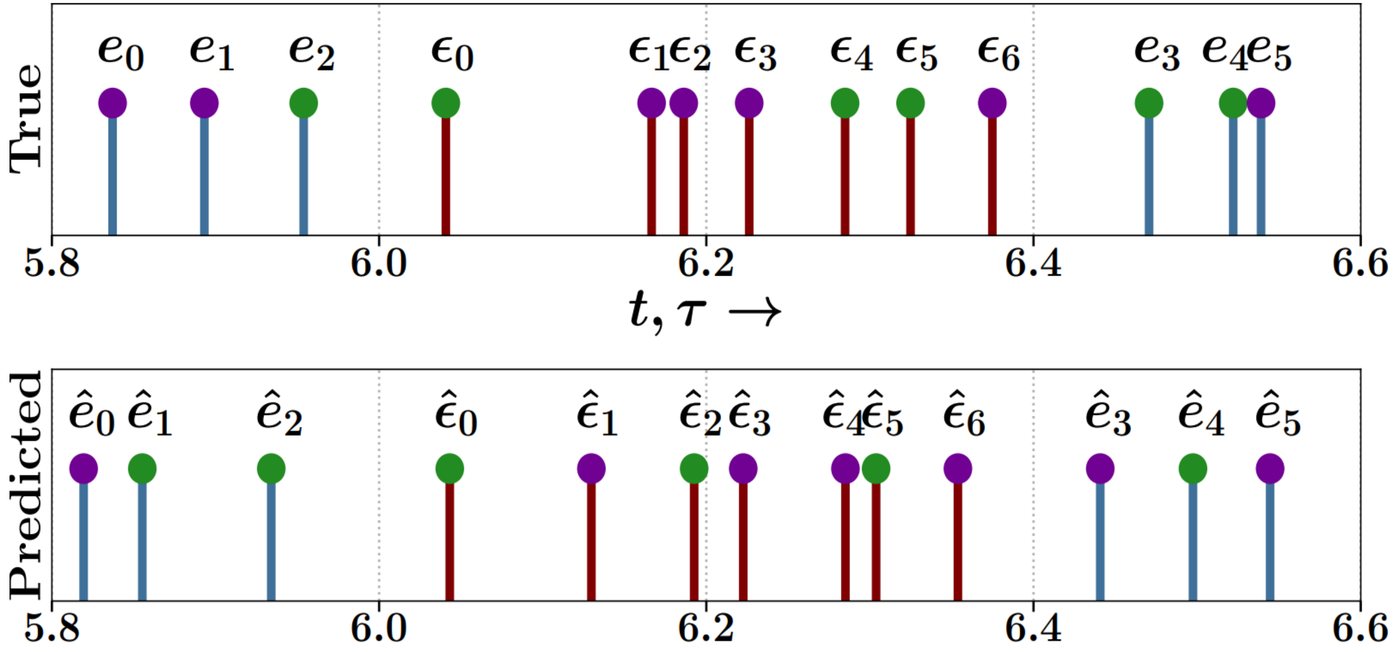
Learning Temporal Point Processes with Intermittent Observations



Presenter: Vinayak Gupta

Vinayak Gupta[1], Srikanta Bedathur[1], Sourangshu Bhattacharya[2], and Abir De[3]

IIT Delhi[1], IIT Kharagpur[2], IIT Bombay[3]

Data Streams in Real World

Stream of events occurring in continuous timestamps.

- **Frequently observed in:**
  - Social Networks (posts and comments).
  - Healthcare (hospital visits, checkups).
  - Finance (stock prices and market trends).
  - Online Shopping (purchases, reviews) and many more ...

Gupta et al. Learning Temporal Point Processes with Intermittent Observations.

Data Streams in Real World

Stream of events occurring in continuous timestamps.

- **Frequently observed in:**
  - Social Networks (posts and comments).
  - Healthcare (hospital visits, checkups).
  - Finance (stock prices and market trends).
  - Online Shopping (purchases, reviews) and many more ...

- How to model them?
  - **Temporal Point Processes:** $\mathcal{S}_k = \{e_i = (x_i, t_i) | t_i < t_{i+1}\}$

Gupta et al. Learning Temporal Point Processes with Intermittent Observations.

Data Streams in Real World

Stream of events occurring in continuous timestamps.

- **Frequently observed in:**
  - Social Networks (posts and comments).
  - Healthcare (hospital visits, checkups).
  - Finance (stock prices and market trends).
  - Online Shopping (purchases, reviews) and many more ...

- How to model them?
  - **Temporal Point Processes:**

$$\mathcal{S}_k = \{e_i = (\underbrace{x_i}_{\text{Discrete Mark}}, \overbrace{t_i}^{\text{Continuous Time}}) | t_i < t_{i+1}\}$$

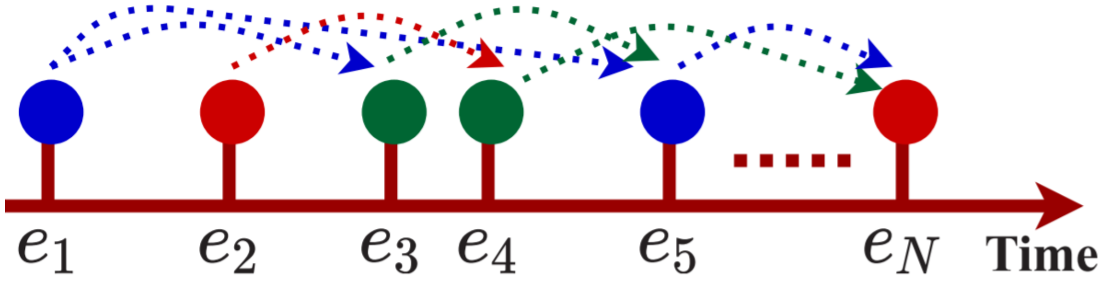
Gupta et al. Learning Temporal Point Processes with Intermittent Observations.

Data Streams in Real World

Stream of events occurring in continuous timestamps.

