|\mathbb{X})\pi(\lambda_k)}{\sum_{\lambda'_k \in \mathcal{Y}} L(\lambda'_k|\mathbb{X})\pi(\lambda'_k)} \ .$$
- The MAP estimation of λ_k is the gaussian naive Bayes classifier : $\arg\max_{\lambda_k \in \mathcal{Y}} \pi(\lambda_k|\mathbb{X})$.

Application on Poisson processes

Bayes classifier of "Discrete Poisson process"

For a Poisson process $\{X_t, t \in [0, T]\}$ with rate $\lambda, \lambda > 0$, we discreetize the process in n times as $\{X_t, t \in \{t_1, t_2, \cdots, t_n\}\}$, $X_{t_i}, i \in [\![1, n]\!]$, with $t_{i-1} < t_i$. Hence $X_{t_i} \sim Poisson(\lambda t_i)$ and we will use this discretion to make a Bayes classifier of the rate λ .

Bayes classifier of a Poisson distribution

$\mathscr{Y}$ discrete and ordered set

Let $\mathscr{Y} = \{\lambda_1, \cdots, \lambda_K\}$, $p_j = \mathbb{P}(Y = \lambda_j)$ and $N_t^{\lambda_j} \sim Poisson(\lambda_j t)$ which is independent of Y. Assume that $(X_t | Y = \lambda) = N_t^{\lambda}$

$$\mathbb{P}(Y = \lambda_j | X_t = x) = \frac{\mathbb{P}(Y = \lambda_j, X_t = x)}{P(X_t = x)} = \frac{\mathbb{P}(X_t = x | Y = \lambda_j)\mathbb{P}(Y = \lambda_j)}{P(X_t = x)}$$

$$\mathbb{P}(Y = \lambda_j | X_t = x) = \frac{\mathbb{P}(X_t = x | Y = \lambda_j)\mathbb{P}(Y = \lambda_j)}{P(X_t = x)} =$$

$$\frac{\mathbb{P}(N_t^{\lambda_j} = x)\mathbb{P}(Y = \lambda_j)}{P(X_t = x)}$$

$$\arg\max_{\lambda_j} \mathbb{P}(Y = \lambda_j | X_t = x) = \arg\max_{\lambda_j} \mathbb{P}(X_t = x | Y = \lambda_j) =$$

$$\arg\max_{\lambda_j} \mathbb{P}(N_t^{\lambda_j} = x)\mathbb{P}(Y = \lambda_j)$$

$$\arg\max_{\lambda_j} \mathbb{P}(Y = \lambda_j | X_t = x) = \arg\max_{\lambda_j} \frac{(\lambda_j t)^x}{x!} e^{-\lambda_j t} p_j$$

Bayes classifier of a Poisson distribution

- if Y is uniformly distributed

$$\arg\max_{\lambda_j} \mathbb{P}(Y = \lambda_j | X_t = x) = \arg\max_{\lambda_j} \frac{(\lambda_j t)^x}{x!} e^{-\lambda_j t}$$

Let $f_t(\lambda) = \frac{(\lambda t)^x}{x!} e^{-\lambda t}$

$$f_t'(\lambda) = e^{-\lambda t}\left(\frac{(\lambda t)^{x-1} t}{(x-1)!} - \frac{(\lambda t)^x t}{x!}\right) = e^{-\lambda t}\frac{(\lambda t)^{x-1} t}{(x-1)!}\left(1 - \frac{\lambda t}{x}\right)$$

λ	0	$\frac{x}{t}$	$+\infty$
f_t	0 ↗	$f_t(\frac{x}{t})$	↘ 0

If there exists an m s.t $x = \lambda_m t$ then

$$\arg\max_{\lambda_j} \mathbb{P}(Y = \lambda_j | X_t = x) = \lambda_m \text{ and } \lambda_m = \frac{x}{t}$$

Bayes classifier of a Poisson distribution

- if Y is not uniformly distributed

  $$\arg\max_{\lambda_j} \mathbb{P}(Y = \lambda_j | X_t = x) = \arg\max_{\lambda_j} \frac{(\lambda_j t)^x}{x!} e^{-\lambda_j t} p_j$$

  Which is λ_m such that $\forall j \neq m$ $$\left(\frac{\lambda_m}{\lambda_j}\right)^x e^{(\lambda_j - \lambda_m)t} > \frac{p_j}{p_m}$$

- **Applications** For a "discrete Poisson process" in $\{X_{t_i}, i \in [\![1, n]\!]\}$ we can predict a λ for each time in $\{t_1, t_2, \cdots, t_n\}$ and take the rounding value of :
  - The mean predicted λ
  - The most predicted λ

Bayes classifier of Poisson process

Proposition 2 :Bayes classifier of Poisson process, $\mathscr{Y}$ discrete

Consider $\{X_t, t \in [0; T]\}$ a Poisson process and
$\mathbb{X} = (X_{t_1}, X_{t_2}, ..., X_{t_n-1}, X_T)$ with
$0 < t_1 < t_2 < \cdots < t_{n-1} < T$ and $X_{t_i} = x_i$ for $i \in [\![1, n-1]\!]$, $X_T = x_n$.
Let $\mathscr{Y} = \{\lambda_j, j \in [\![1, K]\!]\}$ and $x = (x_1, x_2, \cdots, x_n)$.

Bayes classifier of Poisson process

Case 1: Y uniformly distributed in $\mathscr{Y}$

The Bayes classifier of the rate λ is $\widehat{\lambda} = \frac{x_n}{T}$, if there exists an $\lambda_m \in \mathscr{Y}$ such that $x_n = \lambda_m t_n$. We have $\arg\max_{\lambda_j \in \mathscr{Y}} \mathbb{P}(Y = \lambda_j | \mathbb{X} = x) = \lambda_m$ and $\lambda_m = \frac{x_n}{T}$.

Otherwise :

- If $\forall \lambda_j \in \mathscr{Y}, x_n > \lambda_j T$ Then
  $\arg\max_{\lambda_j \in \mathscr{Y}} \mathbb{P}(Y = \lambda_j | \mathbb{X} = x) = \arg\max_{\lambda_j \in \mathscr{Y}} \lambda_j T = \max_{\lambda_j \in \mathscr{Y}} \lambda_j$
- If $\forall \lambda_j \in \mathscr{Y}, x_n < \lambda_j T$ Then
  $\arg\max_{\lambda_j \in \mathscr{Y}} \mathbb{P}(Y = \lambda_j | \mathbb{X} = x) = \arg\min_{\lambda_j \in \mathscr{Y}} \lambda_j T = \min_{\lambda_j \in \mathscr{Y}} \lambda_j$
- If $\forall \lambda_j \in \mathscr{Y}, \lambda_j \leq \lambda_m, x_n > \lambda_j T$ and if $\lambda_j > \lambda_m, x_n < \lambda_j T$ Then
  $\arg\max_{\lambda_j \in \mathscr{Y}} \mathbb{P}(Y = \lambda_j | \mathbb{X} = x) = \arg\max_{\lambda_j \in \{\lambda_m, \lambda_{m+1}\}} (\lambda_j)^{x_n} e^{-\lambda_j T}$

Bayes classifier of Poisson process

Case 2: Y not uniformly distributed in $\mathscr{Y}$

Let $\mathbb{P}(Y=\lambda_j)=p_j$, the Bayes classifier of the rate λ is $\arg\max_{\lambda_j\in\mathscr{Y}}\mathbb{P}(Y=\lambda_j|\mathbb{X}=x)$

Which is λ_m such that $\forall j\neq m$ $\quad \left(\frac{\lambda_m}{\lambda_j}\right)^{x_n} e^{(\lambda_j-\lambda_m)T} > \frac{p_j}{p_m}$

Proof 1:

Since a Poisson process is represented by the values of jump times then we can take $\mathbb{X}=(X_{t_1},\cdots,X_{t_{n-1}},X_T)$ with $t_i, i\in[\![1,n-1]\!]$ the jump times (we can also assume $\{t_i, i\in[\![1,n-1]\!]\}$ is any point process in $[0;T]$) .

Let $x=(x_1,\cdots,x_n)$

Bayes classifier of Poisson process

$$\mathbb{P}(Y=\lambda_j|\mathbb{X}=x)=\mathbb{P}(X_{t_1}=x_1,\cdots,X_{t_{n-1}}=x_{n-1},X_T=x_n|Y=\lambda_j)\mathbb{P}(Y=\lambda_j)$$

$$\mathbb{P}(X_{t_1}=x_1,\cdots,X_{t_{n-1}}=x_{n-1},X_T=x_n|Y=\lambda_j)$$
$$=\mathbb{P}(N^{\lambda_j}_{t_1}=x_1,N^{\lambda_j}_{t_2}=x_2,\cdots,N^{\lambda_j}_{t_{n-1}}=x_{n-1},N^{\lambda_j}_{T}=x_n)$$
$$=\mathbb{P}(N^{\lambda_j}_{t_1}=x_1,N^{\lambda_j}_{t_2}-N^{\lambda_j}_{t_1}=x_2-x_1,\cdots,N^{\lambda_j}_{T}-N^{\lambda_j}_{t_{n-1}}=x_n-x_{n-1})$$

Because if we know $N^{\lambda_j}_{t_1}=x_1$ we will have the others

$N^{\lambda_j}_{t_i}=x_i, i\in[\![2;n]\!]$

$$=\prod_{i=1}^{n}\mathbb{P}(N^{\lambda_j}_{t_i}-N^{\lambda_j}_{t_{i-1}}=x_i-x_{i-1})$$ ($(N^{\lambda_j}_{t_i}-N^{\lambda_j}_{t_{i-1}})_i$ are i.i.d with $N_{t_0}=0$ and $x_0=0$)

Bayes classifier of Poisson process

$$= \prod_{i=1}^{n} \frac{(\lambda_j(t_i - t_{i-1}))^{x_i - x_{i-1}}}{(x_i - x_{i-1})!} e^{-\lambda_j(t_i - t_{i-1})}$$

$$= \lambda_j^{\sum_{i=1}^{n}(x_i - x_{i-1})} e^{-\lambda_j \sum_{i=1}^{n}(t_i - t_{i-1})} \prod_{i=1}^{n} \frac{(t_i - t_{i-1})^{(x_i - x_{i-1})}}{(x_i - x_{i-1})!}$$

$$\mathbb{P}(X_{t_1} = x_1, \cdots, X_{t_{n-1}} = x_{n-1}, X_T = x_n | Y = \lambda_j) =$$

$$\lambda_j^{x_n} e^{-\lambda_j T} \prod_{i=1}^{n} \frac{(t_i - t_{i-1})^{(x_i - x_{i-1})}}{(x_i - x_{i-1})!}$$

Let $h_X(\lambda) = \lambda^{x_n} e^{-\lambda T} \prod_{i=1}^{n} \frac{(t_i - t_{i-1})^{(x_i - x_{i-1})}}{(x_i - x_{i-1})!}$

Note that $Y|\mathbb{X} = x$ is discrete .

