Simulation of Hawkes Process

Suhas Shastry

11/29/2018

Poisson Process

Poisson process is an elementary counting process used to fit the stochastic data. This model in two variants has found myriads of applications. Simple homogeneous Poisson process defined by,

$$P\{X(t) = n\} = e^{-\lambda t} \frac{(\lambda t)^n}{n!} \quad n = 0, 1, 2, \ldots$$

has a fundamental role in queuing theory. The rate λ is constant through out the time. Inhomogeneous Poisson process is an improved process where rate λ varies with time t. It is defined by

$$P\{X(t) = n\} = e^{-\Lambda(t)} \frac{(\Lambda(t))^n}{n!} \quad n = 0, 1, 2, \ldots$$

This process has applications from biology, ecology, chemistry to teletraffic engineering. Both the variants are plotted in Figure 1. Red dotted vertical line in the figure indicates the arrival of an event.

When stochastic data has clustered chunks, or if occurrence of an event invokes many other events immediately, modeling Poisson process would not be a good fit. In these scenarios, Hawkes process is an appropriate model.

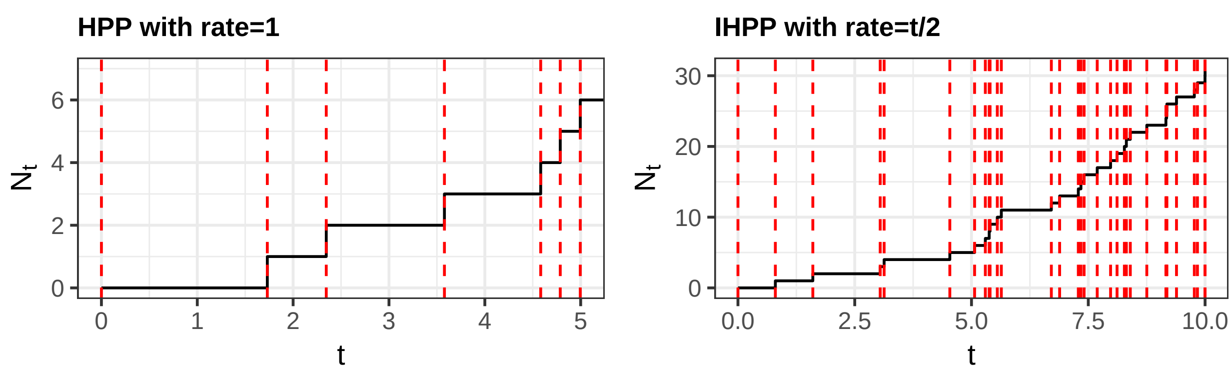

Figure 1: Poisson Process

Hawkes Process

Hawkes process is a non-Markovian extension of Poisson process [1] which is used to model self-excitation process. Exponential decaying process is defined by

$$\Lambda(t) = \lambda_0 + \sum_{t_i < t} \alpha e^{-\beta(t - t_i)}$$

where, λ_0 is a base rate of arrival process, α and β are positive constants. This process is simulated using Thinning procedure defined by Ogata in [2]. Simulated Hawkes process is depicted in Figure 2.

Log-likelihood of this process is defined by,

$$l = \sum_{i=1}^{k} log\left[\lambda_0 + \alpha \sum_{j=1}^{i-1} e^{-\beta(t_i - t_j)}\right] - \lambda_0 t_k + \frac{\alpha}{\beta} \sum_{i=1}^{k} \left[e^{-\beta(t_k - t_i)} - 1\right]$$