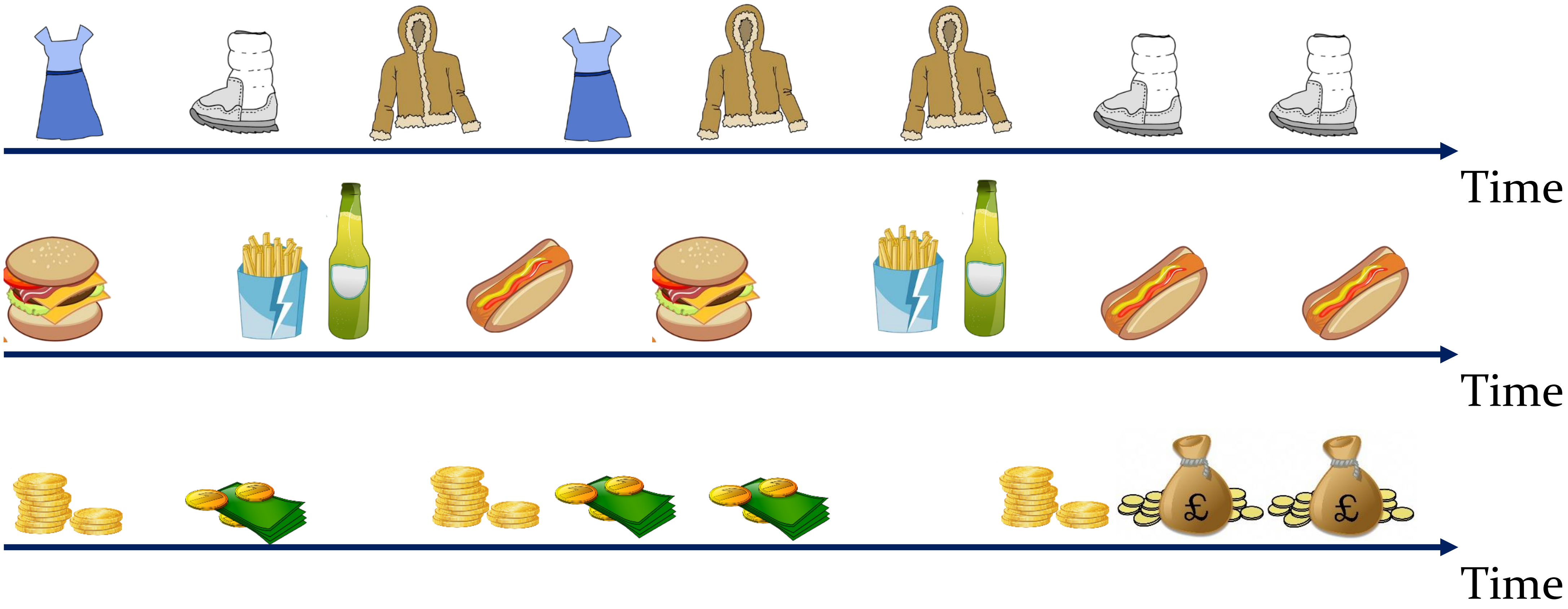
- **Frequently observed in:**
  - Social Networks (posts and comments).
  - Healthcare (hospital visits, checkups).
  - Finance (stock prices and market trends).
  - Online Shopping (purchases, reviews) and many more ...

- How to model them?
  - **Temporal Point Processes:** $\mathcal{S}_k = \{e_i = (x_i, t_i) | t_i < t_{i+1}\}$



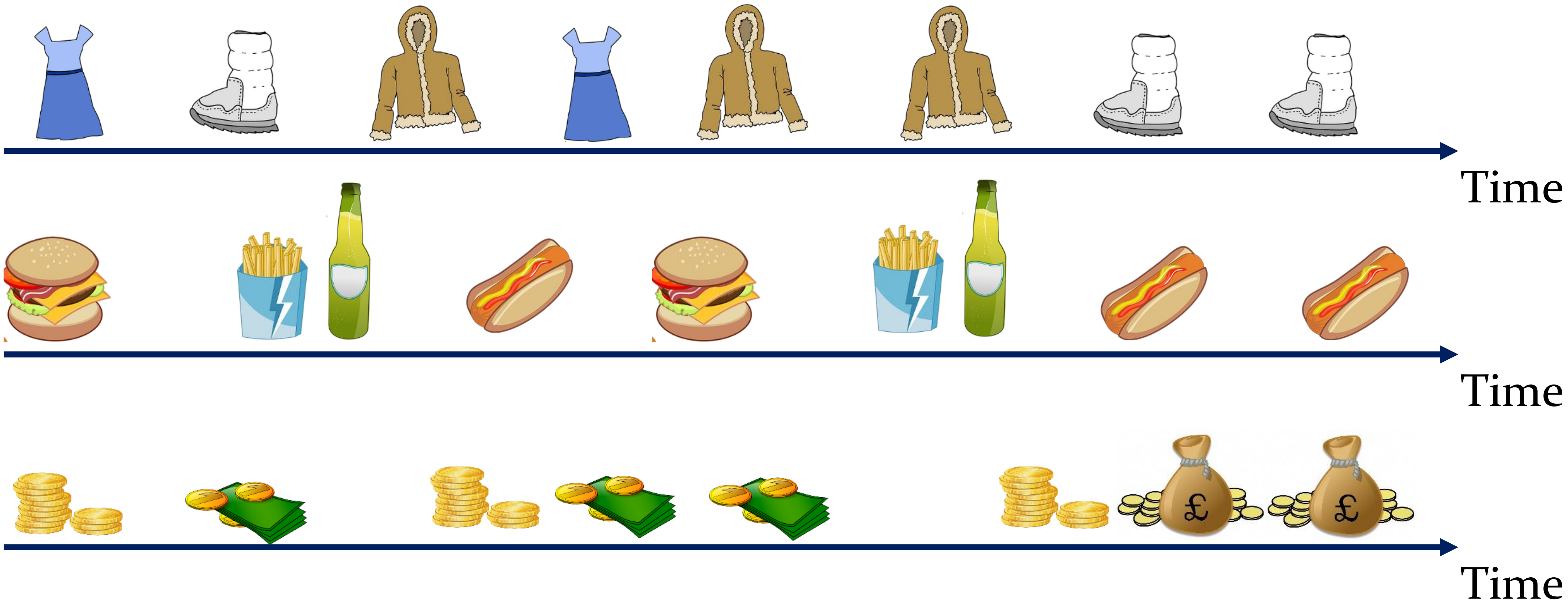
Gupta et al. Learning Temporal Point Processes with Intermittent Observations.

Real World Data



Gupta et al. Learning Temporal Point Processes with Intermittent Observations.