Compute the derivative of h

$$h'_X(\lambda) = \left(x_n \lambda^{x_n - 1} e^{-\lambda T} - T \lambda^{x_n} e^{-\lambda T} \right) \prod_{i=1}^{n} \frac{(t_i - t_{i-1})^{(x_i - x_{i-1})}}{(x_i - x_{i-1})!}$$

Bayes classifier of Poisson process

$$h'_X(\lambda) = \Big(x_n - T\lambda\Big)\lambda^{x_n-1}e^{-\lambda T}\prod_{i=1}^{n}\frac{(t_i - t_{i-1})^{(x_i-x_{i-1})}}{(x_i - x_{i-1})!}$$

λ	0	$\frac{x_n}{T}$	$+\infty$
h_X	0	$h_X(\frac{x_n}{T})$	0

Bayes classifier of Poisson process

1. Y uniformly distributed in $\mathscr{Y}$

   If there exists an m s.t $x_n = \lambda_m T$ then

   $\arg\max_{\lambda_j \in \mathscr{Y}} \mathbb{P}(Y = \lambda_j | \mathbb{X} = x) = \lambda_m$ and $\lambda_m = \dfrac{x_n}{T}$

   Otherwise :

   - If $\forall \lambda_j \in \mathscr{Y}, x_n > \lambda_j T$ Then

     $\arg\max_{\lambda_j \in \mathscr{Y}} \mathbb{P}(Y = \lambda_j | \mathbb{X} = x) = \arg\max_{\lambda_j \in \mathscr{Y}} h_X(\lambda_j) = \max_{\lambda_j \in \mathscr{Y}} \lambda_j$

   - If $\forall \lambda_j \in \mathscr{Y}, x_n < \lambda_j T$ Then

     $\arg\max_{\lambda_j \in \mathscr{Y}} \mathbb{P}(Y = \lambda_j | \mathbb{X} = x) = \arg\max_{\lambda_j \in \mathscr{Y}} h_X(\lambda_j) = \min_{\lambda_j \in \mathscr{Y}} \lambda_j$

   - If $\forall \lambda_j \in \mathscr{Y} \leq \lambda_m, x_n > \lambda_j T$ and $\forall \lambda_j > \lambda_m, x_n < \lambda_j T$ Then

     $\arg\max_{\lambda_j \in \mathscr{Y}} \mathbb{P}(Y = \lambda_j | \mathbb{X} = x) = \arg \max_{\lambda_j \in \{\lambda_m, \lambda_{m+1}\}} (\lambda_j)^{x_n} e^{-\lambda_j T}$

2. Y not uniformly distributed in $\mathscr{Y}$

   $\arg\max_{\lambda_j \in \mathscr{Y}} \mathbb{P}(Y = \lambda_j | \mathbb{X} = x) = \arg\max_{\lambda_j \in \mathscr{Y}} \mathbb{P}(Y = \lambda_j) h_X(\lambda_j)$

   Which is λ_m such that $\forall j \neq m \qquad p_m h_X(\lambda_m) > p_j h_X(\lambda_j)$

   $$\iff \frac{h_X(\lambda_m)}{h_X(\lambda_j)} > \frac{p_j}{p_m} \iff \left(\frac{\lambda_m}{\lambda_j}\right)^{x_n} e^{(\lambda_j - \lambda_m)T} > \frac{p_j}{p_m} \square$$

Proof 2:

We consider $\mathbb{X} = \{X_{t_1}, \cdots, X_{t_{n-1}}, X_{t_n}, X_T\}, T_i$ an gamma distribution of the i^{th} jump. Hence $X_{T_i} = i$. Let $x = (1, \cdots, n-1, n, n)$ thus

$$\mathbb{P}(Y = \lambda_j | \mathbb{X} = x) = \mathbb{P}(X_{t_1} = 1, \cdots, X_{t_{n-1}} = n-1, X_{t_n} = n, X_T = n | Y = \lambda_j)\mathbb{P}(Y = \lambda_j)$$

$$= \mathbb{P}(t_1 \in [T_1; T_2[, \cdots t_n \in [T_n, T], N_T^{\lambda_j} = n)\mathbb{P}(Y = \lambda_j)$$

$$= \mathbb{P}\big((T_1, \cdots, T_n) \in A, N_T^{\lambda_j} = n | Y = \lambda_j\big)\mathbb{P}(Y = \lambda_j)$$

Where $A =]0; t_1] \times]t_1; t_2] \times \cdots \times]t_{n-1}; t_n]$

Since

$$\mathbb{P}\big((T_1, \cdots, T_n) \in A, N_T^{\lambda_j} = n | Y = \lambda_j\big) =$$

$$\int_A e^{-\lambda_j T} \lambda_j^n \mathbb{1}_{0<t_1<t_2<\cdots<t_n\leq T} dt_1 dt_2 \cdots dt_n$$

By (3) in proposition 1.

Bayes classifier of Poisson process

$$\mathbb{P}\big((T_1,\cdots,T_n)\in A, N_T^{\lambda_j}=n|\,Y=\lambda_j\big)=$$
$$e^{-\lambda_j T}\lambda_j^n\int_A \mathbb{1}_{0<t_1<t_2<\cdots<t_n\leq T}\,dt_1\,dt_2\cdots dt_n$$
$$\mathbb{P}(Y=\lambda_j|\mathbb{X}=x)=\mathbb{P}(Y=\lambda_j)e^{-\lambda_j T}\lambda_j^n$$
$$\int_A \mathbb{1}_{0<t_1<t_2<\cdots<t_n\leq T}\,dt_1\,dt_2\cdots dt_n$$

Note that $Y|((T_1,\cdots,T_n)\in A, N_T^{\lambda_j}=n)$ is discrete.
Since we have $0<t_1<t_2<\cdots<t_n\leq T$:

$$\int_A dt_1\,dt_2\cdots dt_n=\prod_{i=1}^{n}(t_i-t_{i-1})$$

$$\arg\max_{\lambda_j}\mathbb{P}(Y=\lambda_j|\mathbb{X}=x)=\arg\max_{\lambda_j}e^{-\lambda_j T}\lambda_j^n\mathbb{P}(Y=\lambda_j)\prod_{i=1}^{n}(t_i-t_{i-1})$$

with $t_0=0$ and we will have the same result with

$$h_X(\lambda_j)=e^{-\lambda_j T}\lambda_j^n\prod_{i=1}^{n}(t_i-t_{i-1})\ \square$$

Maximum Likelihood Estimation of the rate λ

Maximum Likelihood Estimation MLE

Theorem: The maximum likelihood estimation of the rate λ for a Poisson process $\{X_t, t \in [0; T]\}$ using the distribution of the inter-arrival times I_i or the arrival times T_i is $\widehat{\lambda} = \frac{N_T}{T}$ with N_T the value of the process at time T.

Proof 1: Using the distribution of the inter-arrival times I_i
Since $I_j \sim exp(\lambda)$ i.i.d and the density of $(I_1, \cdots, I_n, N_T = n)$ is $f_{(I_1,\cdots,I_n,N_T=n)}(i_1, \cdots, i_n, N_T = n) = \lambda^n e^{-\lambda T} \mathbb{1}_{(t_n \leq T)}$ by (2) in proposition 1. Hence his likelihood is
$L(\lambda | I_1, \cdots, I_n, N_T = n) = \lambda^n e^{-\lambda T} \mathbb{1}_{(t_n \leq T)}$

We assume that $t_n \leq T$, it's our case then
$L'(\lambda | I_1, \cdots, I_n, N_T = n) = \lambda^{n-1}(n - \lambda T)e^{-\lambda T}$
$L'(\lambda | I_1, \cdots, I_n, N_T = n) = 0 \iff \lambda = \frac{n}{T}$ hence
$\widehat{\lambda} = \arg\max_{\lambda} L(\lambda | I_1, \cdots, I_n, N_T = n) = \frac{n}{T}$

Maximum Likelihood Estimation of the rate λ

Proof 2: Using the distribution of the arrival times T_i

Since $T_i \sim Gamma(i,\lambda)$, $f_{T_i}(t_i) = \dfrac{\lambda e^{-\lambda t_i}(\lambda t_i)^{i-1}}{(i-1)!}$, $t_i \geq 0$, since the density of

$(T_1,\cdots,T_n,N_T=n)$ is

$f_{(T_1,\cdots,T_n,N_T=n)}(t_1,\cdots,t_n) = \lambda^n e^{-\lambda T}\mathbb{1}_{(0<t_1<\cdots<t_n\leq T)}$ by (3) then his likelihood is:

$L(\lambda|T_1,\cdots,T_n,N_T=n) = f_{(T_1,\cdots,T_n,N_T=n)}(t_1,\cdots,t_n) =$

$\lambda^n e^{-\lambda T}\mathbb{1}_{(0<t_1<\cdots<t_n\leq T)}$

If we have $0<t_1<\cdots<t_n\leq T$, our case:

$L'(\lambda|T_1,\cdots,T_n,N_T=n) = \lambda^{n-1}(n-\lambda T)e^{-\lambda T}$

$L'(\lambda|T_1,\cdots,T_n,N_T=n) = 0 \iff \lambda = \dfrac{n}{T}$

Since $N_T = n$, then $\widehat{\lambda} = \dfrac{N_T}{T}$ $\square$

Maximum Likelihood Estimation of the rate λ

Proposition 3: $\frac{N(t)}{t} \to \lambda$ almost surely when $t \to +\infty$.

