Mutually Regressive Point Processes

Ifigeneia Apostolopoulou [1], Scott Linderman [2], Kyle Miller [3] and Artur Dubrawski [1,3]

[1]Machine Learning Department, Carnegie Mellon University [2]Department of Statistics, Stanford University, [3]AutonLab

Introduction

Many phenomena and applications involve asynchronous events such as social media dynamics, neuronal activity, or high frequency financial markets. *Point Processes (PPs)* can model distributions of sequences of events.

- The *Cox Process* is characterized by a stochastic intensity function but it does not capture temporal correlations explicitly.
- The *Linear Hawkes Process (HP)* can capture purely excitatory temporal interactions.
- *Non-Linear Hawkes Processes* allow for both excitatory and inhibitory interactions but their inference is generally intractable.
- *Poisson Process Generalized Linear Models (PP-GLMs)* are efficient, discrete-time approximations of Non-Linear Hawkes Processes, but they can turn out to be poor generative models. Moreover, the estimated regression coefficients may vary widely depending on the boundaries chosen for aggregation.

Contributions

- We develop the first class of **Bayesian** point process models—**Mutually Regressive Point Processes (MR-PP)**—that allow for **nonlinear temporal interactions** while still admitting an efficient, fully-Bayesian inference algorithm in **continuous time**.
- **Soft relational constraints** can be established by a joint prior.
- MR-PP recovers **physiological** neuronal **dynamics**.
- MR-PP achieves **competitive predictive likelihood**.

Hierarchical Mutually Regressive Point Process

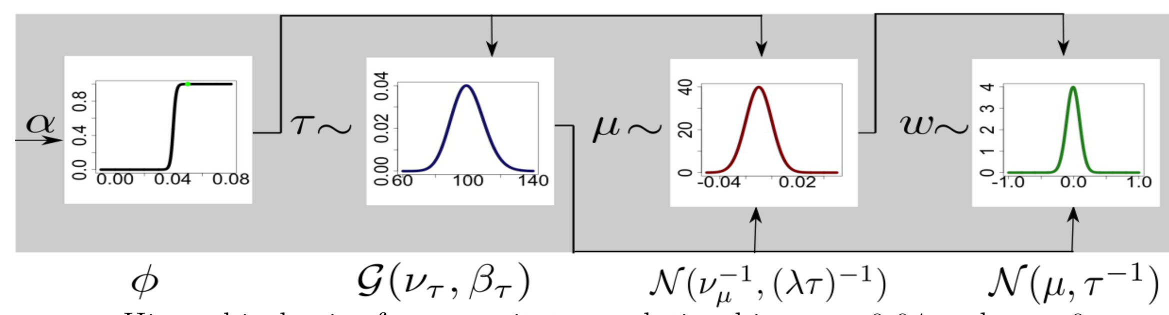
A Gaussian prior for the weights is assumed:

$$\boldsymbol{w}_n \sim \mathcal{N}(\boldsymbol{\mu}_n, \boldsymbol{\Sigma}_n).$$

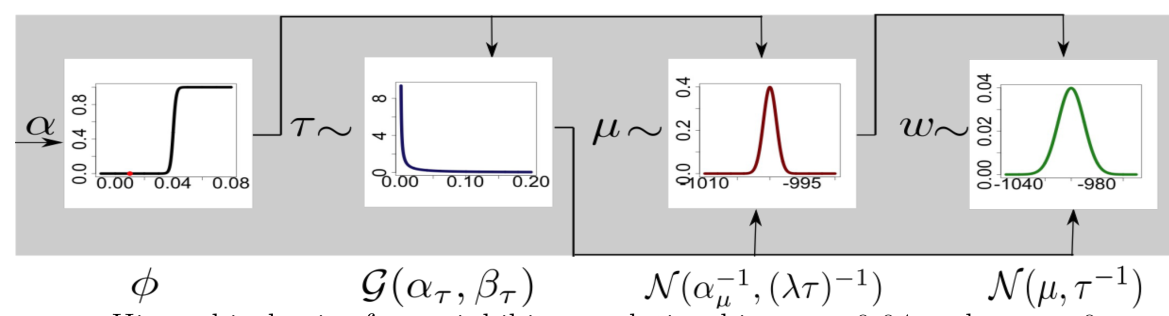
An **inverse relationship between the excitatory and the repulsive effect** of type m on type n can be imposed with a ***Sparse Normal-Gamma*** prior defined as:

$$\tau_{m,n} \sim \text{Gamma}(\nu_\tau \phi(\alpha_{m,n}) + \alpha_\tau, \beta_\tau),$$
$$\mu_{m,n} \sim \mathcal{N}(-(\nu_\mu \phi(\alpha_{m,n}) + \alpha_\mu)^{-1}, (\lambda_\mu \tau_{m,n})^{-1}),$$

where the constants are positive and ϕ is a monotonically increasing function.



