

Estimation of Multivariate Hawks Process

Yunxuan Xiao

June 8, 2019

1 Introduction

Multivariate Hawkes process is a multi-dimensional extension of the univariate case. Apart from the self-excitation feature, it also has the mutual-exciting between events in different dimensions. In this experiment, to estimate the parameters of multivariate process, we utilize EM algorithm to maximize the log likelihood. Since we have already implemented simulation process in project1, now we will use the simulated event sequences to estimate original model parameters. In the following section, I will first introduce the principle of our experiments and then illustrate our experiment results.

2 Multivariate Hawks Process

Define $\{t_i\}$ as the simulated time sequence. d_i refers to the dimension which the i -th event belongs to. For a multivariate Hawks Process, the intensity function of m -th dimension can be defined as follow:

$$\lambda_d(t) = \mu_d + \sum_{\{i=1, t_i < t\}}^N \alpha_{dd_i} e^{-\beta(t-t_i)} \quad (1)$$

The log-likelihood:

$$\begin{aligned} \log L &= \sum_{d=1}^M \left\{ \sum_{i, d_i=d}^N \log \lambda_{d_i}(t_i) - \int_0^T \lambda_d(t) dt \right\} \\ &= \sum_{i=1}^n \log \left(\mu_{d_i} + \sum_{t_j < t_i} \alpha_{d_i d_j} e^{-\beta(t_i - t_j)} \right) - T \sum_{d=1}^M \mu_d - \sum_{d=1}^M \sum_{j=1}^n \alpha_{dd_j} G(T - t_j) \end{aligned}$$

where $G(t) = \int_0^t e^{-\beta t} dt = -\frac{1}{\beta}(e^{-\beta t} - 1)$.

According to Jensen's inequality, we can derive the lower bound of log-likelihood:

$$\log L \leq \sum_{i=1}^N (p_{ii} \log \frac{\mu_{d_i}}{p_{ii}} + \sum_{j=1}^{i-1} p_{ij} \log \frac{\alpha_{d_i d_j} e^{-\beta(t_i - t_j)}}{p_{ij}}) \quad (2)$$

$$- T \sum_{d=1}^M \mu_d - \sum_{d=1}^M \sum_{j=1}^n \alpha_{dd_j} G_{dd_j}(T - t_j) \quad (3)$$

2.1 EM Algorithm

We can derive a trivial solution where the equation holds, which can maximizes the log-likelihood. Intuitively, p_{ij} can be interpreted as the probability that the i -th event is influenced by a previous event j in the event sequence and p_{ii} is the probability that i -th event is sampled from the base intensity.

E-step

$$p_{ii}^{(k+1)} = \frac{\mu_{d_i}^{(k)}}{\mu_{d_i}^{(k)} + \sum_{j=1}^{i-1} \alpha_{d_i d_j}^{(k)} e^{-\beta(t_i - t_j)}}$$

$$p_{ij}^{(k+1)} = \frac{\alpha_{d_i d_j}^{(k)} e^{-\beta(t_i - t_j)}}{\mu_{d_i}^{(k)} + \sum_{j=1}^{i-1} \alpha_{d_i d_j}^{(k)} e^{-\beta(t_i - t_j)}}$$

M-step Then do partial differential equation for μ and α to update the parameters:

$$\mu_d^{(k+1)} = \frac{1}{T} \sum_{i=1, d_i=d}^n p_{ii}^{(k+1)}$$

$$\alpha_{uv}^{(k+1)} = \frac{\sum_{i=1, d_i=v}^n \sum_{j=1, d_j=v}^{i-1} p_{ij}^{(k+1)}}{\sum_{j=1, d_j=v}^n G(T - t_j)}$$

$$\beta^{(k+1)} = \frac{\sum_{i>j} p_{ij}^{(k+1)}}{\sum_{i>j} (t_i - t_j) p_{ij}^{(k+1)}}$$

Then iterate until convergence, the parameters will be fitted.

3 Experiment

3.1 Single Sequence Setting

We first estimate on single event sequence with variant length. Five sequences with different length, 100, 500, 1000, 3000 and 5000 are chosen as training cases. We set the maximum iteration as 30 and visualized the relative error and negative log likelihood through the iterations.