Proof: We have $\mathbb{E}[\frac{N(t)}{t}] = \frac{\lambda t}{t} = \lambda$. $\square$

Maximum Likelihood Estimation for classification of λ

Let $\mathscr{Y} = \{\lambda_1, \cdots, \lambda_K\}$, instead computing MLE in $\mathbb{R}$ we compute it in $\mathscr{Y}$ to have the classifier which maximizes the likelihood. We have with the likelihood of arrival or inter-arrival times :

$\arg\max_{\lambda_j \in \mathscr{Y}} L(\lambda_j) = \arg\max_{\lambda_j \in \mathscr{Y}} h(\lambda_j)$ with $h(\lambda) = \lambda^n e^{-\lambda T}$ and $N_T = n$

If there exists an m s.t $n = \lambda_m T$ then $\arg\max_{\lambda_j \in \mathscr{Y}} L(\lambda_j) = \lambda_m$ and

$$\lambda_m = \frac{n}{T}$$

Maximum Likelihood Estimation of the rate λ

Otherwise :

- If $\forall \lambda_j \in \mathscr{Y}, n > \lambda_j T$ Then
  $\arg\max_{\lambda_j \in \mathscr{Y}} L(\lambda_j) = \arg\max_{\lambda_j \in \mathscr{Y}} h(\lambda_j) = \arg\max_{\lambda_j \in \mathscr{Y}} \lambda_j T = \max_{\lambda_j \in \mathscr{Y}} \lambda_j$
- If $\forall \lambda_j \in \mathscr{Y}, n < \lambda_j T$ Then
  $\arg\max_{\lambda_j \in \mathscr{Y}} L(\lambda_j) = \arg\max_{\lambda_j \in \mathscr{Y}} h(\lambda_j) = \arg\min_{\lambda_j \in \mathscr{Y}} \lambda_j T = \min_{\lambda_j \in \mathscr{Y}} \lambda_j$
- If $\forall \lambda_j \in \mathscr{Y} \leq \lambda_m, n > \lambda_j T$ and $\forall \lambda_j > \lambda_m, n < \lambda_j T$ Then
  $\arg\max_{\lambda_j \in \mathscr{Y}} L(\lambda_j) = \arg \max_{\lambda_j \in \{\lambda_m, \lambda_{m+1}\}} (\lambda_j)^n e^{-\lambda_j T}$

Maximum A Posteriori estimation of the rate λ

Maximum A Posteriori with the inter-arrival times I_n

- The posterior distribution is $\pi(\lambda|I_1,...,I_n,N_T=n)=\mathbb{P}(Y=\lambda|I_1=i_1,...,I_n=i_n,N_T=n)$. Let $\mathscr{Y}=\{\lambda_1,\cdots,\lambda_K\}$, the MAP classifier in $\mathscr{Y}$ is
$$\arg\max_{\lambda_j\in\mathscr{Y}}\pi(\lambda_j|I_1,...,I_n,N_T=n)=$$
$$\arg\max_{\lambda_j\in\mathscr{Y}}\frac{L(\lambda_j|I_1,\cdots,I_n,N_T=n)\mathbb{P}(Y=\lambda_j)}{\sum_j L(\lambda_j|I_1,\cdots,I_n,N_T=n)\mathbb{P}(Y=\lambda_j)}$$
$$\arg\max_{\lambda_j\in\mathscr{Y}}\pi(\lambda_j|I_1,...,I_n,N_T=n)=\arg\max_{\lambda_j\in\mathscr{Y}}L(\lambda_j|I_1,\cdots,I_n,N_T=n)\mathbb{P}(Y=\lambda_j)$$
- If λ is uniformly distributed it's the MLE classifier :
$$\arg\max_{\lambda_j\in\mathscr{Y}}\pi(\lambda_j|I_1,...,I_n,N_T=n)=\arg\max_{\lambda_j\in\mathscr{Y}}L(\lambda_j|I_1,...,I_n,N_T=n)$$

Maximum A Posteriori estimation of the rate λ

- Otherwise let $\mathbb{P}(Y = \lambda_j) = p_j$:

  $\arg\max_{\lambda_j \in \mathscr{Y}} \pi(\lambda_j | I_1, ..., I_n, N_T = n) = \arg\max_{\lambda_j \in \mathscr{Y}} p_j(\lambda_j)^n e^{-\lambda_j T}$

  Which is λ_m such that $\forall j \neq m$ $\left(\frac{\lambda_m}{\lambda_j}\right)^n e^{(\lambda_j - \lambda_m) t_n} > \frac{p_j}{p_m}$

Maximum A Posteriori with the arrival times T_n

- The posterior distribution is
  $\pi(\lambda|T_1,..,T_n,N_T=n)=\mathbb{P}(Y=\lambda|T_1=t_1,..,T_n=t_n,N_T=n)$.
  Let $\mathscr{Y}=\{\lambda_1,\cdots,\lambda_K\}$, the MAP classifier in $\mathscr{Y}$ is
  $$\arg\max_{\lambda_j\in\mathscr{Y}}\pi(\lambda_j|T_1,...,T_n,N_T=n)=$$
  $$\arg\max_{\lambda_j\in\mathscr{Y}}\frac{L(\lambda_j|T_1,...,T_n,N_T=n)\mathbb{P}(Y=\lambda_j)}{\sum_j L(\lambda_j|T_1,...,T_n,N_T=n)\mathbb{P}(Y=\lambda_j)}$$
  $$\arg\max_{\lambda_j\in\mathscr{Y}}\pi(\lambda_j|T_1,...,T_n,N_T=n)=$$
  $$\arg\max_{\lambda_j\in\mathscr{Y}}L(\lambda_j|T_1,...,T_n,N_T=n)\mathbb{P}(Y=\lambda_j)$$
- If λ is uniformly distributed it's the MLE classifier :
  $$\arg\max_{\lambda_j\in\mathscr{Y}}\pi(\lambda_j|T_1,...,T_n,N_T=n)=$$
  $$\arg\max_{\lambda_j\in\mathscr{Y}}L(\lambda_j|T_1,...,T_n,N_T=n)$$

Maximum A Posteriori with the arrival times T_n

- Otherwise let $\mathbb{P}(Y = \lambda_j) = p_j$

  $\arg\max_{\lambda_j \in \mathscr{Y}} \pi(\lambda_j | T_1, ..., T_n, N_T = n) =$

  $\arg\max_{\lambda_j \in \mathscr{Y}} p_j \lambda_j^n e^{-\lambda_j T} \mathbb{1}_{(0<t_1<\cdots<t_n\leq T)}$

  if $0 < t_1 < \cdots < t_n$, our case:

  $\arg\max_{\lambda_j \in \mathscr{Y}} \pi(\lambda_j | T_1, ..., T_n, N_T = n) = \lambda_m$ such that

  $$\forall j \neq m \qquad \left(\frac{\lambda_m}{\lambda_j}\right)^n e^{(\lambda_j - \lambda_m)t_n} > \frac{p_j}{p_m}$$

Machine Learning classifier: Random forest, Gaussian naive Bayes

Data: For a Poisson process $\{X_t, t \in [0, T]\}$, we discrete the process in n times as $\{X_t, t \in \{t_1, t_2, \cdots, t_n\}\}$, $X_{t_i}, i \in [1, n]$. The times $t_1, t_2, \cdots, t_n$ of the process will be our features and the parameter λ our label to predict for the machine learning models. And we simulate several process to make our data base.
Application: We will train this data set with models like random forest classifier and gaussian naive Bayes classifier.

Applications on Poisson point processes
Bayes classifier of homogeneous Poisson point process

Proposition 4.1: Bayes classifier of homogeneous Poisson point process
Consider $X = \{\xi_1, \cdots, \xi_n\}$, $\xi_i \sim Uniform(S)$, $S \subset \mathbb{R}^d$ a bounded Borel measurable set, which form a homogeneous Poisson point process $N(S) = \#X \sim Poisson(\lambda|S|)$, the number of points of n existing in S and $|\cdot|$ a non-negative measure.

Bayes classifier of homogeneous Poisson point process

Case 1: $\mathscr{Y}$ discrete and ordered set, λ uniformly distributed

The Bayes classifier of the rate λ in the discrete set $\mathscr{Y}$ is $\widehat{\lambda} = \dfrac{n}{|S|}$,

if there exist $\lambda_j \in \mathscr{Y}$ such that $\lambda_j = \dfrac{n}{|S|}$. We have

$\arg\max_{\lambda_j} \mathbb{P}(Y = \lambda_j | N(S) = n) = \widehat{\lambda}$.