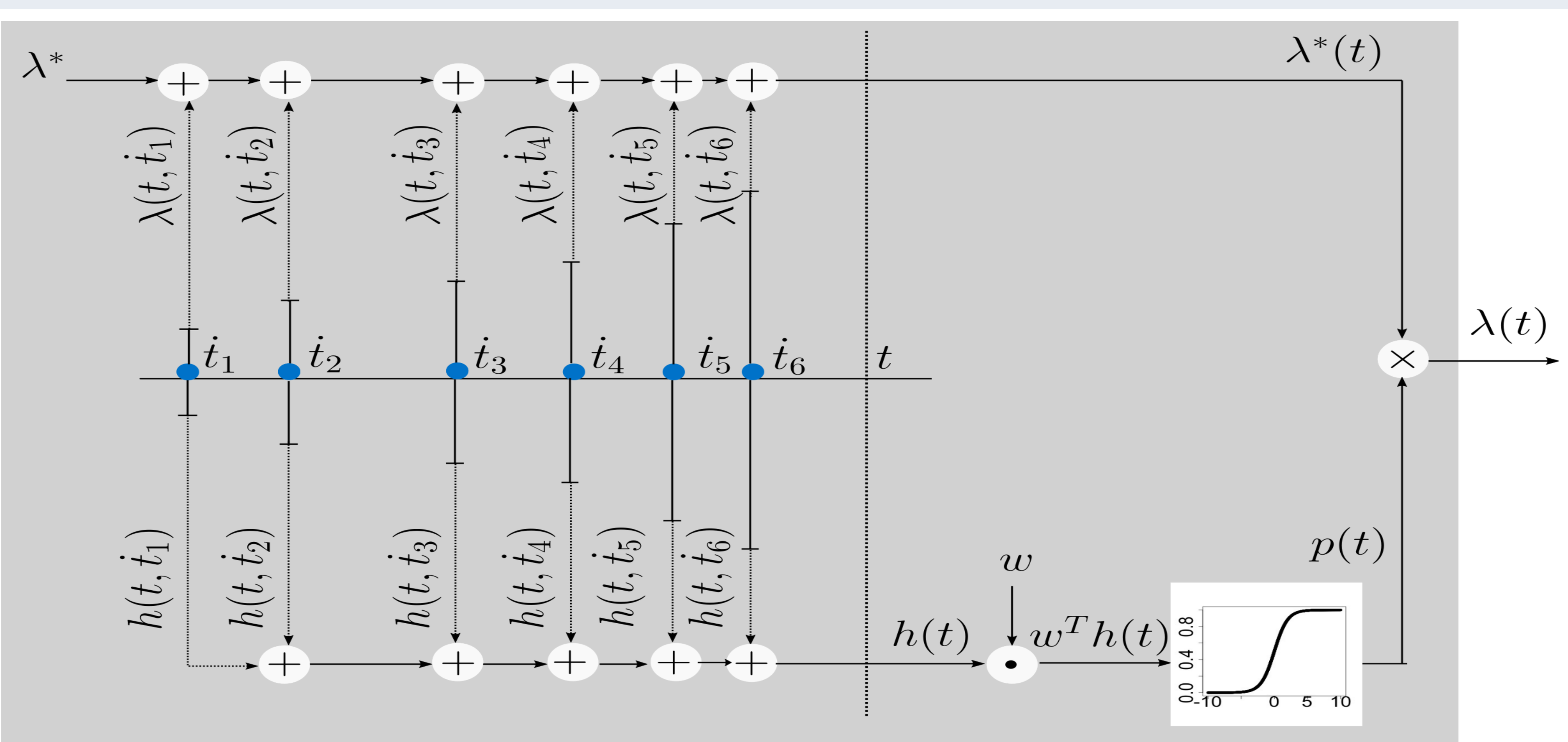
Hierarchical prior for an excitatory relationship: $\alpha > 0.04$ and $w \approxeq 0$



Hierarchical prior for an inhibitory relationship: $\alpha < 0.04$ and $w << 0$

Mutually Regressive Point Process: a generalization of the Hawkes Process

A Point Process $\mathcal{PP}(\lambda(t))$ is characterized by an intensity function $\lambda(t)$, so that in an infinitesimally wide interval $[t, t+dt]$, the probability of the arrival of a new event is $\lambda(t)dt$.

