Poisson process and continuous-time Markov chains

Stochastic Processes - Assignment 2

Javier Martínez Llamas
Santiago Raposo
Ricardo Hortelano

1 Non-homogeneous Poisson Process

The account selected for the extraction of the tweets is *@realmadrid* since it has a constant number of retweets (RT), not extremely high, allowing us to measure the counting process in a reasonable time space. Although library *rtweet* obtains information about the retweets and their time stamps it is worth mentioning that the results detailed here do not reflect the actual behaviour of the account since retweet count is limited to 100.

1.1 Use descriptive statistics and graphics to explore the retweets data set in terms of the number of retweets by time and the time between retweets. Does it fit the hypothesis of a Poisson process?

We know that a Poisson process is a counting process related with the Poisson and Exponential distributions. Being a counting process a stochastic process $\mathbf{N} = \{N_t, t \geq 0\}$ satisfying:

- $N_t = 0$

- N_t is integer valued

- For $s < t, N_s \leq N_t$

- For $s < t, N_t - N_s$ represents the number of events that occur in the time intervals (s, t]

When plotting the tweet with maximum retweets (limited) we can see the behaviour of the retweet counting process.
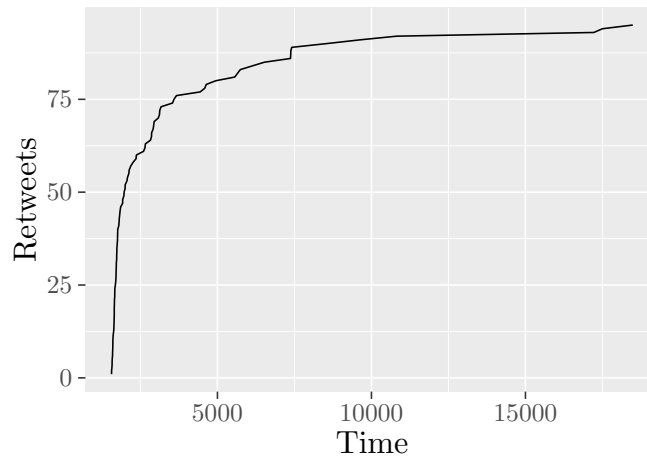

Figure 1: Tweet with maximum RT process

Each retweet counting process is defined by a time series, measuring the time for the next count (in minutes) since the creation of the tweet. That is, starting at $N_t = 0, t = 0$ the arrival times s increases the count by 1 unit.

Assuming that, if a user undoes a retweet, when acquiring the dataset the former time of retweet it is no longer recorded, the count can only increase and never decrease. The real model does not hold this assumption since this restriction is imposed by the internal functioning of Twitter and its API.

i.e. The following sequence of times would define a retweet counting process for a tweet with 6 RT. Where the first event occurs after 1.4 minutes.

$$1.400000 \; 1.616667 \; 1.716667 \; 1.716667 \; 1.900000 \; 2.616667$$

Knowing that, as stated in Section 6.7 of Dobrow RP *Introduction to Stochastic Processes with R*, a nonhomogeneous Poisson process is a counting process $(N_t)_{t \geq 0}$ with intensity function $\lambda(t)$ if

- $N_0 = 0$

- For all $t > 0$, N_t has a Poisson distribution with mean

$$\mathbb{E}[N_t] = \int_0^t \lambda(x)dx$$

- For $0 \leq q < r \leq s < t$, $N_r - N_q$ and $N_t - N_s$ are independent random variables.

1.2 Assuming you can model the number of retweets by time as a non-homogeneous Poisson process with intensity function $\lambda(t) = \theta e^{-\theta t}$, $t > 0$, graphically explore possible values of θ and choose the one that best fit your data. Explain all the considerations you make.

TODO by Ricardo que no tengo ni puta de lo que ha hecho

1.3 Write an R function to simulate data from a non-homogeneous Poisson process during a given time interval. Describe the inputs and outputs of your function.

As stated in section 11.5.1 of Ross, Sheldon *Introduction To Probability Models*, given n events of a non-homogeneous Poisson process by time T the n event times are independent with a common density function

$$f(s) = \frac{\lambda(s)}{m(t)}, \; 0 < s < T, \quad m(T) = \int_0^T \lambda(s)ds$$

By simulating $N(T)$, the number of events by time T, and then simulating $N(T)$ random variables from the previous density function we can generate a NHPP.

TODO Seguir explicando el algoritmo

Translated into the following R function

```
simulateNHPP <- function(intensity_function, time, lambda_bound) {
        X <- numeric(0)

        for ( t in 1:time){
                u <- runif(2)
                accept <- u[2] <= intensity_function(time*u[1]) / lambda_bound
                if (accept)
                        X <- c(X, time * u[1])
        }

        return(sort(X))
}
```