Learning Hawkes Processes Under Synchronization Noise
Supplementary Material

1. Technical Derivations

1.1. Gradient of the Log-Likelihood

We derive the gradient of the log-likelihood of the DESYNC-MHP model with respect to the synchronization noise parameters $\boldsymbol{z}$. It corresponds to the update rule of $\boldsymbol{z}$ in Algorithm 1. First, the gradient of (4) with respect to the noise parameter in the k-th dimension can be written as

$$\nabla_{z^k} \log \mathbb{P}(\tilde{\mathbf{t}}|\boldsymbol{z}, \theta) = \frac{\partial}{\partial z^k} \left[\sum_{i=1}^{d} \left(\sum_{\tau \in \widetilde{N}_i(T)} \log \lambda_i(\tau - z^i) - \int_{t_0 - z^i}^{T - z^i} \lambda_i(t) dt \right) \right],$$

where

$$\lambda_i(t) = \mu_i + \sum_{j=1}^{d} \sum_{\tau \in \mathcal{H}_t^j} \kappa_{ij}(t - \tau) = \mu_i + \sum_{j=1}^{d} \int_0^t \tilde{\kappa}_{ij}(t - \tau) dN_j(\tau).$$

For ease of reading, we dropped the explicit dependence of the intensity $\lambda_i(t)$ on the history $\mathcal{H}_t$ of the process.

Substituting the above intensity into the log-likelihood implies

$$\frac{\partial}{\partial z^k} \sum_{j=1}^{d} \Bigg\{ \sum_{\tau \in \widetilde{N}_j(T)} \log \left(\mu_j + \sum_{i=1}^{d} \sum_{s \in \widetilde{N}_i(\tau)} \alpha_{ji} \tilde{\kappa}(\tau - z^j - s + z^i) \right) - \mu_j (T' - t_0') - \sum_{i=1}^{d} \int_{t_0'}^{T'} \int_0^{T' - s + z^i} \alpha_{ji} \tilde{\kappa}(t) dt d\widetilde{N}_i(s) \Bigg\}, \tag{1}$$

where $t_0' := t_0 - \min_i z^i$, $T' := T - \max_i z^i$, and

$$\tilde{\kappa}(t) = \frac{e^{-\beta t} + e^{-(\gamma - \beta')t}}{1 + e^{-\gamma t}}.$$

is the smooth approximation of the exponential kernel defined in Equation (11). Note that in the above equation, we approximated the boundary of the integral by $[t_0', T')$ to account for windowing effects.

First, we compute the derivative of $\tilde{\kappa}(\tau - z^k - s + z^i)$ with respect to z^k,

$$\frac{\partial}{\partial z^k} \tilde{\kappa}(u_{ki}) = -\frac{(e^{-\beta u_{ki}} + e^{-(\gamma - \beta')u_{ki}}) \gamma e^{-\gamma u_{ki}}}{(1 + e^{-\gamma u_{ki}})^2} + \frac{\left(\beta e^{-\beta u_{ki}} + (\gamma - \beta') e^{-(\gamma - \beta')u_{ki}} \right)}{1 + e^{-\gamma u_{ki}}}, \tag{2}$$

where $u_{ki} := \tau - z^k - s + z^i$. Further, $\frac{\partial}{\partial z^i} \tilde{\kappa}(u_{ki}) = -\frac{\partial}{\partial z^k} \tilde{\kappa}(u_{ki})$. The other important term is

$$\frac{\partial}{\partial z^k} \int_{t_0'}^{T' - s + z^k} \tilde{\kappa}(t) dt = \tilde{\kappa}(T' - s + z^k). \tag{3}$$

By taking the derivative of the log-function, Equation (1) can be written as follows

$$\sum_{j=1}^{d}\left\{\sum_{\tau\in\widetilde{N}_j(T)}\frac{\sum_{i=1}^{d}\sum_{s\in\widetilde{N}_i(\tau)}\alpha_{ji}\nabla_{z^k}\tilde{\kappa}(\tau-z^j-s+z^i)}{\mu_j+\sum_{i=1}^{d}\sum_{s\in\widetilde{N}_i(\tau)}\alpha_{ji}\tilde{\kappa}(\tau-z^j-s+z^i)}-\sum_{i=1}^{d}\int_{t_0'}^{T'}\left(\nabla_{z^k}\int_0^{T'-s+z^i}\alpha_{ji}\tilde{\kappa}(t)dt\right)d\widetilde{N}_i(s)\right\}, \quad (4)$$

Substituting (2) and (3) into (4) implies the result.