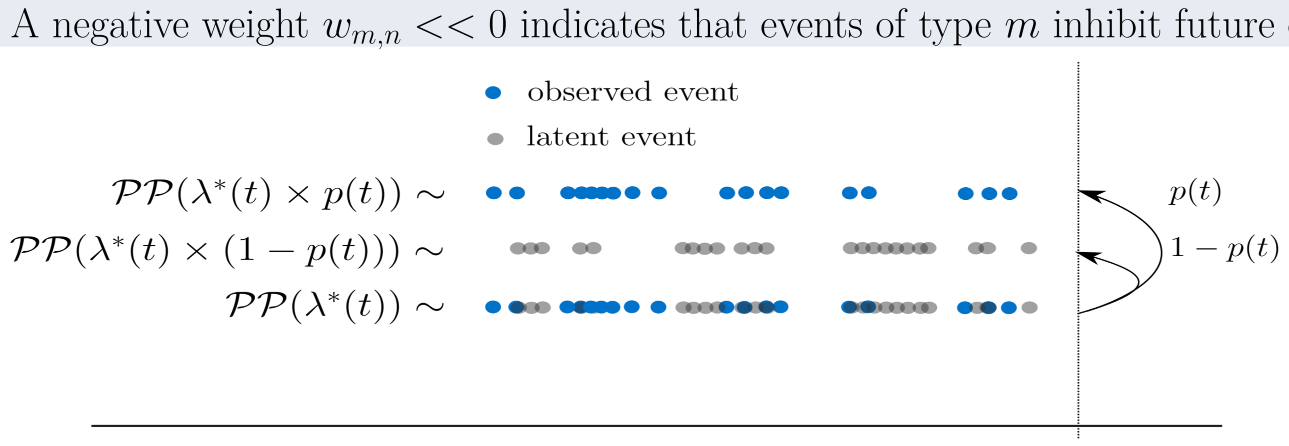


Computational flow of the MR-PP intensity

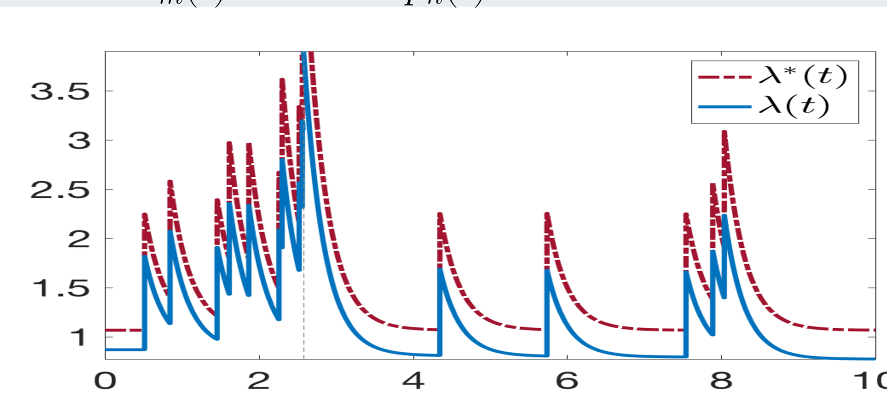
intensity of type $n = 1, \ldots, N$:

$$\lambda_n(t) = \lambda_n^*(t) \times p_n(t)$$
$$\lambda_n^*(t) = \lambda_n^* + \sum_{m=1}^{N} \sum_i \lambda_{m,n}(t, t_i^m)$$
$$t > t_i^m,\ \lambda_n^* \geq 0,\ \lambda_{m,n}(t, t_i^m) \geq 0$$
$$\lambda_{m,n}(t, t_i^m) = \alpha_{m,n} \times \phi(t, t_i^m)$$
$$p_n(t) = \sigma(\boldsymbol{w}_n^T \boldsymbol{h}(t))$$
$$\sigma(x) = (1 + e^{-x})^{-1}$$
$$h_m(t) = \sum_i h(t, t_i^m)$$
$$h(t, t_i^m) \geq 0$$

A negative weight $w_{m,n} << 0$ indicates that events of type m inhibit future events of type n since $h_m(t)$ decreases $p_n(t)$.