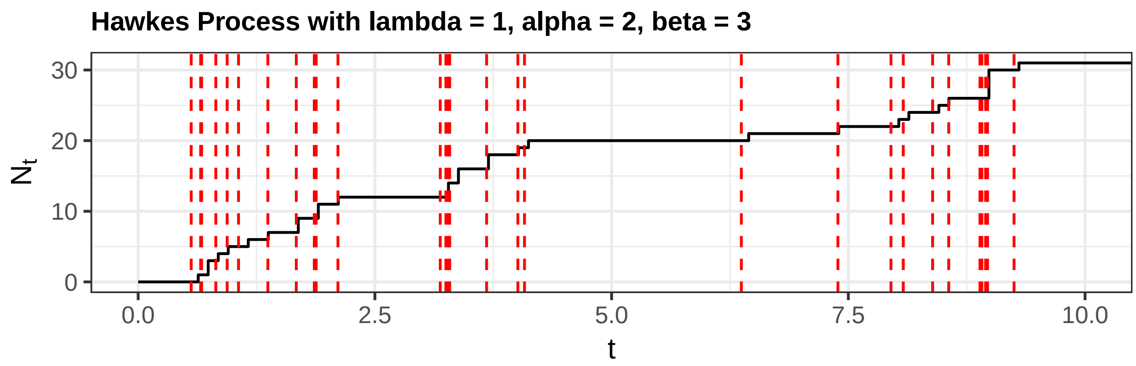

Figure 2: Hawkes Process

For the given data, parameters α, β and λ_0 are obtained by maximizing log-likelihood function. Hawkes process with parameters, $(\lambda_0 = 10, \alpha = 20, \beta = 30)$ was simulated and MLE equation was optimized in R. Below are the MLE estimates of the parameters.

```
## [1]  7.224359 19.273938 25.854809
```

Application of Hawkes Process

Self-excitation of events happen in various real life scenarios. Earthquakes usually occur in clusters, where a mainshock immediately trigger many aftershocks. Hawkes process is applied to account immigration and births for population estimation. But recently, application of Hawkes process in modeling crime dataset has increased. Hence I choose, crime dataset on Burglary in Chicago city in the area Beat 423 during 2017-18. Plot of data, both arrival process and histogram are depicted in Figure 3.

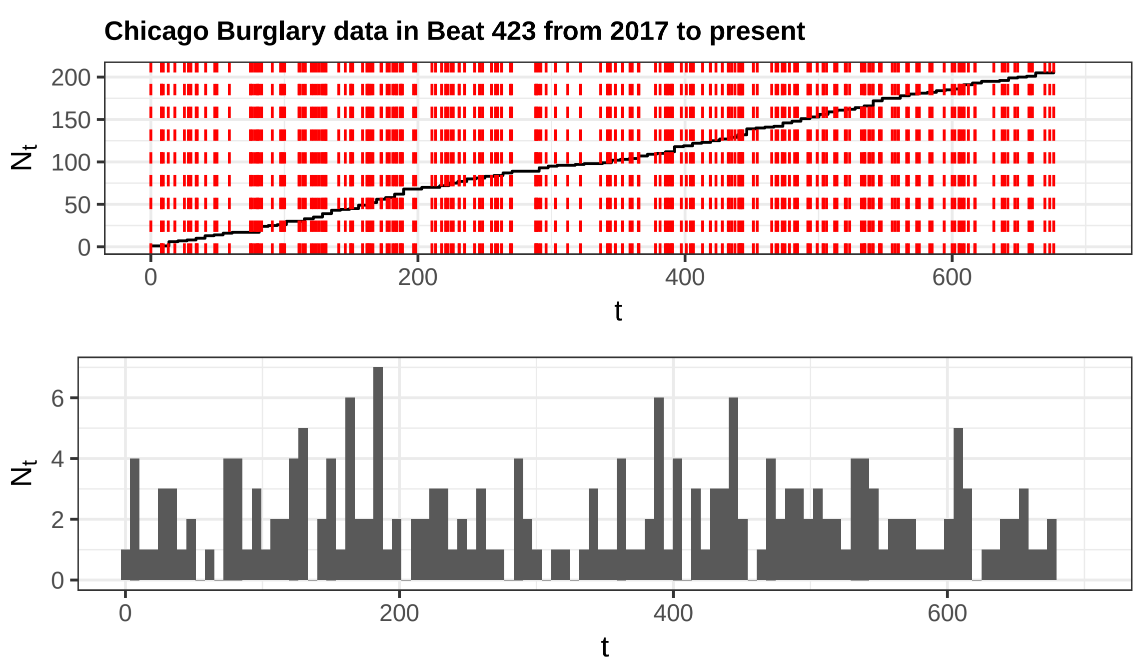

Figure 3: Chicago Burglary data

MLE estimators for above dataset are

```
## [1] 0.29168892 0.04131991 0.79237880
```

Hawkes process was simualated with parameters ($\lambda_0 = 0.29, \alpha = 0.04, \beta = 0.79$) and Figure 4. is the arrival process and histogram of the simulated process.

