Learning Joint Intensity in a Multivariate Poisson Process on Statistical Manifolds

Simon Luo[1,2], Feng Zhou[3], Lamiae Azizi[1,2] and Mahito Sugiyama[4,5]

[1] School of Mathematics and Statistics, The University of Sydney [2] Data Analytics for Resources and Environments (DARE), Australian Research Council

[3] Tsinghua University [4] National Institute of Informatics [5] JST, PRESTO

Thirty-fourth Annual Conference on Neural Information Processing Systems (NeurIPS 2020) Workshop on Deep Learning through Information Geometry

Summary

- Unify the framework between Poisson Processes and Generalized Additive Models (GAMs)
- Formulate an efficient approximation of the joint intensity function in a point process
- Approximate high-order intensity functions using projections in low dimensional space
- Estimate high-order intensity functions by minimizing the Kullback-Leibler Divergence from a set of samples to a distribution on a partial order structure

Poisson Process

$$\log p\left(\{\boldsymbol{t}_i\}_{i=1}^N \mid \lambda(\boldsymbol{t})\right) = \sum_{i=1}^{N} f(\boldsymbol{t}_i) - \int \exp(f(\boldsymbol{t}))dt$$

Generalized Additive Model

$$\log \lambda_J(\boldsymbol{t}) = \sum_{j \in J} f_j(t^{(j)}) - \beta_J$$

Including terms for Higher-Order Feature Interaction Into the model

$$\log \lambda_J(\boldsymbol{t}) = \sum_{j \in J} f_{\{j\}}(t^{(j)}) + \sum_{j_1, j_2 \in J} f_{\{j_1,j_2\}}(t^{(j_1)}, t^{(j_2)}) + \dots + \sum_{j_1,\dots,j_k \in J} f_{\{j_1,\dots,j_k\}}(t^{(j_1)}, \dots, t^{(j_k)}) - \beta_J$$

Kolmogrov-Arnold Representation Theorem

Any multivariate continuous function can be represented as a superposition of lower–dimensional functions

$$\lambda_J = \frac{1}{\exp(\beta_J)} \exp\left(\sum_{I \subseteq J} f_I(\boldsymbol{t}^{(I)})\right) \propto \exp\left(\sum_{I \subseteq J} f_I(\boldsymbol{t}^{(I)})\right)$$

The log-likelihood with higher order interaction terms

$$\log p\left(\{\boldsymbol{t}\}_{i=1}^N \mid \lambda(\boldsymbol{t})\right) = \sum_{i=1}^{N} \exp\left(\sum_{I \subseteq J, |I| \le k} f_I(\boldsymbol{t}^{(I)})\right) - \beta'$$

Log-Linear Model on a Partially Ordered Set (poset)

The log-linear model is defined over a partial order set (poset) $(S, \preccurlyeq)$