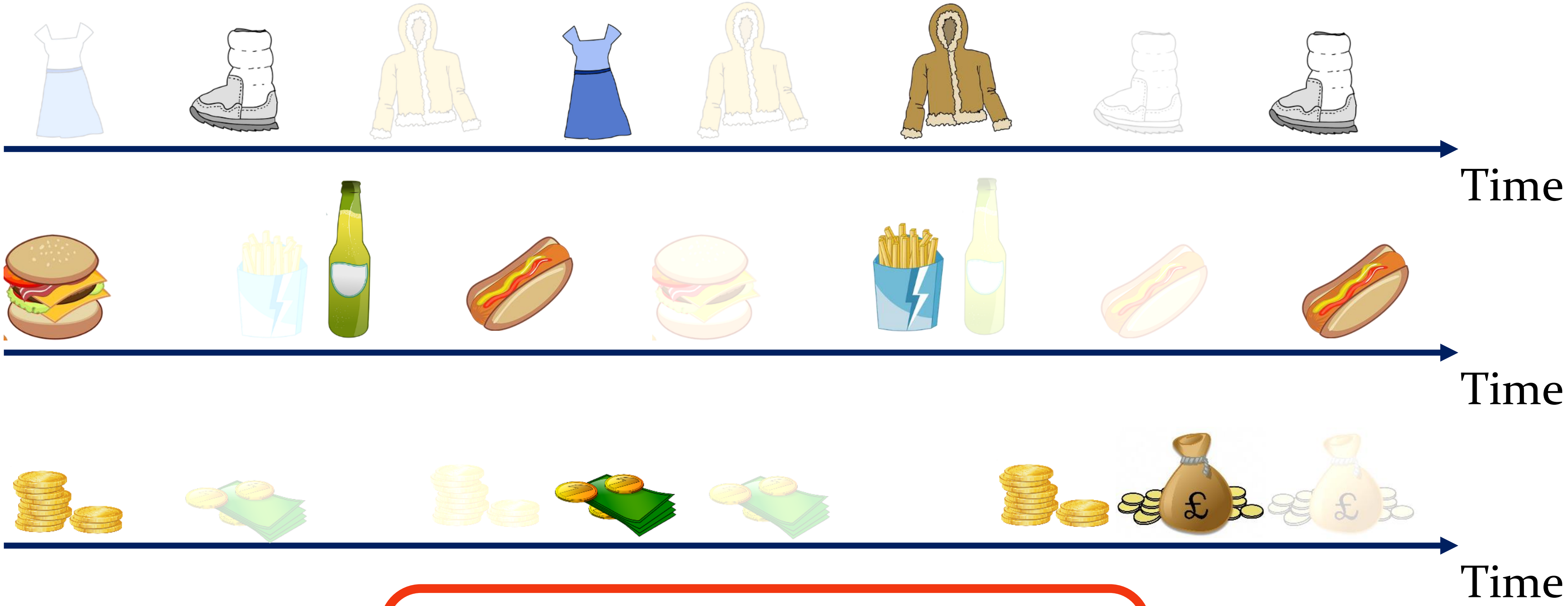
Real World Data



Existing sequential models assume a complete observation scenario. i.e. the event sequence is completely *observed* with no *missing* events. **Rarely Encountered!**

Gupta et al. Learning Temporal Point Processes with Intermittent Observations.

Real World Data



Some reasons for missing data:

1. Different payment procedure.
2. User-awareness for privacy.
3. Privacy restrictions for data collection.

Gupta et al. Learning Temporal Point Processes with Intermittent Observations.

Proposed Model: IMTPP

The proposed model, **IMTPP** (**I**ntermittently-observed **M**arked **T**emporal **P**oint **P**rocesses) learns the dynamics of both observed and missing events:

- ✓ *Coupled* MTPPs for observed and missing events.
- ✓ Missing events as latent variables.
- ✓ Incrementally update on a new event.

Gupta et al. Learning Temporal Point Processes with Intermittent Observations.

Proposed Model: IMTPP

The proposed model, **IMTPP** (**I**ntermittently-observed **M**arked **T**emporal **P**oint **P**rocesses) learns the dynamics of both observed and missing events:

- ✓ *Coupled* MTPPs for observed and missing events.
- ✓ Missing events as latent variables.
- ✓ Incrementally update on a new event.

Notations:

- Sequence of *observed* events. $\mathcal{S}_k = \{e_i = (x_i, t_i) | i \in [k], t_i < t_{i+1}\}$
- Generated *missing* events: $\mathcal{M}_r = \{\epsilon_j = (y_j, \tau_j) | j \in [k], \tau_j < \tau_{j+1}\}$

Gupta et al. Learning Temporal Point Processes with Intermittent Observations.

Proposed Model: IMTPP

The proposed model, **IMTPP** (**I**ntermittently-observed **M**arked **T**emporal **P**oint **P**rocesses) learns the dynamics of both observed and missing events:

- ✓ *Coupled* MTPPs for observed and missing events.
- ✓ Missing events as latent variables.
- ✓ Incrementally update on a new event.

Notations:

- Sequence of *observed* events. $\mathcal{S}_k = \{e_i = (x_i, t_i) | i \in [k], t_i < t_{i+1}\}$
- Generated *missing* events: $\mathcal{M}_r = \{\epsilon_j = (y_j, \tau_j) | j \in [k], \tau_j < \tau_{j+1}\}$
- **Observed TPP:** $p(e_{k+1} | \mathcal{S}_k, \mathcal{M}_{\bar{k}})$
- **Missing TPP Prior:** $p(\epsilon_r | \mathcal{S}_k, \mathcal{M}_{r-1})$
- **Missing TPP Posterior:** $q(\epsilon_{r+1} | e_{k+1}, \mathcal{S}_k, \mathcal{M}_{r-1})$

Gupta et al. Learning Temporal Point Processes with Intermittent Observations.

Proposed Model: IMTPP

The proposed model, **IMTPP** (**I**ntermittently-observed **M**arked **T**emporal **P**oint **P**rocesses) learns the dynamics of both observed and missing events:

✓ *Coupled* MTPPs for observed and missing events.

✓ Missing events as latent variables.

✓ Incrementally update on a new event.

Notations:

- Sequence of *observed* events. $\mathcal{S}_k = \{e_i = (x_i, t_i) | i \in [k], t_i < t_{i+1}\}$
- Generated *missing* events: $\mathcal{M}_r = \{\epsilon_j = (y_j, \tau_j) | j \in [k], \tau_j < \tau_{j+1}\}$
- **Observed TPP:** $p(e_{k+1} | \mathcal{S}_k, \mathcal{M}_{\bar{k}})$
- **Missing TPP Prior:** $p(\epsilon_r | \mathcal{S}_k, \mathcal{M}_{r-1})$
- **Missing TPP Posterior:** $q(\epsilon_{r+1} | e_{k+1}, \mathcal{S}_k, \mathcal{M}_{r-1})$

Gupta et al. Learning Temporal Point Processes with Intermittent Observations.

Proposed Model: IMTPP

The proposed model, **IMTPP** (**I**ntermittently-observed **M**arked **T**emporal **P**oint **P**rocesses) learns the dynamics of both observed and missing events:

- ✓ *Coupled* MTPPs for observed and missing events.
- ✓ Missing events as latent variables.
- ✓ Incrementally update on a new event.