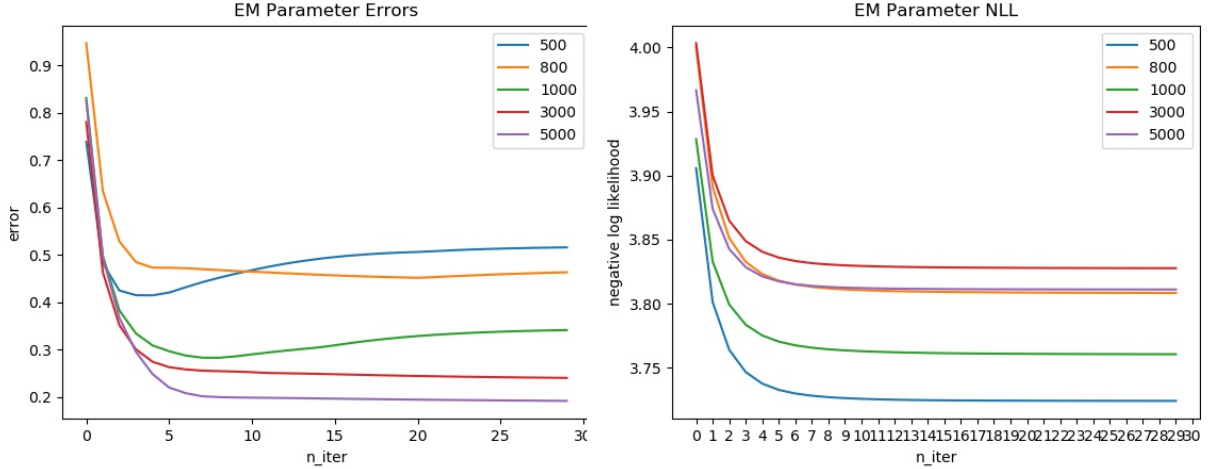


Figure 1: Metrics: Relative Error and Negative log likelihood. The number in legend bar refers to the number of events in a sampled sequence. In the left figure, we notice that as the sequence gets longer, the relative error after convergence declines. The best relative error is 0.19 when the sequence length is 5000.

| sequence length | 100 | 500 | 800 | 1000 | 3000 | 5000 |
|-----------------|------|------|------|------|------|-------------|
| best iter error | 0.84 | 0.41 | 0.45 | 0.28 | 0.24 | 0.19 |

Table 1: Lowest relative error in single-seq setting.

3.2 Multi-sequence Setting

We also tried to use multiple generated sequences to improve the generalization ability. By sampling 20 sequences for each length, we run EM algorithm on each input sequence independently and compute the average of these parameters(Ensemble) to reduce the variance of data. Limited by time we only tries sequence with length from 100 to 1000. It shows that our multi-seq ensemble method significantly reduced the relative error.

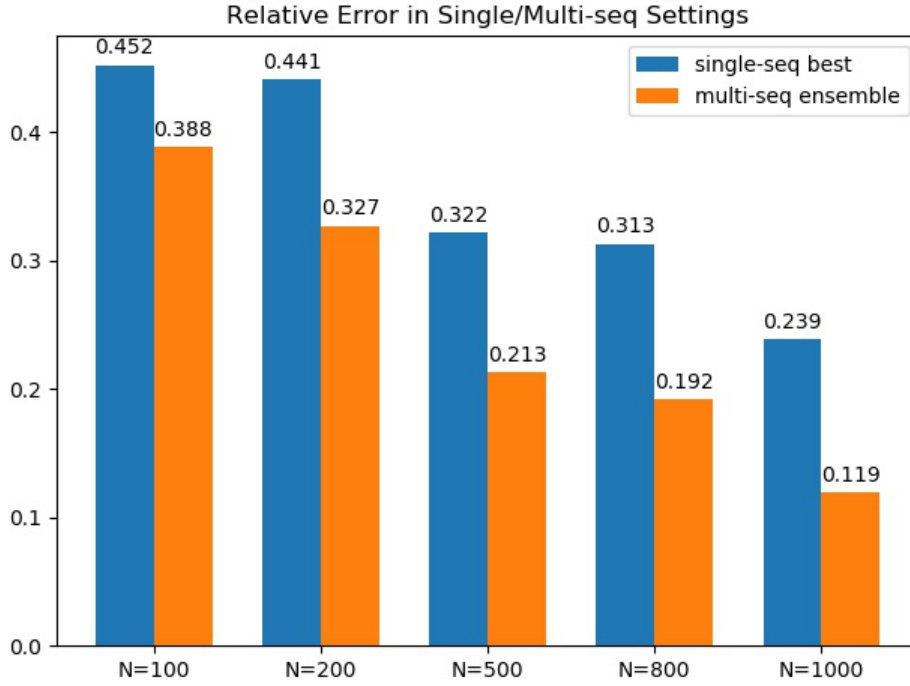


Figure 2: N refers to the total number of events for the sampled sequences. In each group, the blue bar refers to the best error among 20 generated sequence, while the red bar refers to the result after ensemble. There is a huge improvement after taking average for each parameter. The best relative score under Multi-seq setting achieves **0.119** when $N = 1000$.

| sequence length | 100 | 200 | 500 | 800 | 1000 |
|-----------------|-------|-------|-------|-------|--------------|
| best iter error | 0.338 | 0.327 | 0.213 | 0.192 | 0.119 |

Table 2: Lowest relative error in multi-seq setting.

3.3 How to run?

Please refer to **Readme.md** for more information about running the code.