Otherwise :

- If $\forall \lambda_j, n > \lambda_j |S|$ Then
  $\arg\max_{\lambda_j} \mathbb{P}(Y = \lambda_j | N(S) = n) = \arg\max_{\lambda_j} \lambda_j |S| = \max_j \lambda_j$
- If $\forall \lambda_j, n < \lambda_j |S|$ Then
  $\arg\max_{\lambda_j} \mathbb{P}(Y = \lambda_j | N(S) = n) = \arg\min_{\lambda_j} \lambda_j |S| = \min_j \lambda_j$
- If $\forall j \leq m \ n > \lambda_j |S|$ otherwise , $n < \lambda_j |S|$ Then
  $$\arg\max_{\lambda_j} \mathbb{P}(Y = \lambda_j | N(S) = n) = \arg \max_{\lambda_j \in \{\lambda_m, \lambda_{m+1}\}} \frac{(\lambda |S|)^n}{n!} e^{-\lambda |S|}$$

Bayes classifier of homogeneous Poisson point process

Case 2: $\mathscr{Y}$ discrete and ordered set, Y no uniformly distributed

Let $\mathbb{P}(Y = \lambda_j) = p_j$, the Bayes classifier of the rate λ is

$\arg\max_{\lambda_j \in \mathscr{Y}} \mathbb{P}(Y = \lambda_j | N(S) = n)$

Which is λ_m such that $\forall j \neq m$ $\quad \left(\frac{\lambda_m}{\lambda_j}\right)^n e^{(\lambda_j - \lambda_m)|S|} > \frac{p_j}{p_m}$

Proof : See **5.1.1**

Bayes classifier of homogeneous Poisson point process

Proposition 4.2: Bayes classifier of homogeneous Poisson point process

Consider $X_k = \{\xi_1, \cdots, \xi_{n(k)}\}$, $k \in [\![1; K]\!]$, $\xi_i \sim Uniform(S_k)$, $S_1, ..., S_K \subset \mathbb{R}^d$ a collection of disjoint bounded Borel sets which form a homogeneous Poisson point processes

$N(S_k) = \#X_k \sim Poisson(\lambda |S_k|)$. Consider $n = \sum_{k=1}^{K} n(k)$ and

$|S| = \sum_{k=1}^{K} |S_k|$.

Case 1: $\mathscr{Y}$ discrete and ordered set, Y uniformly distributed

The Bayes classifier of the rate λ in the discrete set $\mathscr{Y}$ is

$\widehat{\lambda} = \dfrac{\sum_{k=1}^{K} n(k)}{\sum_{k=1}^{K} |S_k|} = \dfrac{n}{|S|}$, if there exist $\lambda_j \in \mathscr{Y}$ such that $\lambda_j = \dfrac{n}{|S|}$.

We have $\arg\max_{\lambda_j} \mathbb{P}(Y = \lambda_j | N(S_1) = n(1), ..., N(S_K) = n(K)) = \widehat{\lambda}$.

Bayes classifier of homogeneous Poisson point process

Otherwise :

- If $\forall j, n > \lambda_j|S|$ Then
  $\arg\max_{\lambda_j} \mathbb{P}(Y = \lambda_j | N(S_1) = n(1), ..., N(S_K) = n(K)) =$
  $\arg\max_{\lambda_j} \lambda_j|S| = \max_j \lambda_j$
- If $\forall \lambda_j, n < \lambda_j|S|$ Then
  $\arg\max_{\lambda_j} \mathbb{P}(Y = \lambda_j | N(S_1) = n(1), ..., N(S_K) = n(K)) =$
  $\arg\min_{\lambda_j} \lambda_j|S| = \min_j \lambda_j$
- If $\forall j \leq m, n > \lambda_j|S|$ and otherwise, $n < \lambda_j|S|$ Then
  $\arg\max_{\lambda_j} \mathbb{P}(Y = \lambda_j | N(S_1) = n(1), ..., N(S_K) = n(K)) =$
  $\arg \max_{\lambda_j \in \{\lambda_m, \lambda_{m+1}\}} (\lambda_j)^n e^{-\lambda_j|S|}$

Bayes classifier of homogeneous Poisson point process

Case 2: $\mathscr{Y}$ discrete and ordered set, Y no uniformly distributed

Assume that $\mathscr{Y} = \{\lambda_1, \cdots, \lambda_J\}$ and $\mathbb{P}(Y = \lambda_j) = p_j$.

The Bayes classifier of the rate λ in the discrete set $\mathscr{Y}$:

$$\arg\max_{\lambda_j} \mathbb{P}(Y = \lambda_j | N(S_1) = n(1), ..., N(S_K) = n(K)) =$$

$$\arg\max_{\lambda_j} p_j \lambda_j^n e^{-\lambda_j |S|} \prod_{k=1}^{K} \frac{|S_k|}{n(k)!}$$

Which is λ_m such that $\forall j \neq m$ $\quad \left(\frac{\lambda_m}{\lambda_j}\right)^n e^{(\lambda_j - \lambda_m)|S|} > \frac{p_j}{p_m}$

Proof :

$$\mathbb{P}(Y=\lambda_j|N(S_i)=n(i),i=1,\ldots,K)=$$
$$\mathbb{P}(N(S_i)=n(i),i=1,\ldots,K|Y=\lambda_j)\mathbb{P}(Y=\lambda_j)$$

Case 1: Since Y is uniformly distributed then

$$\mathbb{P}(Y=\lambda_j|N(S_i)=n(i),i=1,\ldots,K)=$$
$$\mathbb{P}(N(S_i)=n(i),i=1,\ldots,K|Y=\lambda_j)$$
$$\arg\max_{\lambda_j}\mathbb{P}(Y=\lambda_j|N(S_1)=n(1),...,N(S_K)=n(K))=$$
$$\arg\max_{\lambda_j}\prod_{i=1}^{K}\frac{(\lambda|S_i|)^{n_i}}{n_i!}e^{-\lambda|S_i|}$$
$$=\arg\max_{\lambda_j}\lambda_j^n e^{-\lambda_j|S|}\prod_{k=1}^{K}\frac{|S_k|}{n(k)!}=\arg\max_{\lambda_j}(\lambda_j)^n e^{-\lambda_j|S|}$$

Case 2: Since Y is not uniformly distributed then

$$\arg\max_{\lambda_j}\mathbb{P}(Y=\lambda_j|N(S_1)=n(1),...,N(S_K)=n(K))=$$
$$\arg\max_{\lambda_j}p_j\prod_{i=1}^{K}\frac{(\lambda|S_i|)^{n_i}}{n_i!}e^{-\lambda|S_i|}$$
$$=\arg\max_{\lambda_j}p_j(\lambda_j)^n e^{-\lambda_j|S|}$$

□

Bayes classifier of inhomogeneous Poisson point process

Proposition 5.1: Bayes classifier of homogeneous Poisson point process

Consider $X = \{\xi_1, \cdots, \xi_n\}$, $\xi_i \sim \frac{\lambda(x_i)}{\Lambda(S)}$, $S \subset \mathbb{R}^d$ a bounded Borel measurable set, which form a homogeneous Poisson point process $N(S) = \#X \sim Poisson(\Lambda(S))$, the number of points of n existing in S.

Let $\mathscr{F} = \{\lambda_1, ..., \lambda_J\}$ be a discrete set of measurable function in $\mathbb{R}^d$, uniformly distributed. The Bayes classifier of the rate λ in the discrete set $\mathscr{F}$ is

$$\widehat{\lambda} = \arg\max_{\lambda \in \mathscr{F}} \mathbb{P}(F = \lambda | N(S) = n) = \arg\max_{\lambda \in \mathscr{F}} \frac{(\Lambda(S))^n}{n!} e^{-\Lambda(S)}.$$

Bayes classifier of inhomogeneous Poisson point process

Proposition 5.2: Bayes classifier of inhomogeneous Poisson point process

Consider $X_k = \{\xi_1, \cdots, \xi_{n(k)}\}$, $k \in [\![1; K]\!]$, $\xi_i \sim \frac{\lambda(x_i^k)}{\Lambda(S_k)}$, $S_1, ..., S_K \subset \mathbb{R}^d$ a collection of disjoint bounded Borel sets which form a inhomogeneous Poisson point processes

$N(S_k) = \#X_k \sim Poisson(\Lambda(S_k))$. Consider $n = \sum_{k=1}^{K} n(k)$ and

$|S| = \sum_{k=1}^{K} |S_k|$.

.

Bayes classifier of inhomogeneous Poisson point process

Case 1: (F uniformly distributed) Let $\mathscr{F} = \{\lambda_1, ..., \lambda_J\}$ be a discrete set of measurable function in $\mathbb{R}^d$, uniformly distributed. The Bayes classifier of the rate λ in the discrete set $\mathscr{F}$ is

$$\widehat{\lambda} = \arg\max_{\lambda \in \mathscr{F}} \mathbb{P}(F = \lambda | N(S_1) = n(1), ..., N(S_K) = n(K)) =$$

$$\arg\max_{\lambda \in \mathscr{F}} \prod_{k=1}^{K} \frac{(\Lambda(S_k))^{n(k)}}{n(k)!} e^{-\Lambda(S_k)}$$

Case 2: F no uniformly distributed

Assume that $\mathbb{P}(F = \lambda_j) = p_j$ and $\mathscr{F}$ no uniformly distributed.

