Variational Generative Modeling of Stochastic Point Processes

Anonymous Author(s)
Affiliation
Address
`email`

Abstract

We consider approximate inference for a class of Cox point processes i.e., point processes with stochastic intensities. Specifically, we consider processes where the Poisson intensity function is modeled as the solution of a stochastic differential equation (SDE). We propose a VAE-like approach where the latent variable is the solution to an SDE, the decoder is fixed (mapping the intensity function to a realization of an inhomogeneous Poisson process), and the encoder maps a point process realization to a posterior path measure corresponding to a diffusion process. Using tools from the theory of *enlargement of filtrations*, we show that the posterior path measure lies in a variational family of SDE path measures. Consequently, evidence lower bound (ELBO) maximization coincides with likelihood maximization. We also introduce hybrid encoder architectures for modeling the drift function of the posterior SDE, conditioned on varying length point process sample paths. Experiments on synthetic data showcase the ability to recover the ground truth measure and highlight the potential of this framework for modeling over-dispersed point processes.

1 Introduction

Variational Autoencoders (VAEs) Dai and Wipf [2019] are highly powerful generative models for Euclidean data, where complex data distributions are typically captured using Gaussian priors and decoders. Many real-world phenomena, however, are more naturally represented as point processes: collection of event times in continuous time, which arise in settings such as queueing systems, neural spikes, and many others.

A particularly flexible class of models is called Cox processes (doubly stochastic Poisson processes), where the intensity itself evolves according to a stochastic process. This formulation accommodates over-dispersion and better reflects uncertainty compared to homogeneous or inhomogeneous Poisson processes with deterministic intensities.

In this work, we propose a VAE-like framework for point processes using Cox generative models, for generative and predictive tasks. Unlike classical VAEs, where the variational family typically excludes the true posterior, our construction admits the true posterior as a member of the variational class, ensuring maximization of the log-likelihood via maximizing ELBO. We further introduce several neural architectures for parametrizing the posterior dynamics, allowing trade-offs between model flexibility, inference accuracy, and computational efficiency.

Submitted to 39th Conference on Neural Information Processing Systems (NeurIPS 2025). Do not distribute.

2 Background and Notations

Throughout the paper, the central object of interest is a one-dimensional[1] point process on a compact interval $[0, T]$, which can either be represented either by its event times $\{\tau_1, \tau_2, ..., \tau_{N_T}\}$ or equivalently as a counting process

$$N_t = \sum_{i=1}^{N_T} \mathbb{1}[\tau_i \leq t].$$

In particular, our focus is on Cox processes (doubly stochastic Poisson processes), in which the underlying (nonnegative) intensity is driven by some stochastic differential equation

$$dZ_t = b(Z_t, t)\,dt + \sigma(Z_t, t)\,dB_t, \tag{1}$$

and the process $N_{0:T} := (N_t)_{0 \leq t \leq T}$ is an inhomogeneous Poisson process with rate Z_t conditioned on a sample path of Z. Recall that an inhomogeneous Poisson process is characterized by the following four properties:

i. $N_0 = 0$;

ii. $N_t - N_s \perp\!\!\!\perp N_v - N_u$ if $s < t \leq u < v$ (independent increments);

iii. $\mathbb{P}(N_{t+h} - N_t = 1) = Z_t h + o(h)$;

iv. $\mathbb{P}(N_{t+h} - N_t \geq 2) = o(h)$.

To describe parametrized models, we write b_θ and u_β for neural networks with parameters θ and β. The prior SDE for the intensity is

$$dZ_t = b_\theta(Z_t, t)\,dt + \sigma(Z_t, t)\,dB_t, \tag{2}$$

which induces a likelihood $P_\theta(N_{0:T})$ for the observed point process. For variational inference, we introduce a controlled SDE representing the approximate posterior:

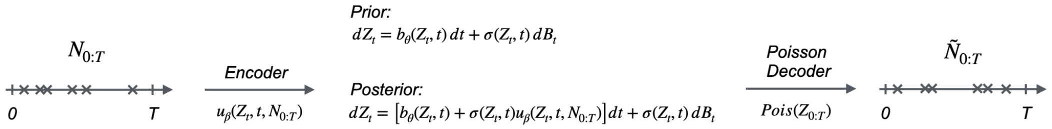
$$dZ_t = \Big[b_\theta(Z_t, t) + \sigma(Z_t, t) u_\beta\left(Z_t, t, \{\tau_i > t\}\right)\Big]dt + \sigma(Z_t, t)\,dB_t, \tag{3}$$

where $\{\tau_i > t\}$ denotes the set of future event times. We use $P_\phi(N_{0:T})$ to denote the corresponding likelihood, with $\phi = (\theta, \beta)$.

3 Methodology

3.1 The VAE structure and Evidence Lower Bound

Similar to a traditional Gaussian VAE, we have an encoder-prior-decoder structure (Fig. 1).



