Introduction to Poisson processes with R

Objectives

The aim of this session is to manipulate and illustrate the notions introduced in the lecture on Poisson processes with the R software.

Load the `R` file **TP-PoissonProcess-etud.R** (to be completed), available on the Moodle page, and fill in the gaps during this session.

1 Homogeneous Poisson processes observed on a fixed window

First, we consider the case of a fixed observation window (and thus a random number of points).

1.1 Simulation

i. Recall the conditional distribution of the arrival times of a homogeneous Poisson process with rate `lambda` on an interval `[0,Tmax]`, given the number of points in that interval.

ii. Write an `R` function `simulPPh1` with arguments `lambda` and `Tmax`, which simulates such a process and returns the corresponding arrival times.

iii. For `lambda=2` and `Tmax=10`, simulate a homogeneous Poisson process and plot both the counting process and the arrival times.
Indication : you may use the `R` functions `plot()` (with the option `type="s"`), `points()` and `lines()`.

1.2 Maximum likelihood estimator

i. Write an `R` function `MLE1` which returns the Maximum Likelihood Estimator of a homogeneous Poisson process `PPh` observed on a fixed window [0,`Tmax`].

ii. Apply it on different simulated data. What do you observe ?

1.3 Asymptotic behavior of the MLE

In this section, we illustrate the asymptotic behavior [1], as T goes to $+\infty$, of the MLE $\hat{\lambda}_T$ of the rate λ of a homogeneous Poisson process on $[0, T]$.

1.3.1 Strong LLN-type result

First let us illustrate the almost-sure convergence :

$$\hat{\lambda}_T \xrightarrow[T\to+\infty]{a.s.} \lambda. \tag{1}$$

Note that in the lecture, we considered a weak LLN-type result since we only proved the convergence in probability. To illustrate (1) :

i. Fix a rate `lambda=2` and a sequence of `Tmax` that tends to $+\infty$, say `Tillustr=1:500`.

ii. For each `Tmax` in `Tillustr`, simulate a homogeneous Poisson process with rate `lambda` on [0,`Tmax`], and compute the MLE.

iii. Plot the obtained values for the MLE as a function of `Tmax`.

What do you observe ?

1. LLN refers to "Law of Large Numbers" and CLT stands for "Central Limit Theorem".

1.3.2 CLT-type result

We now illustrate the following result :

$$\sqrt{T}(\hat{\lambda}_T - \lambda) \xrightarrow[T\to+\infty]{\mathcal{L}} \mathcal{N}(0,\lambda). \tag{2}$$

To do so, fix `lambda=2` and do the following for different values of `Tmax`.

i. Fix the number of simulations `K=1000` and create a vector of size `K` : `Z=rep(0,K)`.

ii. Store in `Z` a `K`-sample with same distribution as $\sqrt{T}(\hat{\lambda}_T - \lambda)$:

```
for(k in 1:K)
{
  pph=simulPPh1(lambda,Tmax)
  mle=MLE1(pph,Tmax)
  Z[k]=sqrt(Tmax)*(mle-lambda)
}
```

iii. **With the density function :** Plot the histogram of the sample `Z` (which approximates the density of the Z_k's), and compare it with the density of the limit distribution $\mathcal{N}(0,\lambda)$:

```
hist(Z,freq=FALSE,main=paste("Tmax",Tmax,sep="="))
curve(dnorm(x,mean=0,sd=sqrt(lambda)),
      col="red",add=TRUE)
```

What do you observe when `Tmax` equals 1, 10, 100 and 500 ?

iv. **With the cdf :** Plot the empirical cumulative distribution function of `Z`, and compare it with the cumulative distribution function of a $\mathcal{N}(0,\lambda)$:

```
plot(ecdf(Z),main=paste("Tmax",Tmax,sep="="))
curve(pnorm(x,mean=0,sd=sqrt(lambda)),
      col="red",lwd=2,add=TRUE)
```

What do you observe when `Tmax` equals 1, 10, 100 and 500 ?