Dual coordinate system (θ, η) of a statistical manifold

•η Expectation parameter

•θ Natural parameter in the exponential family

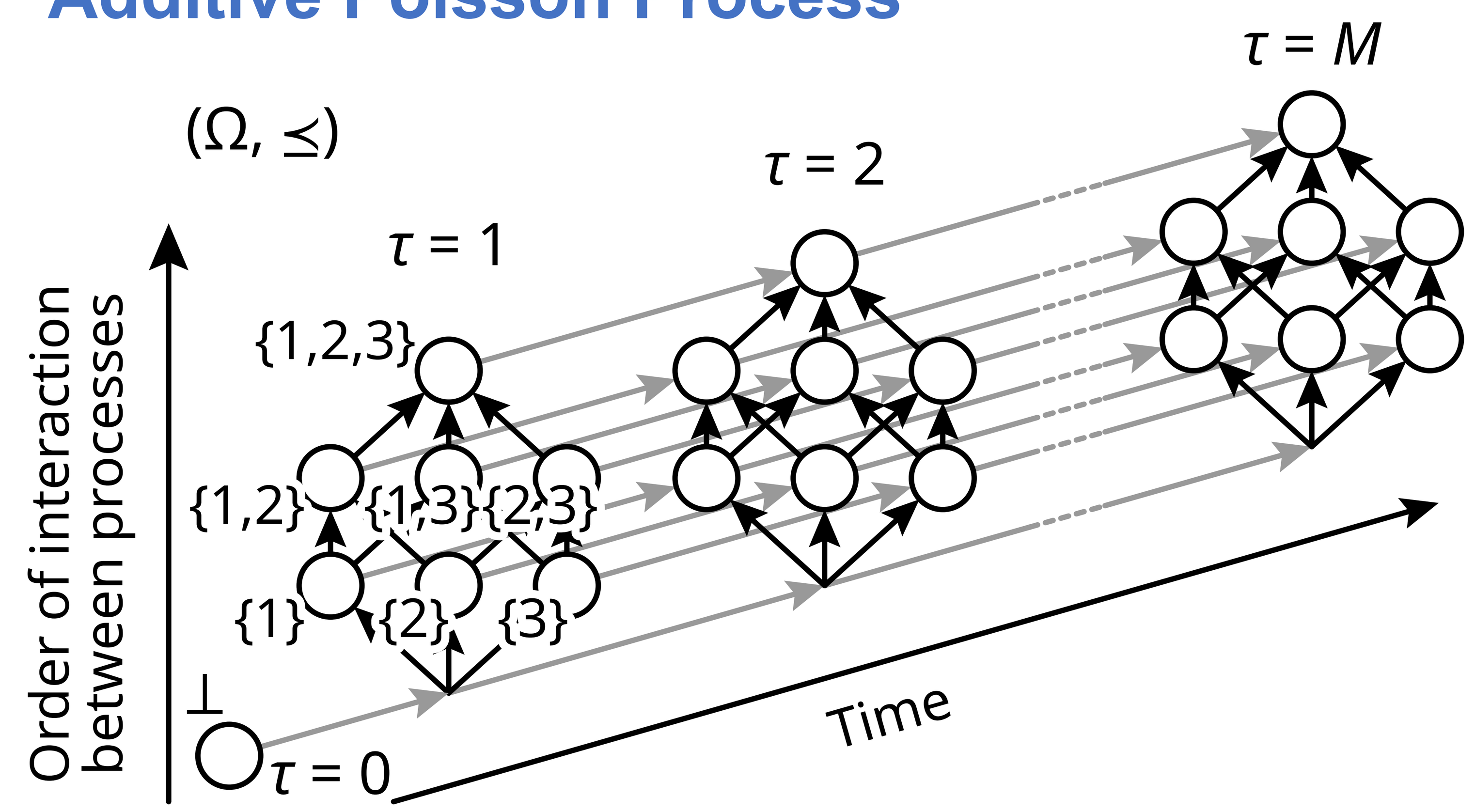
$$\eta_\omega = \sum_{s \in \Omega} \mathbf{1}_{s \succcurlyeq \omega} p(s), \qquad \log p(\omega) = \sum_{s \in \Omega^+} \mathbf{1}_{s \preccurlyeq \omega} \theta_s - \psi(\boldsymbol{\theta}).$$

Solving a convex optimization by minimizing KL divergence with gradient,

$$\frac{\partial}{\partial \theta_s} D_{KL}(\hat{P} \parallel P) = \eta - \hat{\eta} = \Delta\eta$$

Converged once $\frac{\partial}{\partial \theta_s} D_{KL}(\hat{P} \parallel P) = 0$, that means $\eta = \hat{\eta}$

Additive Poisson Process



For a given space,

$$\Omega = \{(J, \tau) \mid J \in 2^{[D]}, \tau \in [M]\},$$

The parameters in the model are

$$S = \{(J, \tau) \in \Omega \mid |J| \le k\}.$$

Rearranging the log-linear model to be the same model used in Kolmogrov-Arnold Representation Theorem