The Bayes classifier of the rate λ in the discrete set $\mathscr{F}$ is

$$\widehat{\lambda} = \arg\max_{\lambda_j} \mathbb{P}(F = \lambda_j | N(S_1) = n(1), ..., N(S_K) = n(K)) =$$

$$\arg\max_{\lambda_j} p_j \prod_{k=1}^{K} \frac{(\Lambda_j(S_k))^{n(k)}}{n(k)!} e^{-\Lambda_j(S_k)} \ .$$

Which is λ_m such that

$$\forall j \neq m \qquad \prod_{k=1}^{K} \left(\frac{\Lambda_m(S_k)}{\Lambda_j(S_k)} \right)^{n(k)} \exp[(\Lambda_j(S_k) - \Lambda_m(S_k)] > \frac{p_j}{p_m}$$

Bayes classifier of inhomogeneous Poisson process

Proposition 5.3 (Lemma) Bayes classifier of an inhomogeneous Poisson process in a separable and bounded metric space
Let $(X, Y) \in S^{\infty} \times \{0, \ldots, M\}$. Let $X|Y = j$ be Poisson processes on S^{∞} with intensities λ_j, $j = 1, \ldots, M$, respectively. Then the Bayes rule classifies a point $x \in S^{\infty}$ into class j if

$$\forall_{i \neq j} \quad \exp\{\mu_i(S) - \mu_j(S)\} \prod_{\xi \in X} \frac{\lambda_j(\xi)}{\lambda_i(\xi)} > \frac{p_i}{p_j}$$

where $p_i = \Pr(Y = i)$, $i = 1, \ldots, M$ and as before,
$\mu_i(S) = \int_S \lambda_i(\xi) d\nu(\xi)$, $i = 1, \ldots, M$.[2]

Results
Analysis of probabilistic models results

We choose λ in $\{1,2,3,4,5,6,7,8,9,10\}$ and N processes for each one .

Table 1: Precisions of Bayes classifier 'discrete' (mean) by λ and T.

T ↓ / λ →	1	2	5	6	8	9	10	Total
1	50.33%	27%	14%	14%	16.67%	22%	27.67%	21.9%
5	75%	47%	31%	27%	27.33	30.33	41.33	37.93
10	81%	57%	38.33%	40%	30.33%	36.33%	53.67%	45.63%
15	87.67%	70.67%	47.33%	46.33%	44.33%	43.67%	52.67%	53.97%
20	89.33%	77.67%	54.33%	49.33%	47.33%	41%	63%	60.83%
25	94%	83%	62%	53.33%	51.67%	54.67%	62.67%	65.43%
30	95%	86%	62%	63%	48.67%	55.67%	69.67%	68.5%
50	99.33%	95.33%	76.33%	73%	67%	73%	74.67%	79.83%
100	100%	99.67%	93.67%	90.67%	86.67%	86%	86.33%	92.47%
200	100%	100%	97.67%	97%	95.67%	92.67%	94%	97.37%

Analysis of probabilistic models results

Table 2: Precisions of Bayes classifier 'discrete' (most) by λ and T.

T ↓ / λ →	1	2	5	6	8	9	10	Total
1	85.67%	4.33%	11.67%	11.67%	0.67%	4.67%	67.67%	22.63%
5	88%	45.33%	34.33%	31.33%	22%	21.33%	68%	42.27%
10	88.67%	60.33%	41%	45%	35%	27.67%	73%	51.37%
15	92%	77%	54%	54.33%	50.67%	39.67%	73%	61.4%
20	92.67%	80.33%	61%	57%	47%	45%	76.67%	66%
25	96.33%	85%	67.67%	58%	58.67%	52.67%	78%	70.47%
30	97%	87%	68.33%	68%	54.67%	55%	80.33%	73.27%
50	99.33%	96.33%	81%	77.33%	72%	72.67%	82.33%	83.13%
100	100%	100%	94.33%	92.33%	88.33%	85.33%	91%	93.6%
200	100%	100%	98.33%	97.67%	96.67%	92.67%	94.33%	97.77%

Analysis of probabilistic models results

Table 3: Precisions of Bayes classifier 'inter-arrival times' by λ and T.

T ↓ / λ →	1	2	5	6	8	9	10	Total
1	42.67%	16.67%	17.67%	16.33%	14.67%	15%	66.33%	23.47%
5	66.33%	52.67%	34.33%	31%	27.33%	31.67%	60%	42.53%
10	79.33%	68.33%	50%	50%	39.67%	39.33%	71.33%	54.87%
15	90.67%	81.33%	61.33%	56.67%	51%	47%	73.33%	65.73%
20	96%	82.67%	66%	62.67%	52.67%	54.67%	78%	68.97%
25	96.67%	91.67%	74%	71%	62.33%	59%	81.33%	77.13%
30	98.33%	95.33%	80%	72.67%	64.33%	60.33%	82.67%	79.3%
45	100%	98%	86.33%	82.67%	72%	72.33%	90%	86.67%
50	100%	97.33%	91%	83.67%	79%	75%	93%	88.37%
55	99.67%	98.67%	92.33%	88.67%	78.67%	38.33%	98.67%	86.9%
100	100%	100%	97.33%	96.33%	92%	89%	95.33%	96.53%
200	100%	100%	100%	99.67%	99.67%	99%	99.33%	99.6%

Analysis of probabilistic models results

Table 4: Precisions of Bayes classifier 'inter times' by λ and T.

T ↓ / λ →	1	2	5	6	8	9	10	Total
1	76.67	00%	19.33%	19.67%	13%	12.33%	72%	26.5%
5	85%	43.33%	36%	27%	28.67%	31.67%	62.67%	43.4%
10	91.33%	62.33%	49.33%	45.67%	38.33%	36.33%	73.33%	54.1%
15	94%	76.67%	58%	55.33%	49%	46.33%	71.67%	64.5%
20	97.67%	89.67%	33.33%	56.67%	48.33%	38.33%	71%	55.1%
25	98%	86.33%	72.67%	68%	58%	59%	80.33%	74.77%
30	99.33%	90%	75%	70.33%	62%	59.33%	83%	77.57%
45	100%	98.67%	86.67%	81.33%	73.67%	73%	89.33%	86.93%
50	100%	97.67%	89.33%	83.67%	79%	81%	91%	88.63%
55	100%	98.67%	92.67%	46.33%	45.33%	49.67%	98.67%	76.83%
100	100%	100%	97.33%	96.67%	92%	88.33%	94%	96.43%
200	100%	100%	100%	99.67%	99.67%	99%	99.33%	99.6%

Confusion matrix

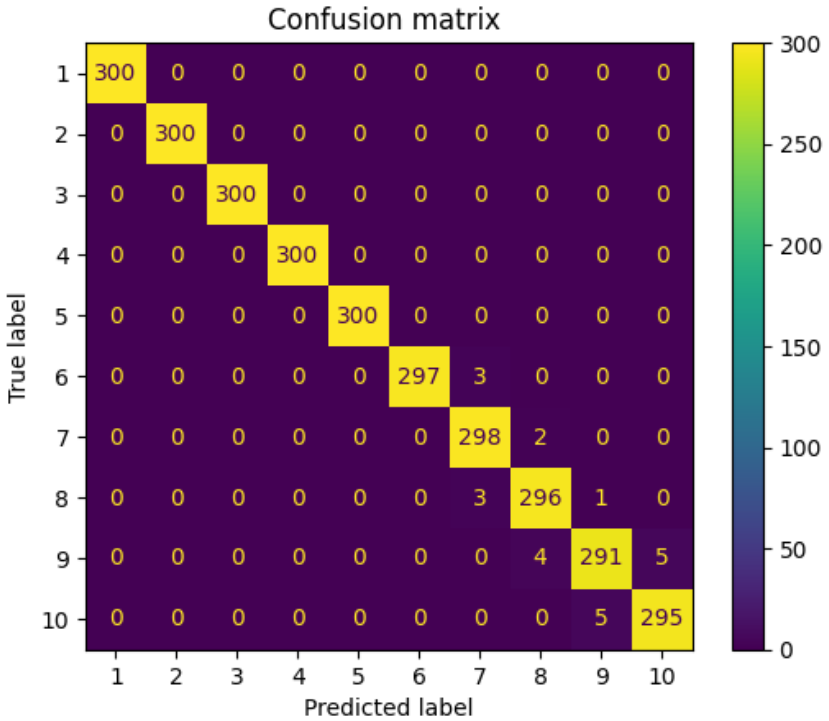


Figure 2: inter times with T=200

Analysis of probabilistic models results

Table 5: Precisions of MLE classifier (round()) by λ and T.

T ↓ / λ →	1	2	5	6	8	9	10	Total
1	72.67%	26.33%	15.33%	14.33%	12.67%	14.33%	56.67%	27.23%
5	89%	57.67%	40%	33.33%	30.33%	30.67%	60.67%	46.67%
10	90.67%	78.67%	40%	54.33%	49.33%	35.66%	69.67%	57.57%
15	95.67%	83.67%	62%	59.67%	48%	48%	49.67%	70.67%
20	98.67%	88.33%	60.33%	64.33%	61%	48.33%	73.33	70.03
25	98.67%	92%	76.67%	64.67%	64%	63.33%	77.33%	76.27%
30	99.67%	95.33%	77.67%	76.67%	67%	61.67%	84%	79.73%
50	100%	99.67%	89.67%	87%	81%	76.67%	86.67%	88.6%
55	100%	98%	86.67%	89.33%	83.67%	45.67%	99.67%	87.13%
100	100%	100%	97.67%	95.67%	92.67%	91%	92.67%	96.53%
200	100%	100%	99.67%	99.67%	99.33%	97.67%	97.67%	99.37%

Analysis of probabilistic models results

Table 6: Precisions of MLE (round()) by λ and T.

T ↓ / λ →	1	2	5	6	8	9	10	Total
1	30.33%	26.33%	15.33%	14.33%	12.67%	14.33%	12.67%	18.6%
5	77.33%	57.67%	40%	33.33%	30.33%	30.67%	24.33%	41.87%
10	82.67%	78.67%	40%	54.33%	49.33%	35.66%	39.67%	53.77%
15	93.33%	83.67%	62%	59.67%	48%	48%	47.33%	58.57 %
20	98.67%	88.33%	60.33%	64.33%	61%	48.33	51.33%	67.7%
25	98%	92%	76.67%	64.67%	64%	63.33%	55.67%	74.03%
30	99.67%	95.33%	77.67%	76.67%	67%	61.67%	61.67%	77.5%
50	100%	99.67%	89.67%	87%	81%	76.67%	75.33%	87.47%
100	100%	100%	97.67%	95.67%	92.67%	91%	90.67%	96.33%
200	100%	100%	99.67%	99.67%	99.33%	97.67%	96.33%	99.23%

Analysis of ML models results

For the random forest classifier we have 47%,100%,100%,100%,... accuracy for training data.