$N_{0:T}$
0
T
Encoder
$u_\beta(Z_t, t, N_{0:T})$
Prior:
$dZ_t = b_\theta(Z_t, t)\,dt + \sigma(Z_t, t)\,dB_t$
Posterior:
$dZ_t = \left[b_\theta(Z_t, t) + \sigma(Z_t, t)u_\beta(Z_t, t, N_{0:T})\right]dt + \sigma(Z_t, t)\,dB_t$
Poisson
Decoder
$Pois(Z_{0:T})$
$\tilde{N}_{0:T}$
0
T

Figure 1: An illustration of the Point Process VAE structure.

We can also derive the Evidence Lower Bound (ELBO):

$$\log P_\theta(N_{0:T}) \geq \mathbb{E}_{Q_{\phi, N_{0:T}}}\left[\log P(N_{0:T}|Z) - \frac{1}{2}\int u_\beta^2(Z_t, t, N_{0:T})\,dt\right] =: \text{ELBO}; \tag{4}$$

see Wang et al. [2021] for a detailed derivation.

[1] This can be naturally generalized to n-dimensional point process for modeling several queues.

3.2 Theoretical Foundations

A natural concern is whether the class of approximate posteriors in our framework is rich enough to contain the true posterior, since this is not the case in classical Gaussian VAEs, where maximization of the ELBO does not imply maximization of the log-likelihood. The good news is, it can be shown that the true posterior intensity can indeed be described via an SDE, with an extra control term in the drift. The tools used to show this come from a branch probability theory called *Enlargement of Filtrations*. Using Jacod's absolute continuity condition (condition (A) in Jacod [2006]), we can prove the following theorem:

Theorem 3.1. *Let $N_{0:T}$ be a one-dimensional point process on the interval $[0, T]$, and let the prior intensity process be described by some SDE*

$$dZ_t = b(Z_t, t)\, dt + \sigma(Z_t, t)\, dB_t,$$

then the "posterior" intensity process conditioned on $N_{0:T}$ can be described via

$$dZ_t = \big[b(Z_t, t) + \sigma(Z_t, t)^2 h(Z_t, t, \{\tau_i > t\})\big] dt + \sigma(Z_t, t)\, dB_t,$$

where

$$h(Z_t, t, \{\tau_i > t\}) := \frac{\partial}{\partial Z_t} \log \mathbb{E}\left[\exp\left(-\int_t^T Z_s\, ds\right) \prod_{\tau_i > t} Z_{\tau_i} \middle| Z_t\right].$$

This result justifies our assumption: the posterior intensity lies in the same family of SDEs as the prior, but with an additional term. Thus, the encoder u_β can be interpreted as learning this drift correction, namely, $\sigma(Z_t, t)h(Z_t, t, \{\tau_i > t\})$.

3.3 Architectural Choices

While the prior drift b_θ only depends on (Z_t, t) and can be modeled with a simple MLP, the posterior drift correction u_β depends on the entire event sequence $\{\tau_i\}$, making it significantly harder to model. To this end, we consider three progressively more expressive architectures for u_β:

1. **[remaining count]** A fully connected neural network with inputs Z_t, t, and $N_T - N_t$;
2. **[count vector]** A fully connected neural network with inputs Z_t, t, and the binned counts $(N_{T/k}, N_{2T/k} - N_{T/k}, ..., N_T - N_{(k-1)T/k})$;
3. **[deep set]** A Deep Sets architecture with boundary conditions enforced.

With the first architecture, all future information is summarized by a single number—the remaining number of events. This mode suffices for learning the prior, but consistently fails to approximate the true posterior dynamics. The second architecture seeks to incorporate all the information of the event sequence by taking the binned counts of the events as inputs. This improves over the first method, but not by much. In particular, u_β still can't learn the posterior dynamics. Inspired by Zaheer et al. [2018], we incorporate a Deep Sets architecture which takes a varying length of inputs together with strictly enforced boundary constraints. Analytically, we know that h satisfies the following properties:

i. for any event time τ,

$$\lim_{s \nearrow \tau} h - \lim_{s \searrow \tau} h = \frac{1}{Z_\tau}$$

(discrete jumps at event times);

ii.

$$h = \mathcal{O}(t - T).$$

Furthermore, u_β needs to be invariant to the order of the arrival times input $\{\tau_i\}$. Therefore, we designed the neural network u_β to take the following form:

$$u_\beta(Z_t, t, \{\tau_i > t\}) = \sigma(Z_t, t) \left(\frac{N_T - N_t}{Z_t} + (t - T)\Big[1 + (T - t)\, \mathit{NN}(Z_t, t, \{\tau_i > t\})\Big] \right),$$

where $\mathit{NN}(Z_t, t, \{\tau_i > t\})$ uses a Deep Sets architecture, i.e.,

$$\mathit{NN}(Z_t, t, \{\tau_i > t\}) = \rho \left(Z_t, t, \sum_{\tau_i > t} \psi(\tau_i) \right)$$

with ρ and ψ being fully connected neural networks.