Notations:

- Sequence of *observed* events. $\mathcal{S}_k = \{e_i = (x_i, t_i) | i \in [k], t_i < t_{i+1}\}$
- Generated *missing* events: $\mathcal{M}_r = \{\epsilon_j = (y_j, \tau_j) | j \in [k], \tau_j < \tau_{j+1}\}$
- **Observed TPP:** $p(e_{k+1} | \mathcal{S}_k, \mathcal{M}_{\bar{k}})$
- **Missing TPP Prior:** $p(\epsilon_r | \mathcal{S}_k, \mathcal{M}_{r-1})$
- **Missing TPP Posterior:** $q(\epsilon_{r+1} | e_{k+1}, \mathcal{S}_k, \mathcal{M}_{r-1})$

Gupta et al. Learning Temporal Point Processes with Intermittent Observations.

Proposed Model: IMTPP

The proposed model, **IMTPP** (**I**ntermittently-observed **M**arked **T**emporal **P**oint **P**rocesses) learns the dynamics of both observed and missing events:

- ✓ *Coupled* MTPPs for observed and missing events.
- ✓ Missing events as latent variables.
- ✓ Incrementally update on a new event.

Notations:

- Sequence of *observed* events. $\mathcal{S}_k = \{e_i = (x_i, t_i) | i \in [k], t_i < t_{i+1}\}$
- Generated *missing* events: $\mathcal{M}_r = \{\epsilon_j = (y_j, \tau_j) | j \in [k], \tau_j < \tau_{j+1}\}$
- **Observed TPP:** $p(e_{k+1} | \mathcal{S}_k, \mathcal{M}_{\bar{k}})$
- **Missing TPP Prior:** $p(\epsilon_r | \mathcal{S}_k, \mathcal{M}_{r-1})$
- **Missing TPP Posterior:** $q(\epsilon_{r+1} | e_{k+1}, \mathcal{S}_k, \mathcal{M}_{r-1})$

Gupta et al. Learning Temporal Point Processes with Intermittent Observations.

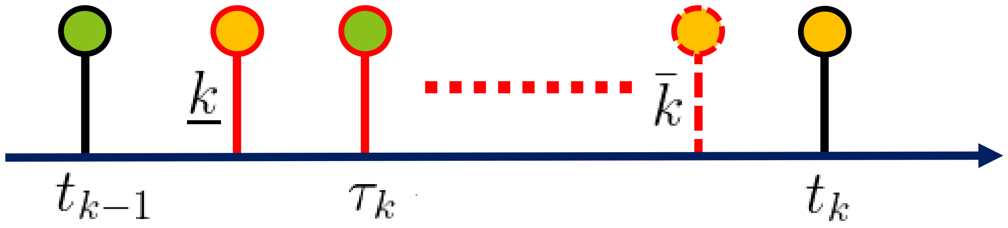
Proposed Model: IMTPP

The proposed model, **IMTPP** (**I**ntermittently-observed **M**arked **T**emporal **P**oint **P**rocesses) learns the dynamics of both observed and missing events:

✓ *Coupled* MTPPs for observed and missing events.

✓ Missing events as latent variables.

✓ Incrementally update on a new event.

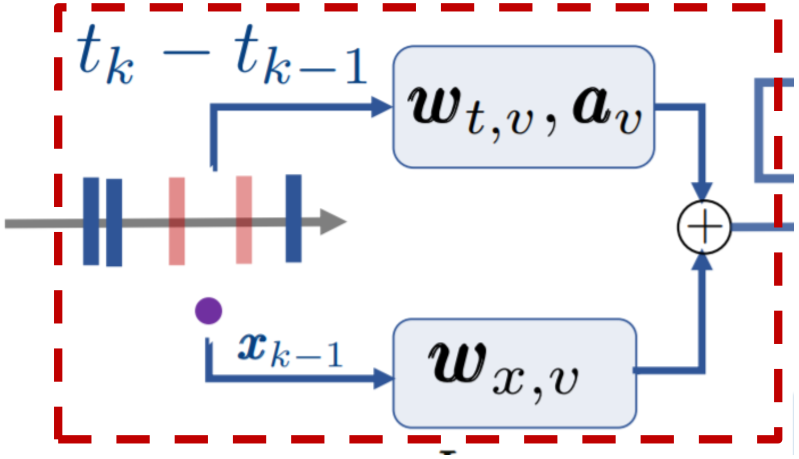


Notations:

- Sequence of *observed* events. $\mathcal{S}_k = \{e_i = (x_i, t_i) | i \in [k], t_i < t_{i+1}\}$
- Generated *missing* events: $\mathcal{M}_r = \{\epsilon_j = (y_j, \tau_j) | j \in [k], \tau_j < \tau_{j+1}\}$
- **Observed TPP:** $p(e_{k+1} | \mathcal{S}_k, \mathcal{M}_{\bar{k}})$
- **Missing TPP Prior:** $p(\epsilon_r | \mathcal{S}_k, \mathcal{M}_{r-1})$
- **Missing TPP Posterior:** $q(\epsilon_{r+1} | e_{k+1}, \mathcal{S}_k, \mathcal{M}_{r-1})$

Gupta et al. Learning Temporal Point Processes with Intermittent Observations.

TPP for Observed Events

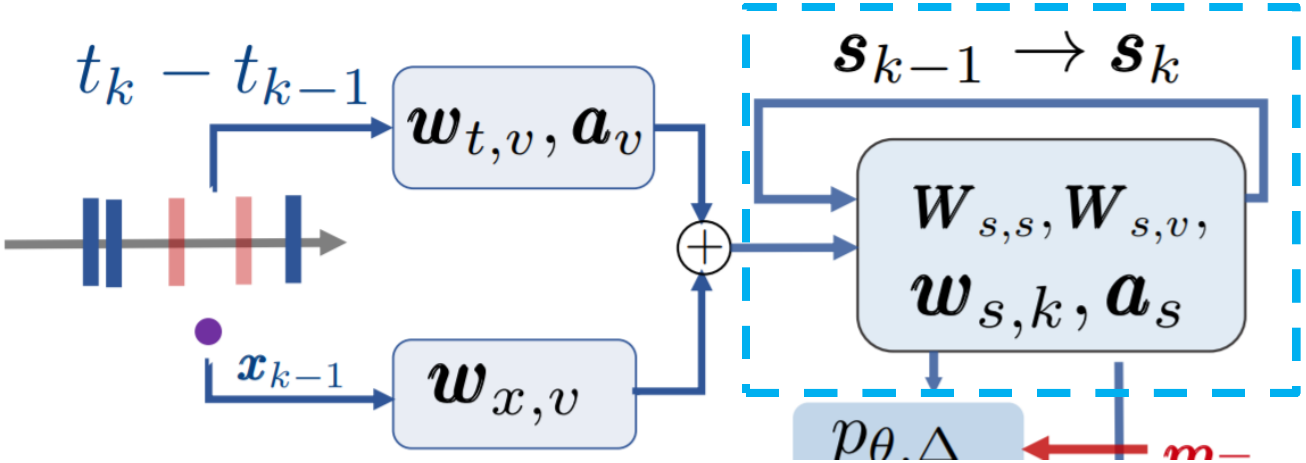


- **Embed input event via a vector.**

$$\boldsymbol{v}_k = \boldsymbol{w}_{t,v} t_k + \boldsymbol{w}_{x,v} x_k + \boldsymbol{w}_{t,\Delta}(t_k - t_{k-1}) + \boldsymbol{a}_v,$$

Gupta et al. Learning Temporal Point Processes with Intermittent Observations.

