Deep Random Splines for Point Process Intensity Estimation of Neural Population Data

Gabriel Loaiza-Ganem[1], Sean M. Perkins[2], Karen E. Schroeder[3], Mark M. Churchland[3] and John P. Cunningham[1]

Department of Statistics[1], Department of Biomedical Engineering[2] and Department of Neuroscience[3], Columbia University

Columbia University in the City of New York

Introduction

- We use random splines to model intensity functions of Poisson Processes.
- We use neural networks to endow splines with rich distributions that avoid some pitfalls of Gaussian Processes (GPs).
- Unlike popular alternatives, we treat time as truly continuous by not discretizing events into bins.
- This allows to recover low-dimensional representations whose dimension is unrelated to the number of time bins (there are no time bins).

Contributions of the paper

- Introducing Deep Random Splines (DRS) as a tool to model random functions.
- Ensuring splines are nonnegative via the method of alternating projections.
- Proposing a Variational AutoEncoder (VAE) with a novel encoder architecture, which recovers low-dimensional latent structure.
- Outperforming alternatives on simulated and both old and new real microelectrode array data.

Acknowledgments and References

We thank the Simons Foundation, Sloan Foundation, McKnight Endowment Fund, NIH NINDS 5R01NS100066, NSF 1707398, and the Gatsby Charitable Foundation for support.

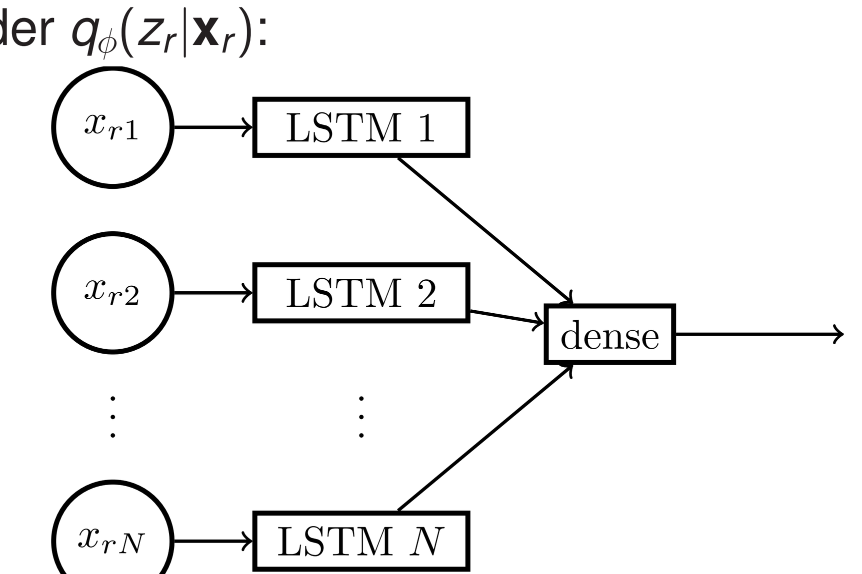
[KW13] Diederik P Kingma and Max Welling, *Auto-encoding variational bayes*, arXiv preprint arXiv:1312.6114 (2013).

Deep Random Splines

- Splines are smooth piecewise polynomial functions of degree d.
- The fixed knot locations $T_1 = t_0 < \cdots < t_I = T_2$ define the pieces and Ψ is the set of parameters of the polynomials in each of those pieces.
- g_ψ denotes a spline on $[T_1, T_2)$ parametrized by $\psi \in \Psi$.
- Simple distributions over Ψ result in oversimplified distributions on functions.
- DRS are a distribution over splines.
- Sampling from a DRS is done by transforming Gaussian noise $Z \in \mathbb{R}^m$ through a neural network $f_\theta : \mathbb{R}^m \to \Psi$, the result being $g_{f_\theta(Z)}$.
- To enforce nonnegativity, we use the method of alternating projections and parameterize splines of degree $2k+1$ with $(Q_1^{(i)}, Q_2^{(i)})_i$ as:
$$p^{(i)}(t) = (t_i - t)[t]^\top Q_1^{(i)}[t] + (t - t_{i-1})[t]^\top Q_2^{(i)}[t]$$
where $[t] = (1, t, \ldots, t^k)^\top$ and $Q_1^{(i)}, Q_2^{(i)} \succeq 0$, for $i = 1, \ldots, I$.