Poisson Thinning



HP vs MR-PP intensity function

Bayesian inference

Latent Events for Tractability

The likelihood of the sequence $\mathcal{T} \triangleq \{t_i\}$ of an event sequence generated by a point process $\mathcal{PP}(\lambda(t))$ in $[0, T]$ is:

$$p(\mathcal{T} \mid \lambda(t)) = \boxed{\exp\left\{-\int_0^T \lambda(t)\,\mathrm{d}t\right\}} \times \prod_i \lambda(t_i).$$

The integral can be intractable for non-linear $\lambda(t)$.

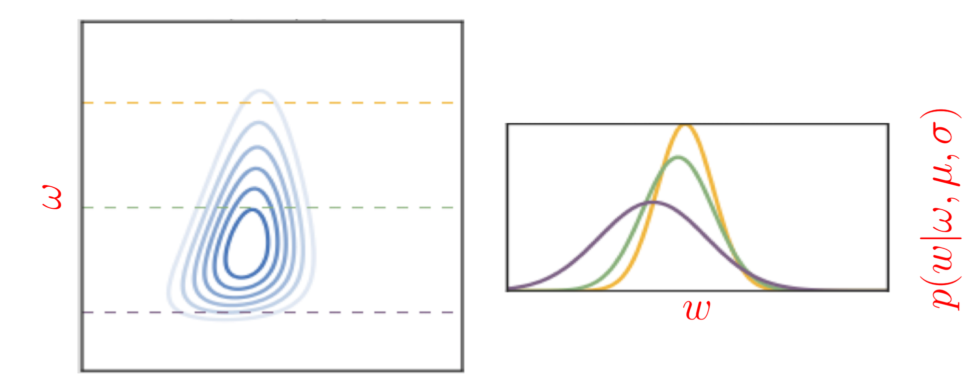
$$\tilde{\mathcal{T}}_n \sim \mathcal{PP}(\lambda_n^*(t) \times (1 - p_n(t)))$$

The joint likelihood of the observed ($\dot{T}_n$) and latent events ($\tilde{T}_n$) is:

$$p(\dot{\mathcal{T}}_n \cup \tilde{\mathcal{T}}_n \mid \lambda_n(t)) \propto \boxed{\exp\left\{-\int_0^T \lambda_n^*(t)\,\mathrm{d}t\right\}} \times \prod_i \lambda_n^*(t_i^n).$$

Pólya-Gamma Augmentation

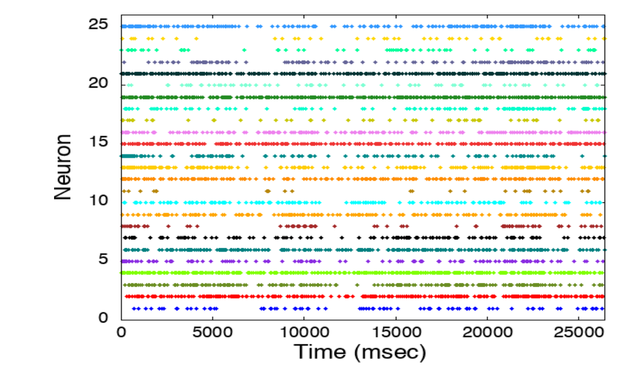
The sigmoidal term **destroys the conjugacy and yields non-Gaussion posteriors** for $\boldsymbol{w}_n$. Pólya-Gamma latent variables ω_i^n associated with each event t_i^n render the model conjugate.



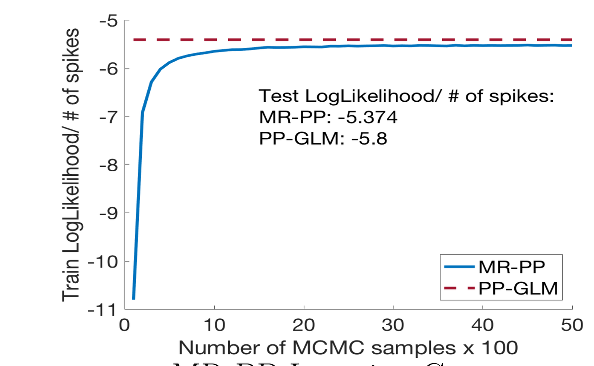
The posterior for sampling ω_i^n is a Pólya-Gamma distribution:

$$p(\omega_i^n \mid \ldots) = \mathcal{PG}_m(\omega_i^n; 1, \boldsymbol{w}_n^T \boldsymbol{h}(t_i^n)).$$