TPP for Observed Events



- **Embed input event via a vector.**

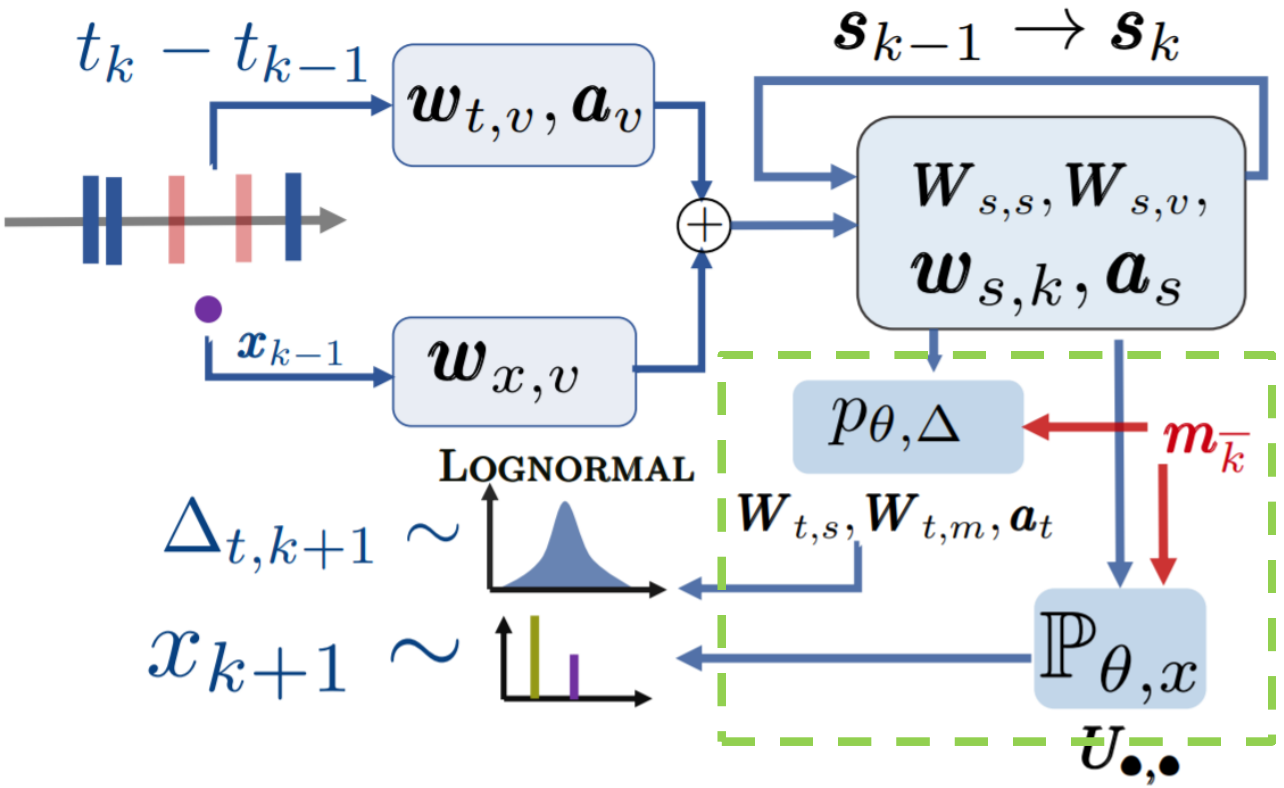
$$\boldsymbol{v}_k = \boldsymbol{w}_{t,v} t_k + \boldsymbol{w}_{x,v} x_k + \boldsymbol{w}_{t,\Delta}(t_k - t_{k-1}) + \boldsymbol{a}_v,$$

- **Update the state of observed TPP.**

$$\boldsymbol{s}_k = \tanh(\boldsymbol{W}_{s,s}\boldsymbol{s}_{k-1} + \boldsymbol{W}_{s,v}\boldsymbol{v}_k + (t_k - t_{k-1})\boldsymbol{w}_{s,k} + \boldsymbol{a}_s)$$

Gupta et al. Learning Temporal Point Processes with Intermittent Observations.

TPP for Observed Events



- **Embed input event via a vector.**

$$\boldsymbol{v}_k = \boldsymbol{w}_{t,v} t_k + \boldsymbol{w}_{x,v} x_k + \boldsymbol{w}_{t,\Delta}(t_k - t_{k-1}) + \boldsymbol{a}_v,$$