3.4 Training

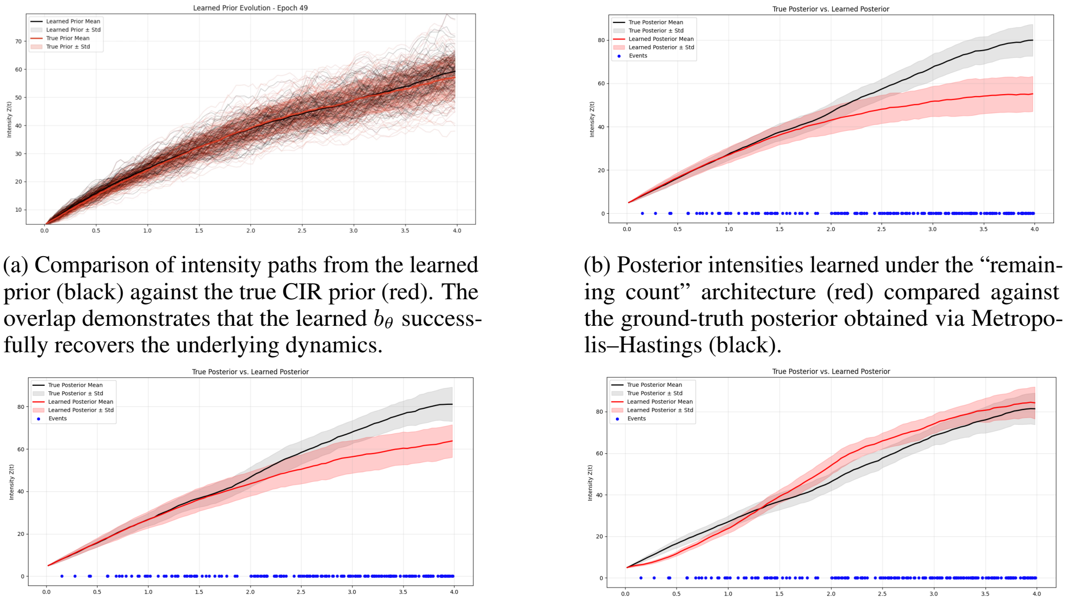
We discretize the SDE using Euler–Maruyama and apply a pathwise reparameterization trick, analogous to the Gaussian VAE. Gradients of both b_θ and u_β are estimated via Monte Carlo with mini-batch size of 16. Optimization is carried out using Adam with learning rate 0.005.

4 Numerical Experiments

We use CIR intensity as the ground truth, since it is a positive-valued process, and is widely used model in finance/queueing We generated 128 samples $\{N^i_{0:T}\}_{i=1}^{128}$ of Cox processes on the interval $[0, 4]$ with CIR intensity:

$$dZ_t = 0.3(80 - Z_t)\,dt + \sqrt{Z_t}\,dB_t, \text{ and } Z_0 = 5. \tag{5}$$

Only $\{N^i_{0:T}\}_{i=1}^{128}$ is used towards training, while the true prior remains unknown to the model. Fig. 2a below shows a comparison between intensity paths generated using the learned prior and using the true prior (Deep Sets architecture is used for u_β). To show the power of Deep Sets architecture in learning the correct posteriors, we also compare the learned posterior intensities in all three modes against the "true posterior" intensities generated via Metropolis-Hastings. For reference, we choose the point process sample with the most number of events (most extreme) for posterior conditioning. The event times are marked as blue dots on the time axis.



(a) Comparison of intensity paths from the learned prior (black) against the true CIR prior (red). The overlap demonstrates that the learned b_θ successfully recovers the underlying dynamics.

(b) Posterior intensities learned under the "remaining count" architecture (red) compared against the ground-truth posterior obtained via Metropolis–Hastings (black).

(c) Posterior intensities learned under the "count vector" (binned counts) architecture (red) compared against the ground-truth posterior (black).

(d) Posterior intensities learned using the Deep Sets architecture with boundary constraints (red) compared against the ground-truth posterior (black).

Figure 2

These results confirm that simple encoders (remaining count, binned counts) cannot capture posterior dynamics, while the Deep Sets architecture obtains close agreement with true posterior intensities.

5 Conclusion

We establish a variational framework for generative modeling of point processes, demonstrating the importance of structured neural encoders for posterior inference. Our results highlight the value of combining modern ML architectures with classical tools from probability and operations research, particularly for uncertainty quantification in event-driven systems. Future directions include multivariate extensions and applications to event-driven decision-making systems (with marked Poisson process).

References

B. Dai and D. Wipf. Diagnosing and enhancing vae models, 2019. URL https://arxiv.org/abs/1903.05789.

J. Jacod. Grossissement initial, hypothese (H') et theoreme de Girsanov. In *Grossissements de filtrations: exemples et applications: Séminaire de Calcul Stochastique 1982/83 Université Paris VI*, pages 15–35. Springer, 2006.

R. Wang, P. Jaiswal, and H. Honnappa. Estimating stochastic poisson intensities using deep latent models. In *Proceedings of the Winter Simulation Conference*, WSC '20, page 596–607. IEEE Press, 2021. ISBN 9781728194998.

M. Zaheer, S. Kottur, S. Ravanbakhsh, B. Poczos, R. Salakhutdinov, and A. Smola. Deep sets, 2018. URL https://arxiv.org/abs/1703.06114.