Table 7: Precisions of Random forest classifier in test data by λ and T.

T ↓ / λ →	1	2	5	6	8	9	10	Total
1	46.67%	18.89%	13.33%	14.44%	22.22%	23.33%	42.22%	22.55%
5	86.56%	52.22%	33.33%	26.67%	32.22%	31.11%	53.33%	41.78%
10	90%	71.11%	34.44%	37.78%	42.22%	50%	64.44%	54.67%
15	95.56%	90%	50%	48.89%	56.67%	52.22%	57.78%	63.11%
20	94.44%	86.67%	68.89%	68.89%	41.11%	48.89%	66.67%	68.67
25	96.67%	88.89%	71.11%	66.67%	62.22%	60%	77.78%	74.11%
30	97.78%	95.56%	70%	76.67%	61.11%	57.78%	82.22%	77.89%
50	100%	95.56%	95.56%	78.89%	77.78%	78.89%	84.44%	87.67%
100	100%	100%	94.44%	97.78%	91.11%	88.89%	95.56%	96.33%
200	100%	100%	100%	100%	98.89%	98.89%	97.78%	99.44%

Confusion matrix

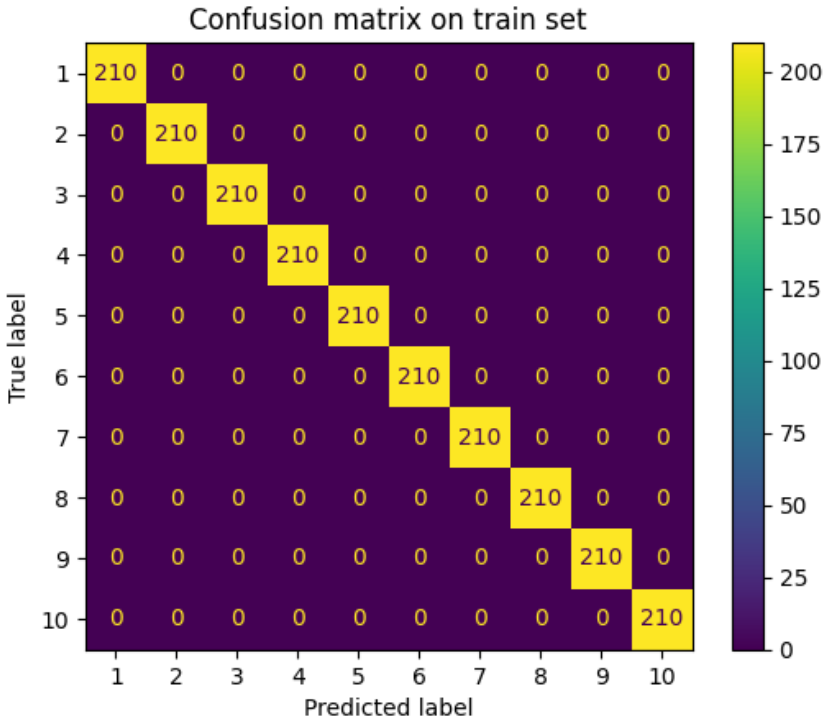


Figure 3: random forest on train set with T=200

Confusion matrix

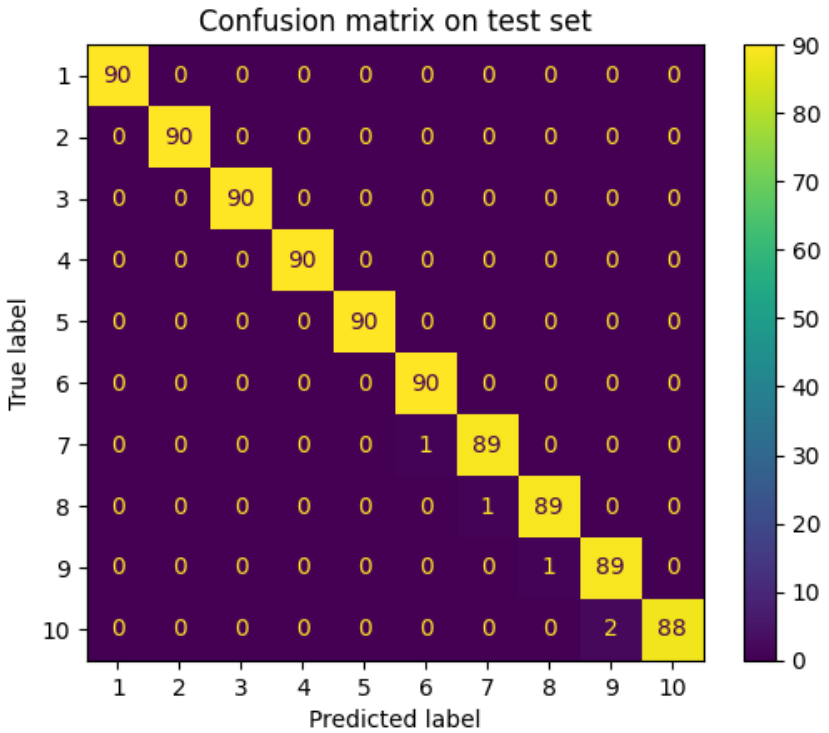


Figure 4: random forest on test set with T=200

Analysis of ML models results

For the gaussian naive Bayes classifier we have 26%,41.86%,52.19%,60.19%,66.05%,70.90%,74.9%,84.57%,94.47% and 98.14% accuracy for training data.

Table 8: Precisions of gaussian naive Bayes classifier in test data by λ and T.

T ↓ / λ →	1	2	5	6	8	9	10	Total
1	75.55%	2%	11.11%	2%	7.77%	18.89%	54.44%	25%
5	85.56%	41.11%	41.11%	27.78%	27.78%	23.33%	53.33%	40%
10	90%	65.56%	40%	45.55%	43.33%	31.33%	68.89%	53%
15	93.33%	75.56%	55.56%	46.67%	52.22%	36.67%	58.89%	60%
20	91.11%	86.67%	63.33%	60%	45.56%	46.67%	75.56%	68%
25	94.44%	93.33%	72.22%	64.44%	56.67%	55.56%	76.67%	72%
30	96.67%	90%	62.22%	64.44%	60%	57.78%	77.78%	74%
50	100%	92.22%	84.44%	71.11%	73.33%	75.56%	77.76%	84.57%
100	100%	100%	96.67%	94.44%	91.11%	86.67%	95.56%	98.14%
200	100%	100%	98.89%	98.89%	95.56%	96.67%	93.33%	98.14%

