Intensity-free Learning of Temporal Point Processes

Oleksandr Shchur*, Marin Biloš*, Stephan Günnemann

Technical University of Munich, Germany

ICLR 2020

* Equal contribution

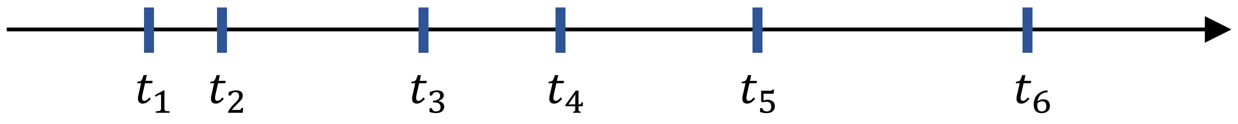
Temporal Point Processes (TPP)

Discrete events in continuous time

Hospital visits

Purchases and transactions

Social media posts ...



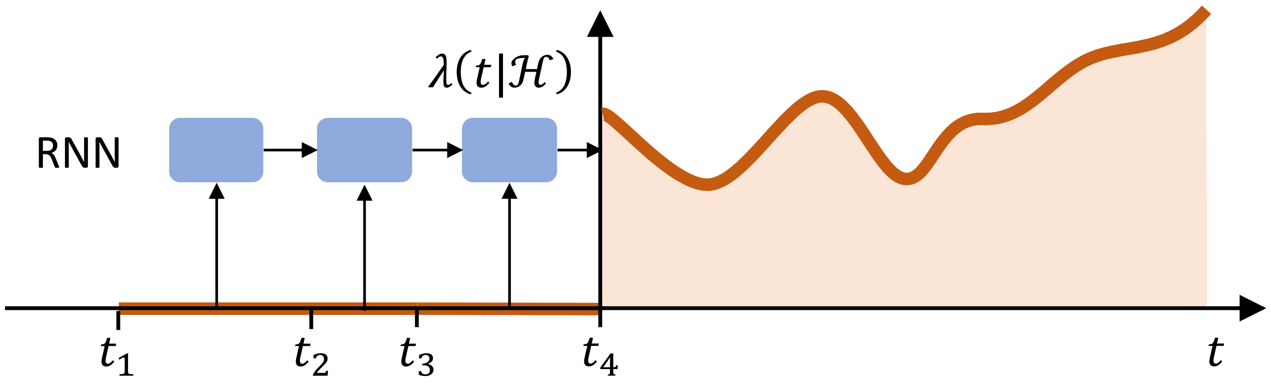
Prediction task

- When will the next event occur?

Conditional intensity function

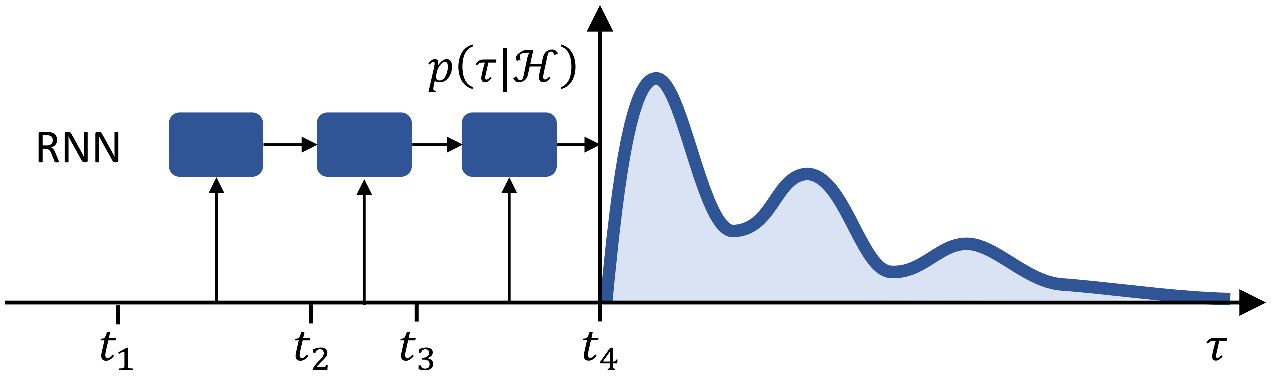
- A TPP is uniquely defined by its conditional intensity $\lambda(t_i|\mathcal{H}_{t_i})$
  - $\lambda(t_i|\mathcal{H}_{t_i})$ defines the rate of event occurrence given the history $\mathcal{H}_{t_i} = \{t_1, \dots, t_{i-1}\}$
- Likelihood

$$\sum_i \log p(t_i|\mathcal{H}_{t_i}) = \sum_i \log \lambda(t_i|\mathcal{H}_{t_i}) - \int_0^T \lambda(t|\mathcal{H}_t) dt$$