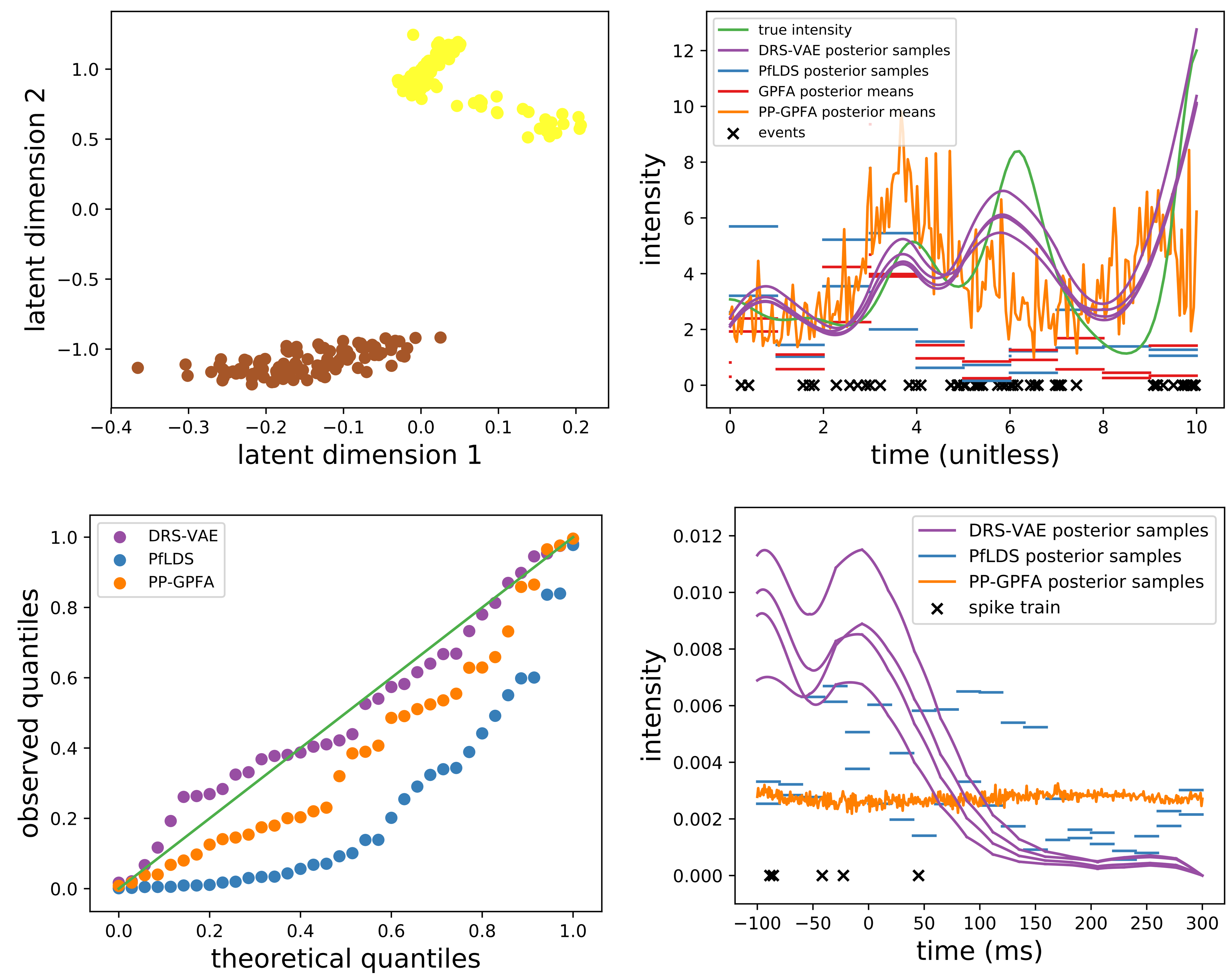
Our Model

- We model each of the R repetitions (trials) of N simultaneous point processes $X_{r,n}$ as:
$$\begin{cases} Z_r \sim \mathcal{N}(0, I_m) \text{ for } r = 1, \ldots, R \\ \psi_{r,n} = f_\theta^{(n)}(Z_r) \text{ for } n = 1, \ldots, N \\ X_{r,n}|\psi_{r,n} \sim \mathcal{PP}_{[T_1,T_2)}(g_{\psi_{r,n}}) \end{cases}$$
- The latent variable is shared across point processes.
- Truly continuous time: latent dimension m does not depend on temporal resolution I.
- We perform inference with a VAE [KW13] , maximizing the ELBO:
$$\sum_{r=1}^{R} E_{q_\phi}[p_\theta(\mathbf{x}_r|z_r)] - KL(q_\phi(z_r|\mathbf{x}_r)||p(z_r))$$
- Computing $p_\theta(\mathbf{x}_r|z_r)$ is tractable due to our spline choice, and we use the following architecture for the encoder $q_\phi(z_r|\mathbf{x}_r)$:



- The input for the encoder are N point processes.
- Each point process is fed to its LSTM and outputs are concatenated at the end.
- Novel encoder allows us to process actual event times (not binned counts).

Experiments



METHOD	SIMULATED DATA ELBO	SIMULATED DATA L^2	SIMULATED DATA p-VALUE	REACHING DATA ELBO	REACHING DATA 15-NN	REACHING DATA SSG/SST	CYCLING DATA ELBO	CYCLING DATA 15-NN	CYCLING DATA SSG/SST
DRS-VAE	**57.1**	$\mathbf{0.11 \pm 0.09}$	—	**−500.8**	**23.7%**	**73.9%**	6372	**55.9%**	**70.0%**
PfLDS	52.3	0.21 ± 0.10	$< 10^{-44}$	−505.7	3.1%	6.2%	**6532**	11.7%	3.2%
GPFA	—	0.21 ± 0.10	$< 10^{-45}$	—	—	—	—	—	—
PP-GPFA	29.0	0.38 ± 0.24	$< 10^{-70}$	−523.2	14.1%	30.5%	6079	51.1%	14.6%

Conclusions

- DRS: rich distribution class over functions and tractable constraints.
- Truly continuous time: better dimensionality reduction.
- We outperform commonly used alternatives.

Email: gl2480@columbia.edu