Confusion matrix

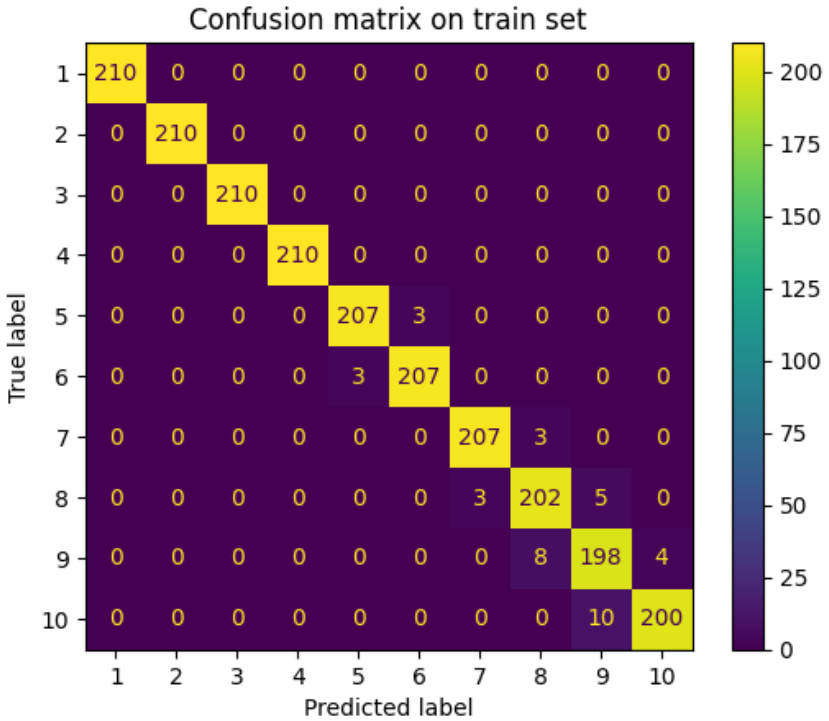


Figure 5: Gaussian naive Bayes on train set with T=200

Confusion matrix

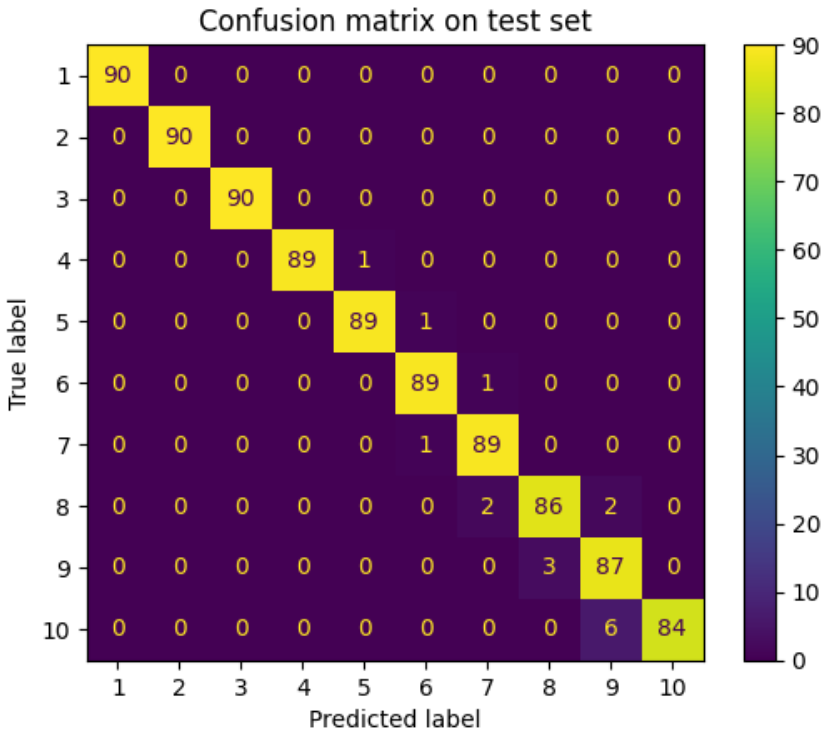


Figure 6: Gaussian naive Bayes on test set with T=200

Discussion

The results show us that the models have the same evolution following T. The precisions increase with T. If T is less than 10 the precisions are under 50% , at $T = 50$ we reach mostly 85% and at $T = 100$ we reach mostly 96%. The Bayes classifier for Poisson process and the MLE classifier reach 99% at $T = 200$. And we know that the MLE converge to the true λ when T tend to infinity.

Inhomogeneous Poisson process in a separable and bounded metric space

Application: We define three gaussian intensities on the space S:=[0, 1] x [0, 1].

Higher intensity around (0.25, 0.25) :

$$\lambda_1(x,y) = 10e^{-\frac{(x-0.25)^2+(y-0.25)^2}{2\times 0.05^2}} + 0.5$$

Higher intensity around (0.75, 0.75):

$$\lambda_2(x,y) = 8e^{-\frac{(x-0.75)^2+(y-0.75)^2}{2\times 0.08^2}} + 0.2$$

Higher intensity around (0.25, 0.75) and (0.75, 0.25):

$$\lambda_3(x,y) = 7e^{-\frac{(x-0.25)^2+(y-0.25)^2}{2\times 0.06^2}} + 7e^{-\frac{(x-0.75)^2+(y-0.75)^2}{2\times 0.06^2}} + 1$$

And we compare the results of classifications Bayes classifier using the intensities to the KNN classifier.

Results

A realization with the intensities λ_1, λ_2 and λ_3 .

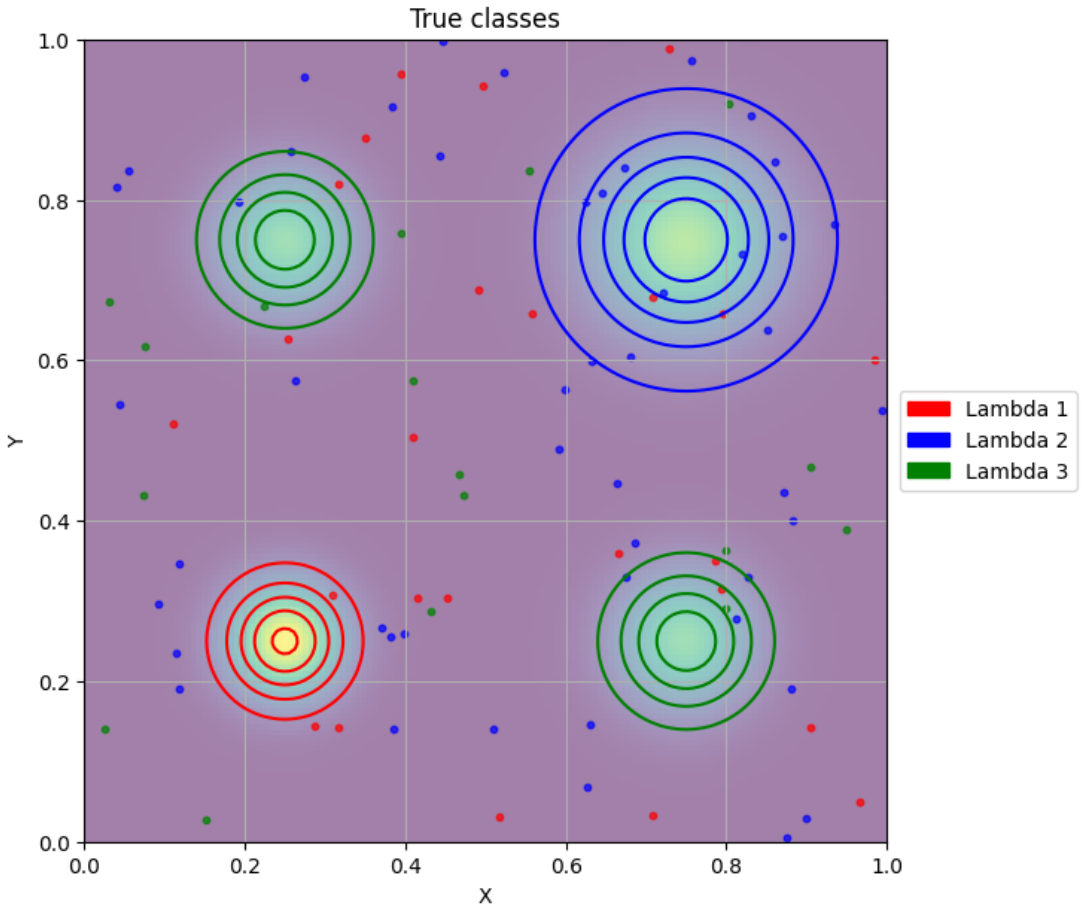


Figure 7: Realization with λ_1, λ_2 and λ_3

Results

We predict classes with the Bayes classifier define in **Proposition 5.3** :