- **Update the state of observed TPP.**

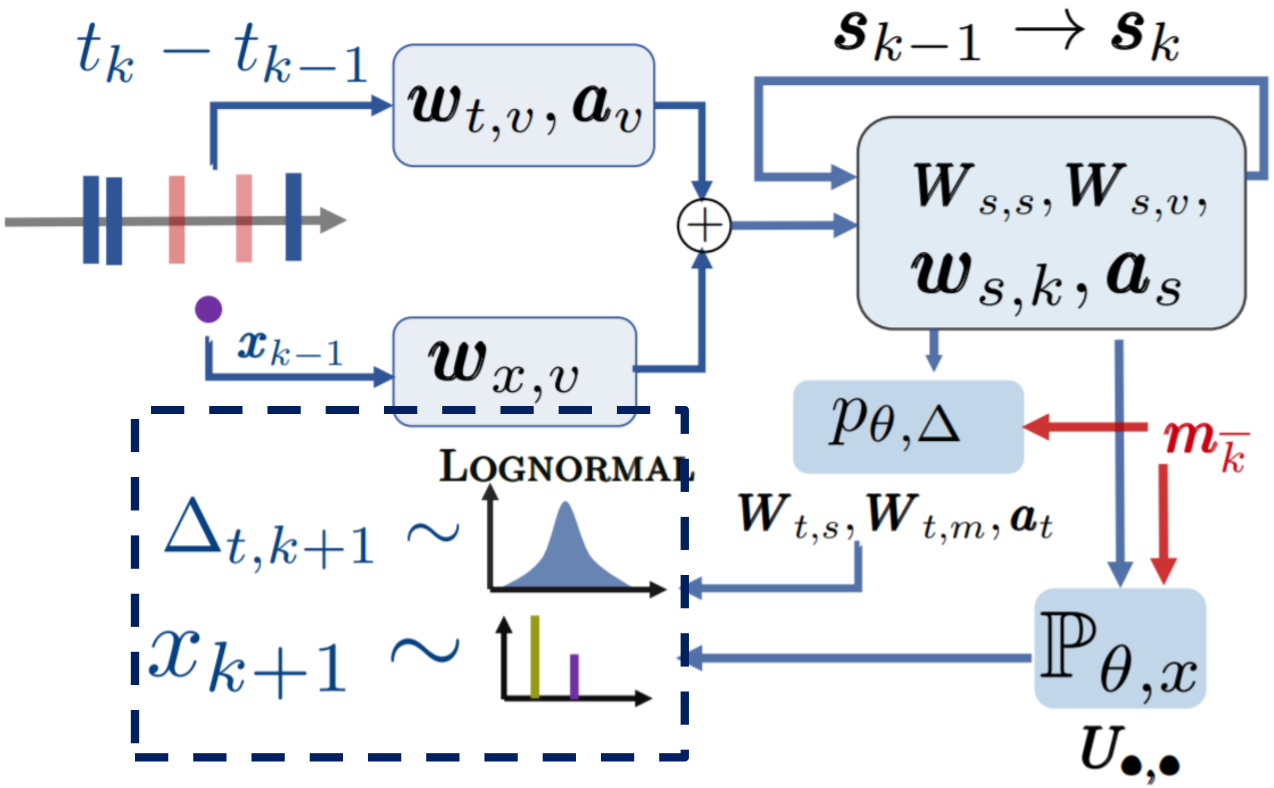
$$\boldsymbol{s}_k = \tanh(\boldsymbol{W}_{s,s}\boldsymbol{s}_{k-1} + \boldsymbol{W}_{s,v}\boldsymbol{v}_k + (t_k - t_{k-1})\boldsymbol{w}_{s,k} + \boldsymbol{a}_s)$$

- **Sample the mark of next event.**

$$\mathbb{P}_{\theta,x}(x_{k+1} = x \mid \Delta_{t,k+1}, \boldsymbol{s}_k, \boldsymbol{m}_{\overline{k}}) = \frac{\exp(\boldsymbol{U}_{x,s}^\top \boldsymbol{s}_k + \boldsymbol{U}_{x,m}^\top \boldsymbol{m}_{\overline{k}})}{\sum_{x' \in \mathcal{C}} \exp(\boldsymbol{U}_{x',s}^\top \boldsymbol{s}_k + \boldsymbol{U}_{x',m}^\top \boldsymbol{m}_{\overline{k}})},$$

Gupta et al. Learning Temporal Point Processes with Intermittent Observations.

TPP for Observed Events



- **Embed input event via a vector.**

$$\boldsymbol{v}_k = \boldsymbol{w}_{t,v} t_k + \boldsymbol{w}_{x,v} x_k + \boldsymbol{w}_{t,\Delta}(t_k - t_{k-1}) + \boldsymbol{a}_v,$$

- **Update the state of observed TPP.**

$$\boldsymbol{s}_k = \tanh(\boldsymbol{W}_{s,s}\boldsymbol{s}_{k-1} + \boldsymbol{W}_{s,v}\boldsymbol{v}_k + (t_k - t_{k-1})\boldsymbol{w}_{s,k} + \boldsymbol{a}_s)$$

- **Sample the mark of next event.**

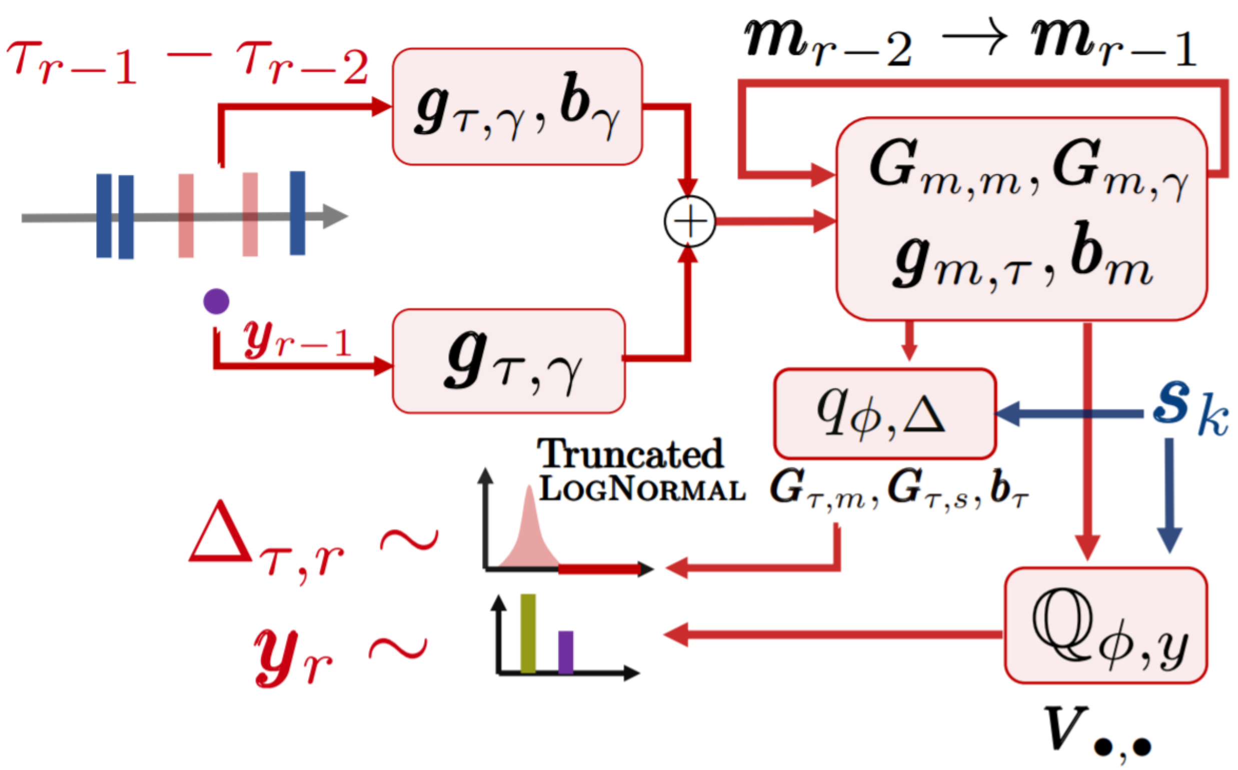
$$\mathbb{P}_{\theta,x}(x_{k+1} = x \mid \Delta_{t,k+1}, \boldsymbol{s}_k, \boldsymbol{m}_{\overline{k}}) = \frac{\exp(\boldsymbol{U}_{x,s}^\top \boldsymbol{s}_k + \boldsymbol{U}_{x,m}^\top \boldsymbol{m}_{\overline{k}})}{\sum_{x' \in \mathcal{C}} \exp(\boldsymbol{U}_{x',s}^\top \boldsymbol{s}_k + \boldsymbol{U}_{x',m}^\top \boldsymbol{m}_{\overline{k}})},$$