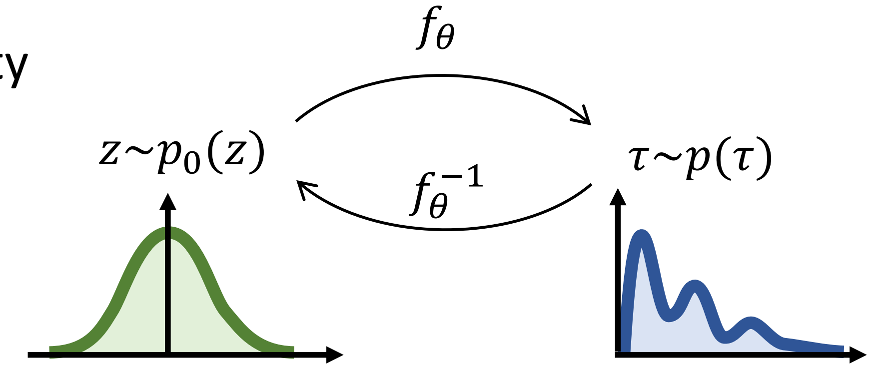
Conditional **density** estimation

- Limitations of modeling the conditional intensity $\lambda(t_i|\mathcal{H}_{t_i})$
  - Often limited flexibility
  - Computing likelihood requires integration
  - No closed-form sampling & expectation
- Solution – Model the **conditional density** $p(\tau|\mathcal{H})$
  - We can use existing tools for neural density estimation!

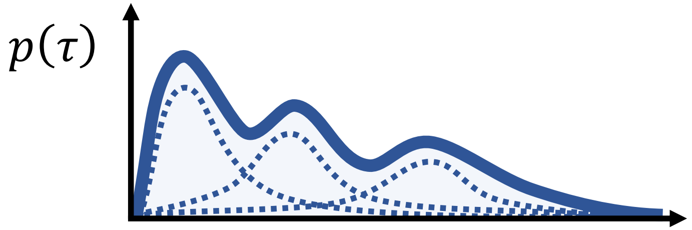


How can we model $p(\tau|\mathcal{H})$?

- Normalizing flows
  - Get flexible distribution by transforming a simple density
  - $p(\tau) = p_0\left(f_\theta^{-1}(\tau)\right)\left|\frac{d}{d\tau} f_\theta^{-1}(\tau)\right|$



- Mixture distribution
  - Convex combination of simple densities
  - $p(\tau) = \sum_k \pi_k \, p_0(\tau|\theta_k)$

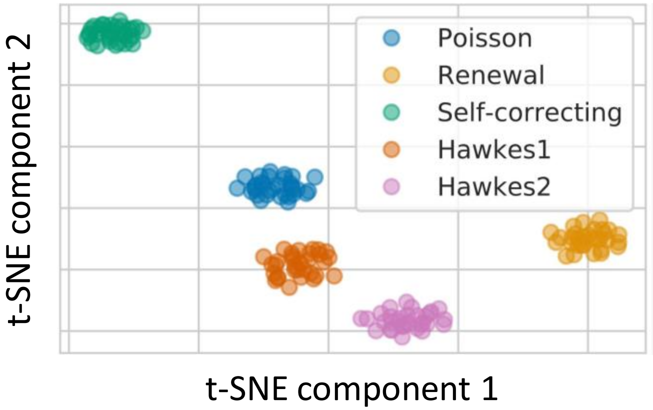


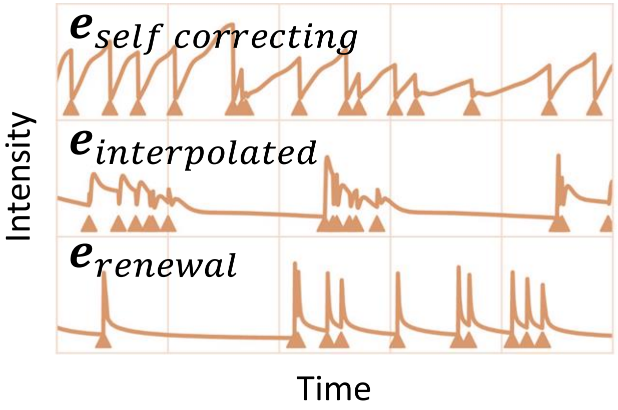
- The parameters θ are produced by an RNN that encodes $\mathcal{H}$

Results

- Tractable, flexible and efficient
  - SOTA results on density estimation
  - Sampling, likelihood, mean – all in closed form

- New possibilities
  - Sequence embedding
  - Conditional generation
  - Training with missing data





Code and datasets: www.daml.in.tum.de/intensity-free-tpp