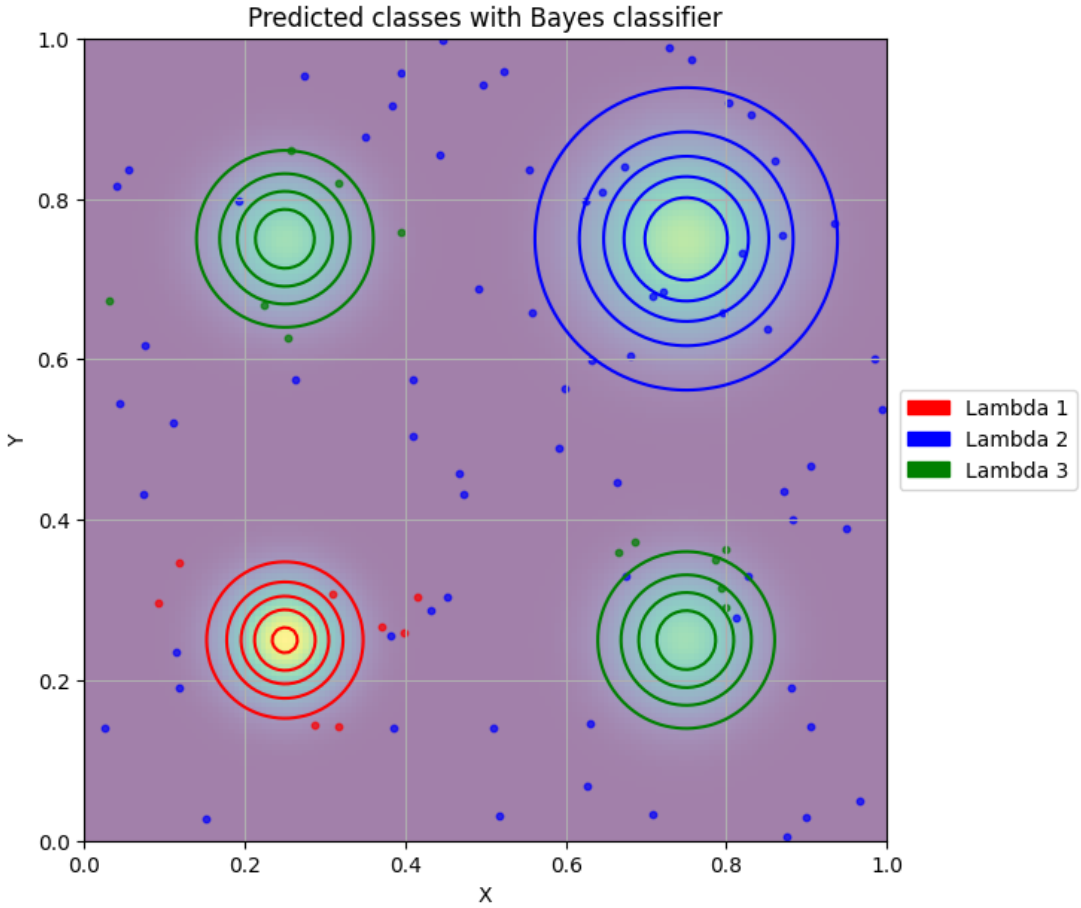


Figure 8: Predicted classes with Bayes classifier

Results

Confusion matrix of the Bayes classifier predictions with 46%

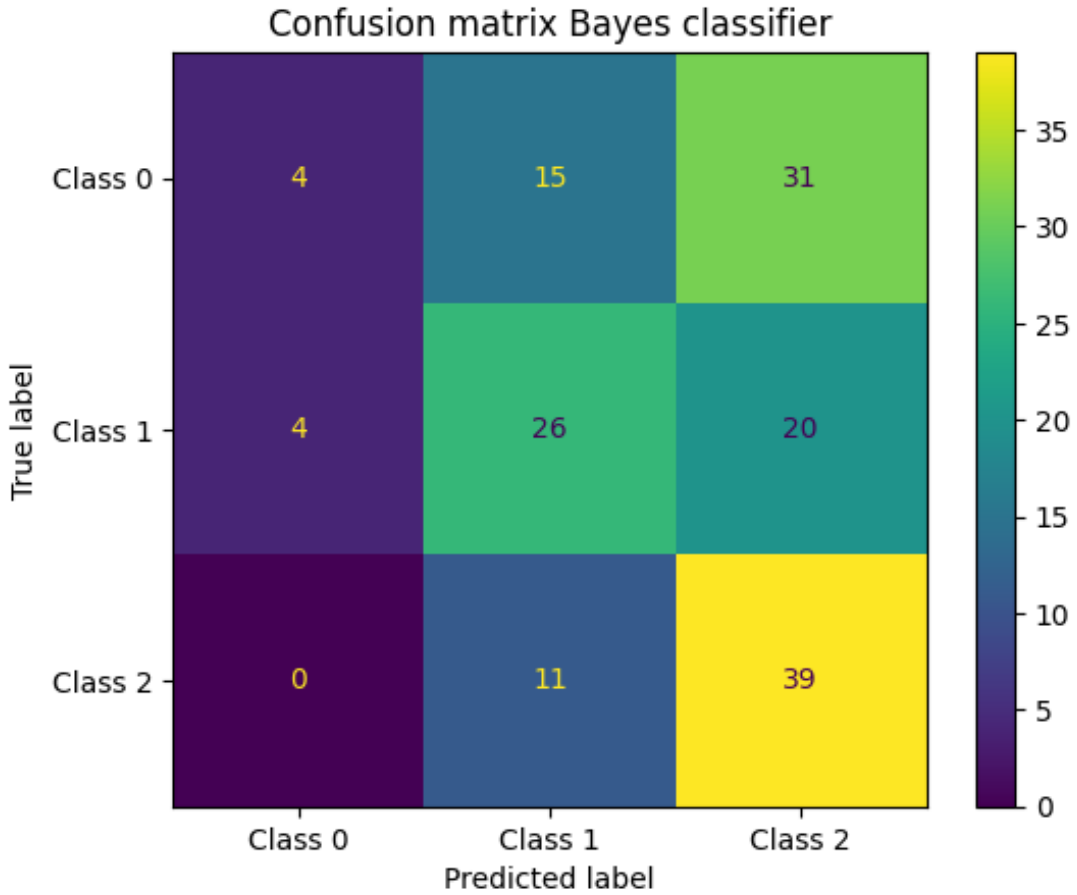


Figure 9: Confusion matrix Bayes classifier

Results

We predict classes with a KNN classifier :

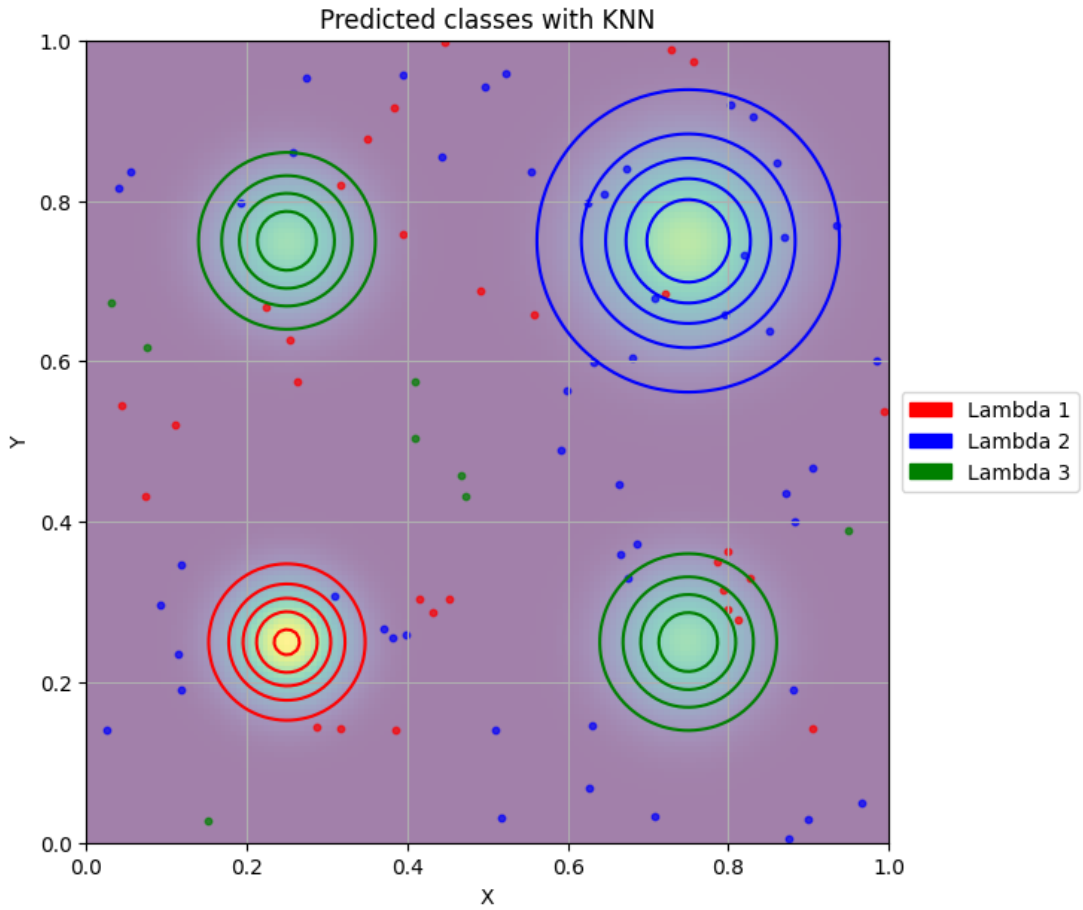


Figure 10: Predicted classes with KNN classifier

Results

Confusion matrix of the KNN classifier predictions with 65.55% accuracy:

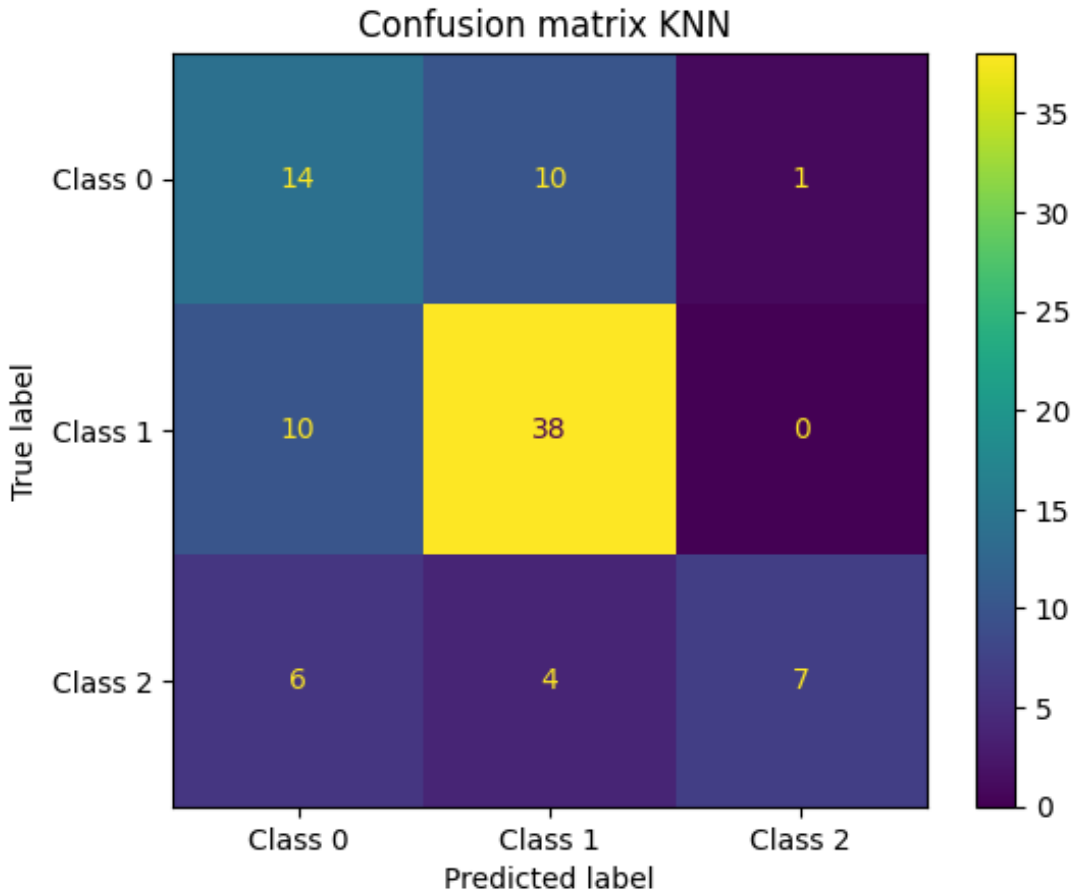


Figure 11: Confusion matrix KNN classifier

Discussion

For the spatial point Poisson classification the KNN classifier has better accuracy but it doesn't give us a density function. And Bayes classifier can have a good accuracy and give a density function which permit to do simulation.

Conclusion

We conclude that for Poisson processes the Bayes classifier is better, and it's the right solution. However the others classifiers are good too,we have almost the same score especially with the MLE classifier. We know that the MLE converge to λ when T tend to infinity.
And for spatial point Poisson processes, if we want to just classify the ML classifier is better. But to have better simulation we have to use the Bayes classifier of densities.

Thank you for your attention!