The gradient of the log-likelihood with respect to α_{kb} for some k and b in $\{1,...,d\}$ is

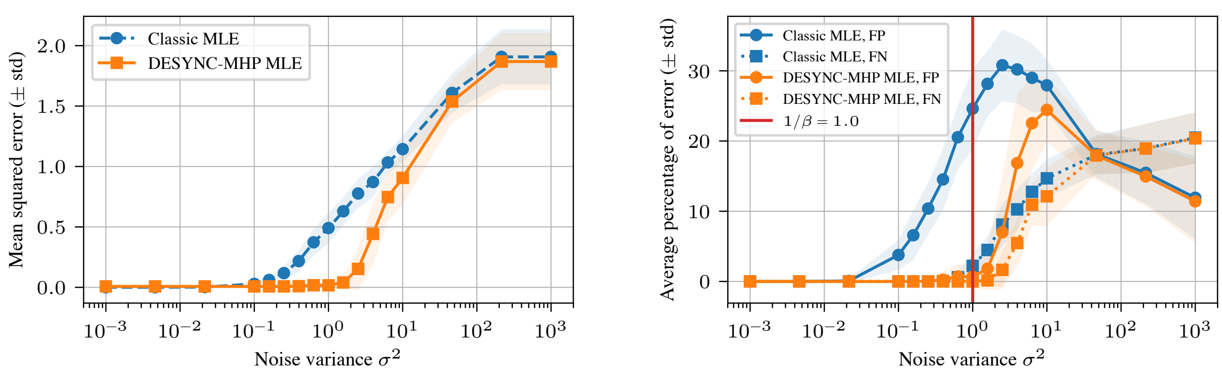
$$\sum_{\tau\in\widetilde{N}_k(T)}\frac{\sum_{s\in\widetilde{N}_b(\tau)}\tilde{\kappa}(\tau-z^k-s+z^b)}{\mu_k+\sum_{s\in\widetilde{N}_b(\tau)}\alpha_{kb}\tilde{\kappa}(\tau-z^k-s+z^b)}-\int_{t_0'}^{T'}\int_0^{T'-s+z^b}\tilde{\kappa}(t)dtd\widetilde{N}_b(s). \quad (5)$$

In (5), we have

$$\int_0^x\tilde{\kappa}(t)dt\approx\frac{1-e^{-\beta x}}{\beta}+\frac{\log 2-\log(1+e^{-(\gamma-\beta')x})}{\gamma-\beta'},$$

when $\gamma\gg\beta$ and $\gamma/(\gamma-\beta')\approx 1$.

2. Additional Experimental Results



(a) Analysis of the average mean squared error as a function of the noise variance.

(b) Analysis of the different error types as a function of the noise variance. FP refers to false positive errors, and FN refers to false negative errors.

Figure 1: Analysis of the sensitivity to the noise scale for various noise variances. ($d=10$ is fixed.)

Sensitivity to the noise scale σ^2. In Section 5, we discussed the accuracy of the DESYNC-MHP ML estimator compared to the classic ML estimator. To show that the DESYNC-MHP ML estimator is also capable of accurately recovering the coefficients $\{\alpha_{ij}^*\}$ and not just whether they are non-zero, we also report the mean squared error

$$\mathrm{MSE}\coloneqq\sum_{i,j}\left(\alpha_{ij}^*-\hat{\alpha}_{ij}\right)^2/\sum_{i,j}{\alpha_{ij}^*}^2$$

between the ground truth and estimated excitation matrices in Figure 1.

In addition, to gain a better understanding of the behavior of both estimators in the higher noise regimes, we also display the evolution of the false positive (FP) and false negative (FN) in Figure 1b. We see that the number of false positive (*i.e.,* wrongly learnt causal relation) first increases and is then followed by an increase in the number of false negative (*i.e.,* edges misidentified for non-existent). It appears that, for increasingly high noise levels, both estimators lead to sparser networks.