1.4 Statistical inference : hypothesis testing

Consider a given rate $\lambda_0 > 0$ to test. Given the observation of a homogeneous Poisson process with (unknown "true") rate λ observed on a fixed window `[0,T]`, we aim at testing

$$\mathcal{H}_0 : \lambda = \lambda_0 \quad \text{against} \quad \mathcal{H}_1 : \lambda \neq \lambda_0.$$

i. Construct a test of $\mathcal{H}_0$ against $\mathcal{H}_1$, and express the corresponding p-value in terms of the standard gaussian cdf.

ii. Write an `R` function `test1`, with arguments the observed homogeneous Poisson process `PPh`, the observation time `Tmax` and the rate `lambda0` to test, and which returns the p-value of the test constructed in the previous question.

iii. To validate this test on simulated data, we estimate for different values of λ, the probability (or proportion)

$$p(\lambda) = \mathbb{P}_\lambda(\text{reject } \mathcal{H}_0).$$

In particular, if λ satisfies $\mathcal{H}_0$ (i.e. $\lambda = \lambda_0$), then $p(\lambda)$ is the size of the test, and otherwise, it is the power of the test against the alternative λ.

The `R` function `plot.level.power1` in the file **TP-PoissonProcess-etud.R** plots for different values of the rate λ (in `TrueLambda`) confidence intervals for the proportion $p(\lambda)$.

Now, fix `alpha=0.05`, `nsimu=1000` and let `lambda0=2` and `TrueLambda=c(2,2.2,2.5,3)`.
Apply this function for `Tmax = 1,10,100` and `500`. Understand and comment the obtained graphs.

2 Homogeneous Poisson processes with fixed number of points

Second, we consider the case of a fixed number of points (and thus a random observation window).

2.1 Simulation

i. Recall the distribution of the interarrival times of a homogeneous Poisson process with rate `lambda`.

ii. Write an `R` function `simulPPh2` with arguments the rate `lambda` and the number of points `n`, which simulates such a process and returns the corresponding arrival times.

iii. As in the "fixed window" case, fix `lambda=2` and `n=20`, simulate a homogeneous Poisson process and plot both the counting process and the arrival times.

2.2 Maximum likelihood estimator

i. Write an `R` function `MLE2` which returns the maximum likelihood estimator of the rate of a homogeneous Poisson process `PPh` observed up to the `n`th point.

ii. Apply it on different simulated data. What do you observe ?

2.3 Asymptotic behavior of the MLE

As in section 1.3, illustrate the asymptotic behavior of the MLE $\hat{\lambda}_n$ of a homogeneous Poisson process with rate λ observed up to the nth point as n goes to $+\infty$, that is

i. (Strong) LLN-type result :

$$\hat{\lambda}_n \xrightarrow[n\to+\infty]{a.s.} \lambda.$$

ii. CLT-type result :

$$\sqrt{n}(\hat{\lambda}_n - \lambda) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0,\lambda^2).$$

2.4 Statistical inference : confidence intervals

i. Recall both asymptotic and non-asymptotic confidence intervals for the (unknown) rate `lambda` of a homogeneous Poisson process `PPh` observed up to the `n`th point.

ii. Write an `R` function with arguments `PPh`, `alpha=0.05` and a boolean `asymptotic`, and which, depending on the value of `asymptotic` (`TRUE` or `FALSE`), returns the corresponding confidence interval for `lambda`.

iii. A for the testing problem, we want to validate the confidence intervals on simulated data.
To do so, illustrate, in both asymptotic and non-asymptotic cases, the fact that the proportion of times the rate belongs to the obtained confidence interval is larger than the fixed confidence level `1-alpha`.
You may fix `lambda=2`, and consider `n=10` or `n=100`.

3 Inhomogeneous Poisson processes

We now aim at simulating inhomogeneous Poisson processes with given intensity function, on a given interval.

i. Recall the thinning algorithm. In particular, when does it apply ?

ii. Write an `R` function `simulPPi` with arguments the intensity function `lambda_fct`, `Tmax` and an upper bound `M`, which simulates an inhomogeneous Poisson processes with intensity function `lambda_fct()` on `[0,Tmax]`, and returns the corresponding arrival times.

iii. Simulate and represent inhomogeneous Poisson processes on $[0, 10]$ with intensity functions :
   - $\lambda_1 : t \mapsto 2 \times \mathbb{1}_{[0,7]}(t) + 8 \times \mathbb{1}_{]8,10]}(t)$.
   - $\lambda_2 : t \mapsto 2t$.
   - $\lambda_3 : t \mapsto \ldots$.