- **Time using a *LogNormal* distribution** [Schhur et al 2020].

$$\textsc{LogNormal}\left(\mu_e(\boldsymbol{s}_k, \boldsymbol{m}_{\overline{k}}), \sigma_e^2(\boldsymbol{s}_k, \boldsymbol{m}_{\overline{k}})\right),$$

Gupta et al. Learning Temporal Point Processes with Intermittent Observations.

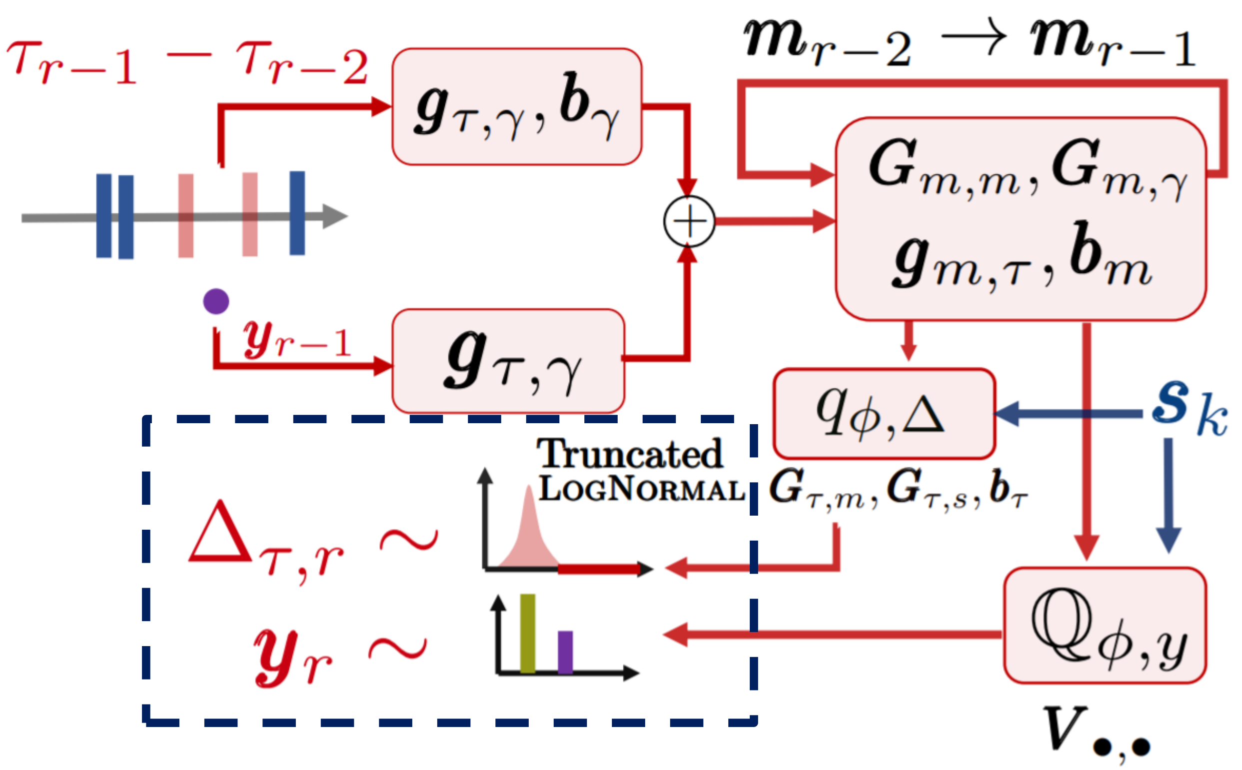
TPP for Missing Events



- **Mostly same as the observed TPP.**

Gupta et al. Learning Temporal Point Processes with Intermittent Observations.

TPP for Missing Events



- **Mostly same as the observed TPP.**
- **Major Distinction:**
  - Sampling the next missing event is bounded by the time of the next observed event.
  - Simple LogNormal won't work here.
  - Sample for a *truncated* LogNormal distribution.

$$\text{LogNormal}\left(\mu_\epsilon(\boldsymbol{m}_{r-1}, \boldsymbol{s}_k), \sigma_\epsilon^2(\boldsymbol{m}_{r-1}, \boldsymbol{s}_k)\right) \odot [\![\tau_{r-1} + \Delta_{\tau,r} < t_{k+1}]\!],$$

  - We follow a similar procedure for Missing TPP prior.

Gupta et al. Learning Temporal Point Processes with Intermittent Observations.

IMTPP: Training Procedure

How is IMTPP trained?

- **Likelihood of observed data demands marginalization with respect to the set of latent missing events.**

IMTPP: Training Procedure

How is IMTPP trained?

- **Likelihood of observed data demands marginalization with respect to the set of latent missing events.**
- **Therefore, we resort to maximizing a variational lower bound of the log-likelihood of the observed stream.**

$$p(\mathcal{S}_K) = \prod_{k=1}^{K-1} \int_{\mathcal{M}_{\overline{k}}} p(e_{k+1} \,|\, \mathcal{S}_k, \mathcal{M}_{\overline{k}})\, p(\mathcal{M}_{\overline{k}})\, d\omega(\mathcal{M}_{\overline{k}})$$

- **Maximixe the ELBO:**

$$\max_{\theta,\phi} \mathrm{ELBO}(\theta, \phi)$$

Gupta et al. Learning Temporal Point Processes with Intermittent Observations.