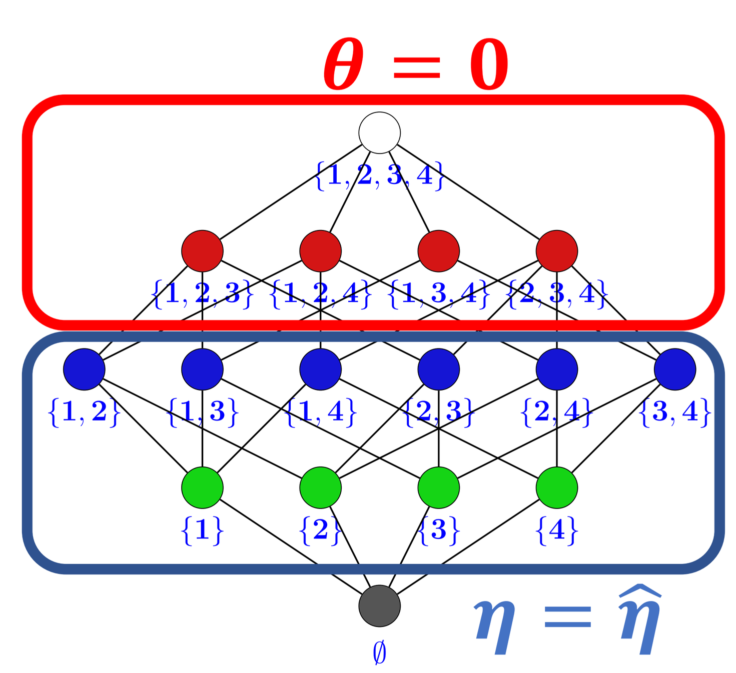
$$p(\omega; \theta) = \frac{1}{\exp\psi(\theta)} \exp\left(\sum_{s \in S} \mathbf{1}_{s \preccurlyeq \omega} \theta_s\right) \propto \exp\left(\sum_{s \in S} \mathbf{1}_{s \preccurlyeq \omega} \theta_s\right).$$

Representing the observations

$$\hat{p}(\omega) = \frac{1}{Z} \sum_{I \subseteq J} \sigma_I(\tau), \qquad Z = \sum_{\omega \in \Omega} \hat{p}(\omega), \qquad \sigma_I(\tau) := \frac{1}{N h_I} \sum_{i=1}^{N} K\left(\frac{\tau^{(I)} - t_i^{(I)}}{h_I}\right).$$

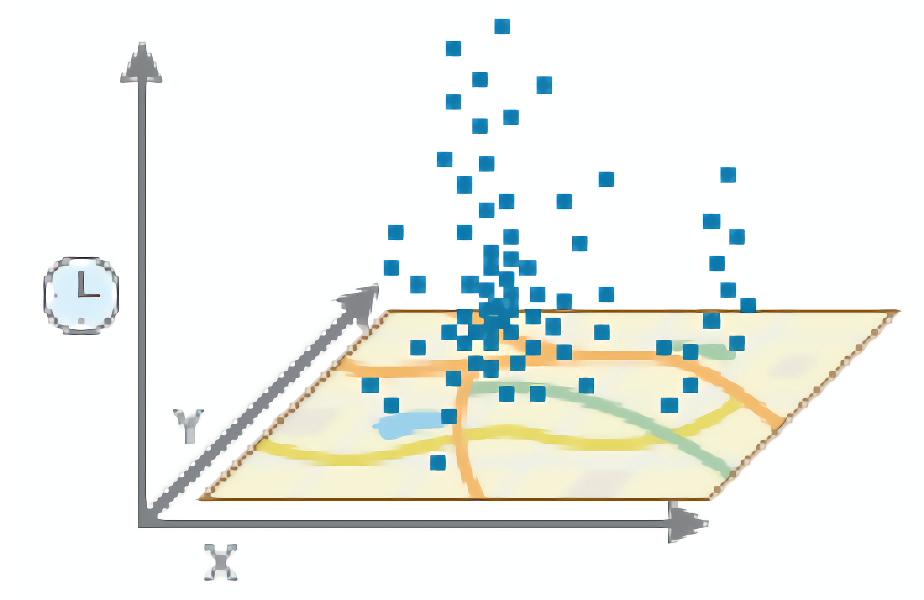
Partial order Structure

$$(J, \tau) \preccurlyeq (J', \tau') \Leftrightarrow J \subseteq J' \quad \text{and} \quad \tau \le \tau'$$