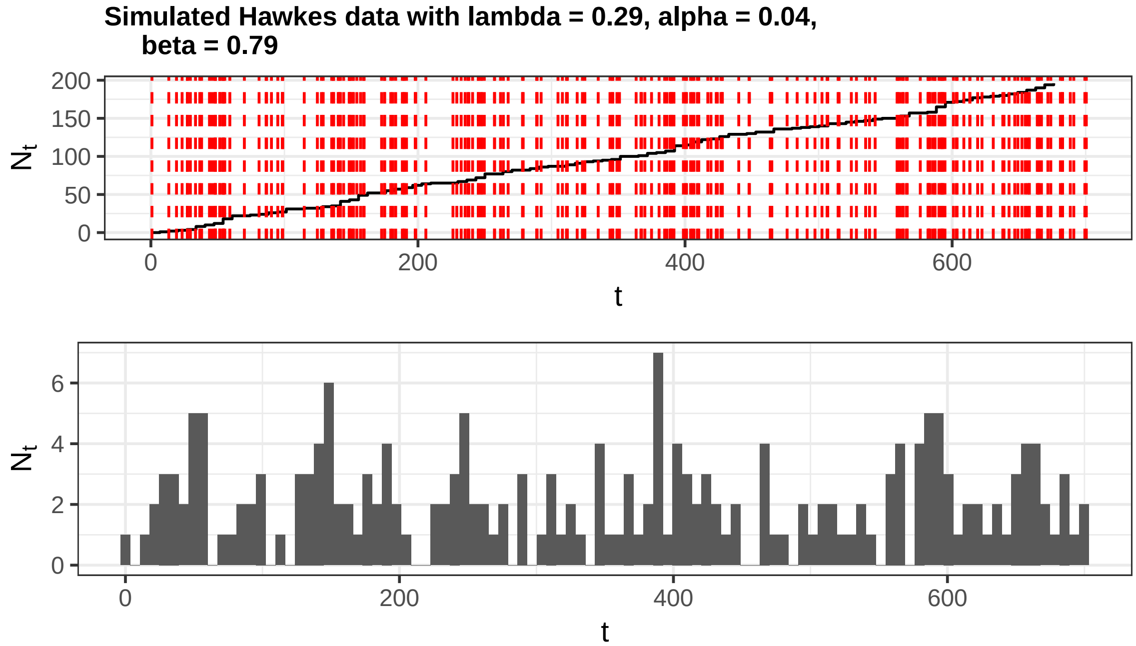

Figure 4: Simlulated Hawkes process

The histograms of Chicago dataset and simulated Hawkes process look similar. Hence Kolmogorov–Smirnov test was used to check the goodness of fit. Hawkes process with same parameters was simulated 1000 times and compared with Chicago burglary data set each time. Figure 5. depicts the histogram of p-values of 1000 KS test. Red dotted line indicates the mark $\alpha = 0.05$.

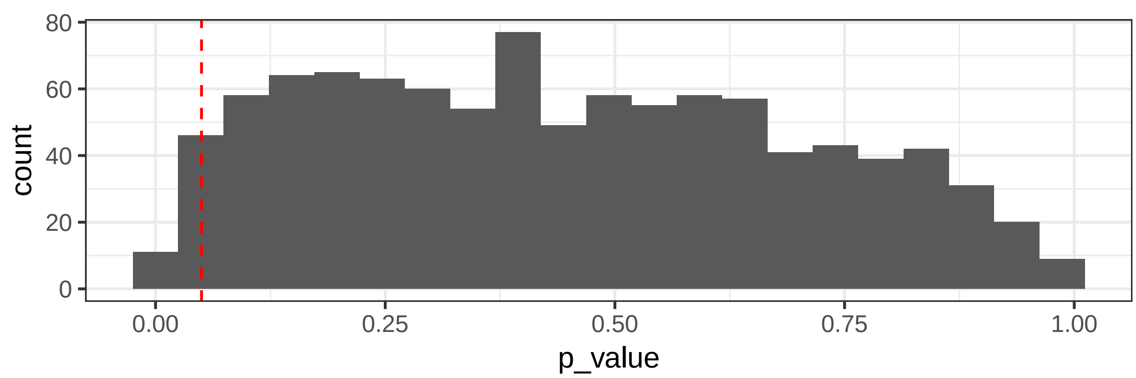

Figure 5: p-values of KS-test

Proportion of p-values less than the mark $\alpha = 0.05$ is

```
mean(p_value<0.05)
```

```
## [1] 0.035
```

As we rejected the null hypothesis only a fraction of times, we conclude that there is a strong evidence against the alternative hypothesis, "Chicago burglary data is not following Hawkes process". In other words, Hawkes process is not a bad fit for Chicago burglary data.

References

1. Patrick J. Laub, Thomas Taimre, Philip K. Pollett, Hawkes Process - https://arxiv.org/abs/1507.02822
2. Y. Ogata, Journal of the American Statistical Association 83(401), 9 (1988)