Multineuronal study



Multineuronal activity in cat visual cortex



MR-PP Learning Curve

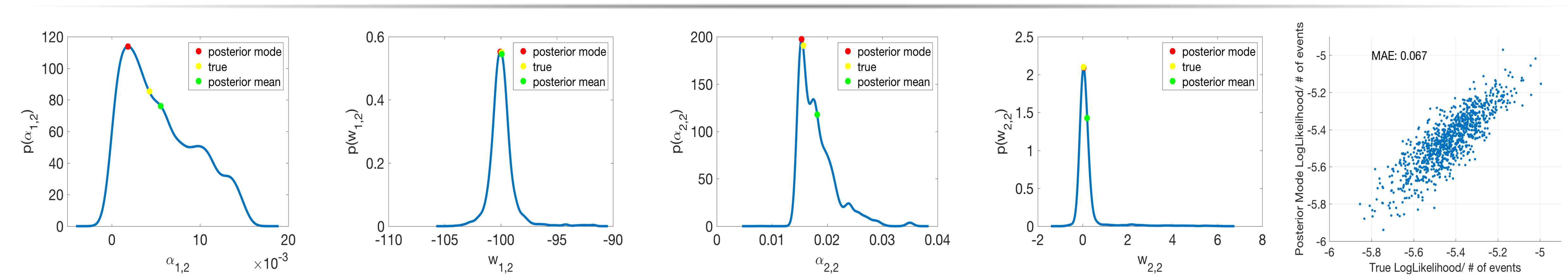
The log-likelihood of the MR-PP for the held-out, second half of the spike train, is larger.

Discussion

- Relevant Application Domains: modeling clinical events, complex biological networks, or sequential events in financial markets
- Model Refinements: variational inference for scalability, hyperparameters selection
- Prediction Tasks: causality, network structure discovery

Work partially supported by DARPA under award FA8750-17-2-013, Onassis Foundation. and A.G. Leventis Foundation.

Synthetic validation

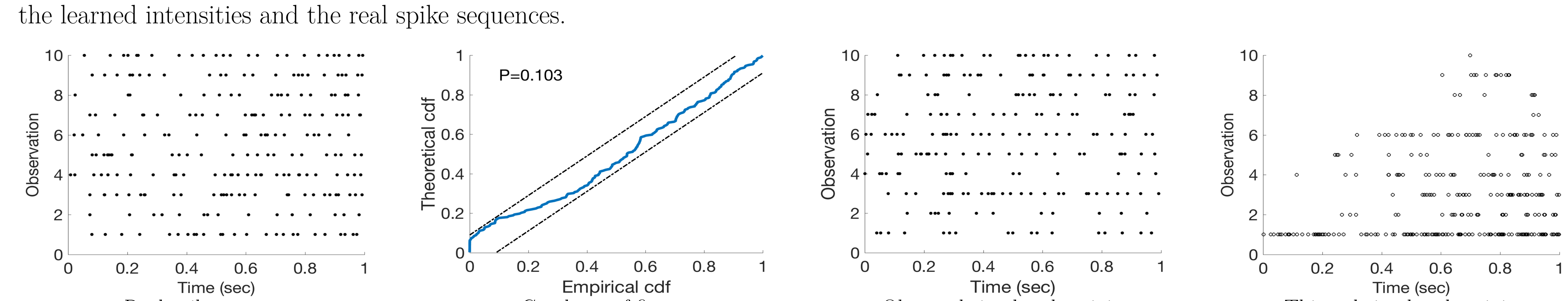


Effects on Type II

- The types are self-excited ($w_{i,i} \approxeq 0$ and $\alpha_{i,i} > 0.015$) and cross-inhibited by the other type ($w_{i,j} << 0$ and $\alpha_{i,j} < 0.015$).
- **Correct temporal correlations between the event types were identified**.
- The **predictive log-likelihood of the learned MR-PP** for 1,000 held-out event sequences is **almost equal to that of the real parameters**.

Analysis of stability of single neuron spiking dynamics

PP-GLMs may yield non-physiological spiking patterns when used as generative models because of explosive firing rates although they pass the goodness-of-fit test. On the other hand, **MR-PP both passes the goodness-of-fit test** ($p - value > 0.05$) and **generates stable spike trains similar to those used for the learning**. We used the the Kolmogorov-Smirnov (KS) test after applying the time-rescaling theorem on the learned intensities and the real spike sequences.



Stability analysis of the MR-PP for spike patterns from monkey cortex