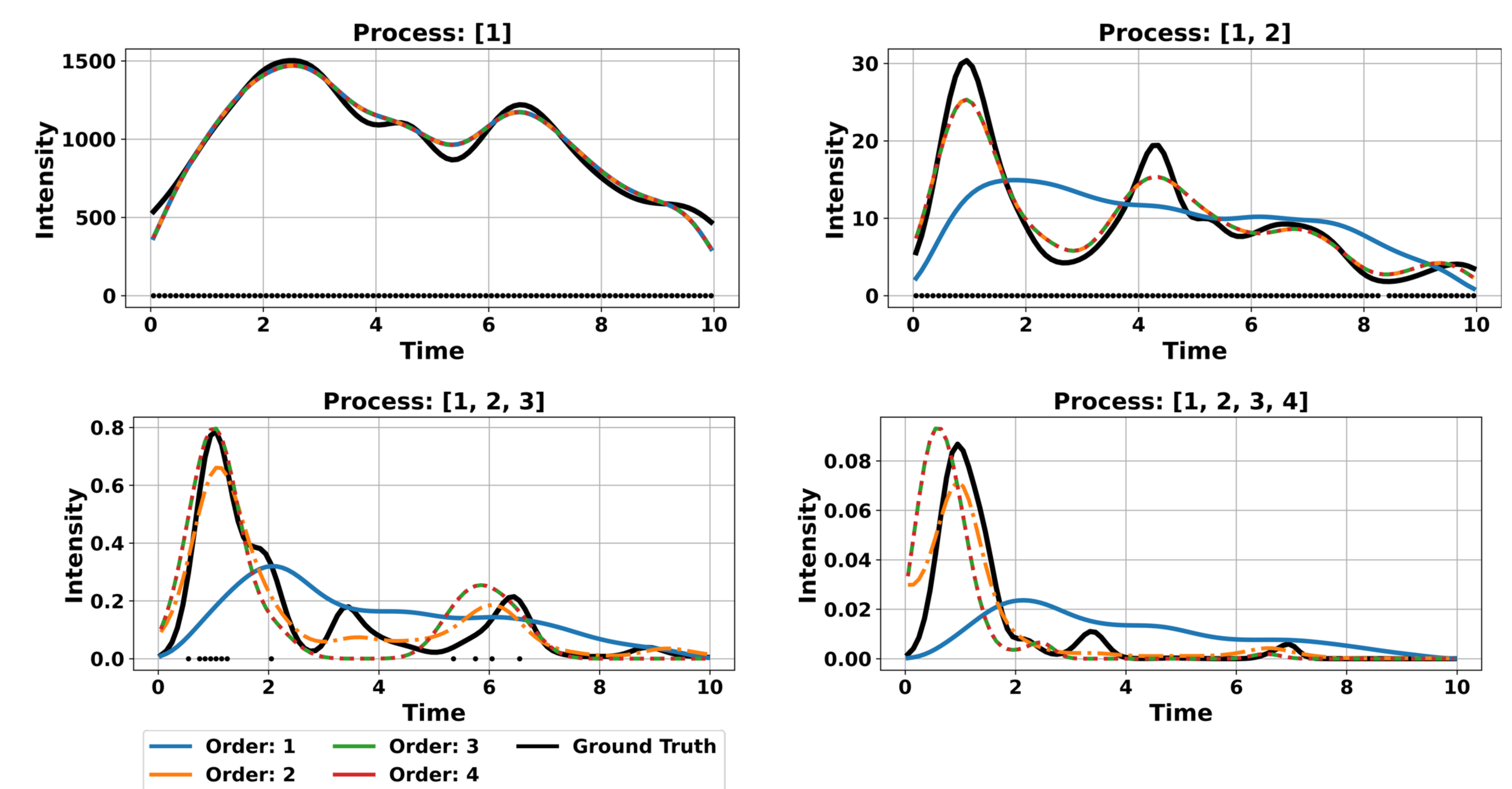


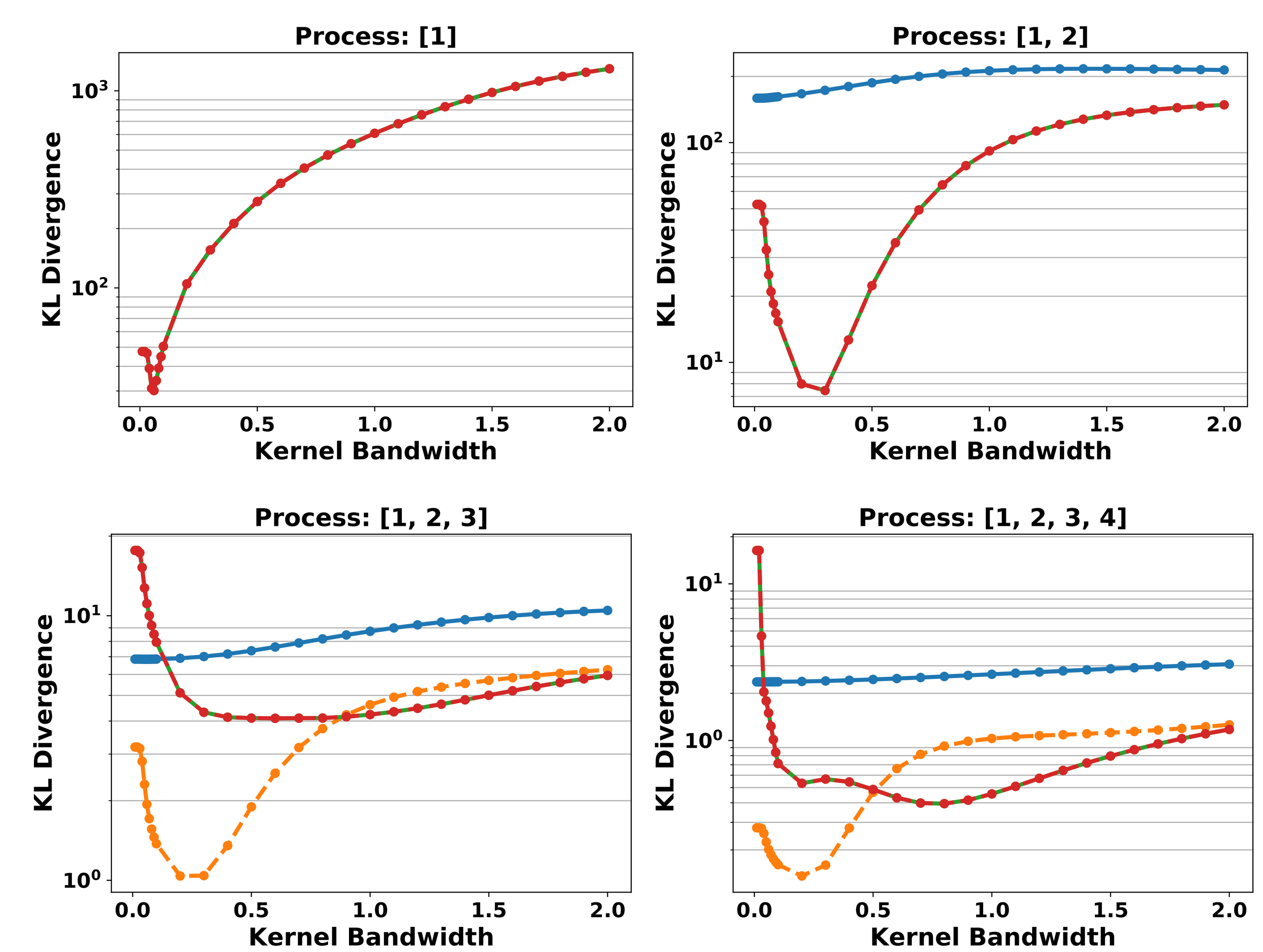
Example for $k = 2$

Example application for 2 spatial dimensions and 1 time dimension



Results





References

- Mahito Sugiyama, Hiroyuki Nakahara, and Koji Tsuda. **Tensor balancing on statistical manifold**, ICML 2017.
- Simon Luo and Mahito Sugiyama, **Bias-variance trade-off in hierarchical probabilistic models using higher-order feature interactions**, AAAI 2019.