Results

	Mean Absolute Error (MAE)						Mean Prediction Accuracy (MPA)					
	Movies	Toys	Taxi	Retweet	SO	Foursquare	Movies	Toys	Taxi	Retweet	SO	Foursquare
HP (Hawkes, 1971)	0.060	0.062	0.220	0.049	0.010	0.098	0.482	0.685	0.894	0.531	0.418	0.523
SMHP (Liniger, 2009)	0.062	0.061	0.213	0.051	0.008	0.091	0.501	0.683	0.893	0.554	0.423	0.520
RMTPP (Du et al., 2016)	$\boxed{0.053}$	$\boxed{0.048}$	$\boxed{0.128}$	$\boxed{0.040}$	**0.005**	$\boxed{0.047}$	0.548	0.734	0.929	$\boxed{0.572}$	0.446	0.605
SAHP (Zhang et al., 2020)	0.072	0.073	0.174	0.081	0.017	0.108	0.458	0.602	0.863	0.461	0.343	0.459
THP (Zuo et al., 2020)	0.068	0.057	0.193	0.047	$\boxed{0.006}$	0.052	0.537	0.724	$\boxed{0.931}$	0.526	**0.458**	**0.624**
PFPP (Mei et al., 2019)	0.058	0.055	0.181	0.042	0.007	0.076	$\boxed{0.559}$	$\boxed{0.738}$	0.925	0.569	0.437	0.582
HPMD (Shelton et al., 2018)	0.060	0.061	0.208	0.048	0.008	0.087	0.513	0.688	0.907	0.558	0.439	0.531
IMTPP (our proposal)	**0.049**	**0.045**	**0.108**	**0.038**	**0.005**	**0.041**	**0.574**	**0.746**	**0.938**	**0.577**	$\boxed{0.451}$	$\boxed{0.612}$

Performance in terms of time prediction error (MAE) and mark prediction accuracy (MPA).

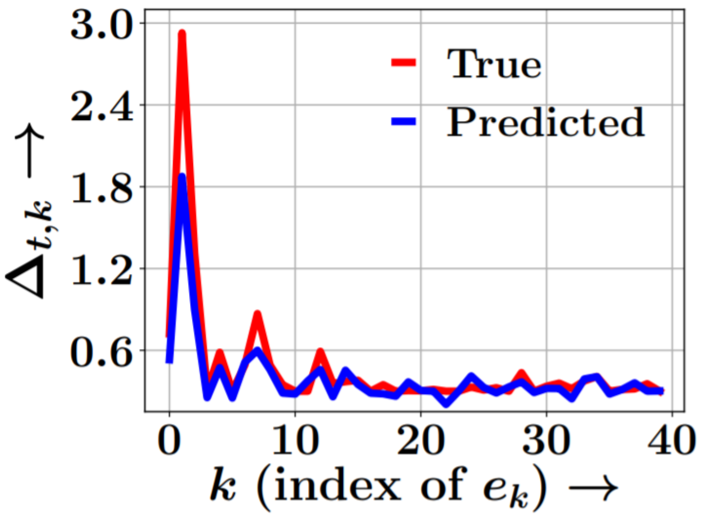
- **Best in terms of time prediction.**
- **Comparable for mark prediction.**

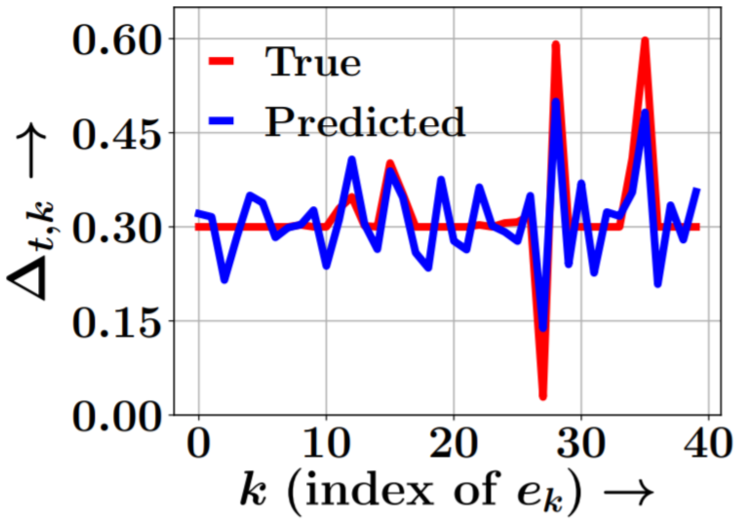
Six Datasets:

- Amazon Movies
- Amazon Toys
- NYC Taxi
- Retweet
- Stack Overflow
- Foursquare

Gupta et al. Learning Temporal Point Processes with Intermittent Observations.

Qualitative Assessment

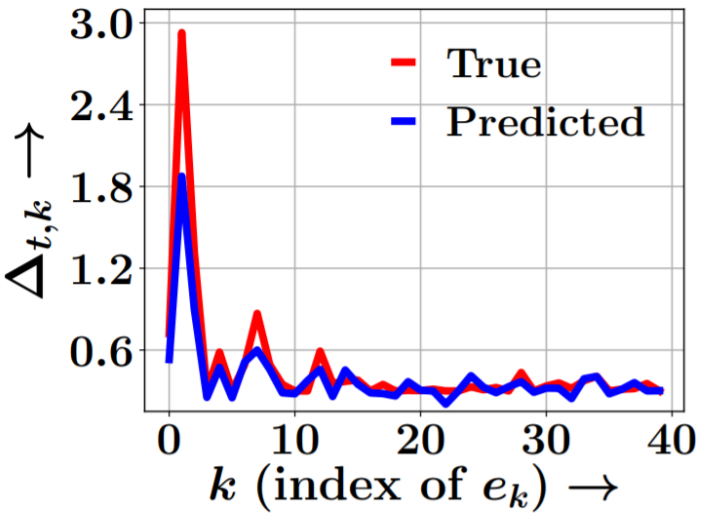


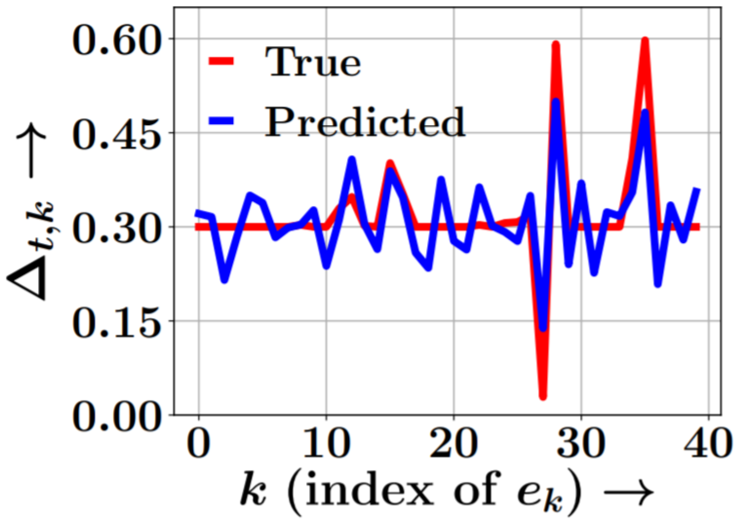


Difference between the actual time and the time predicted by IMTPP for Movies and Toys dataset.

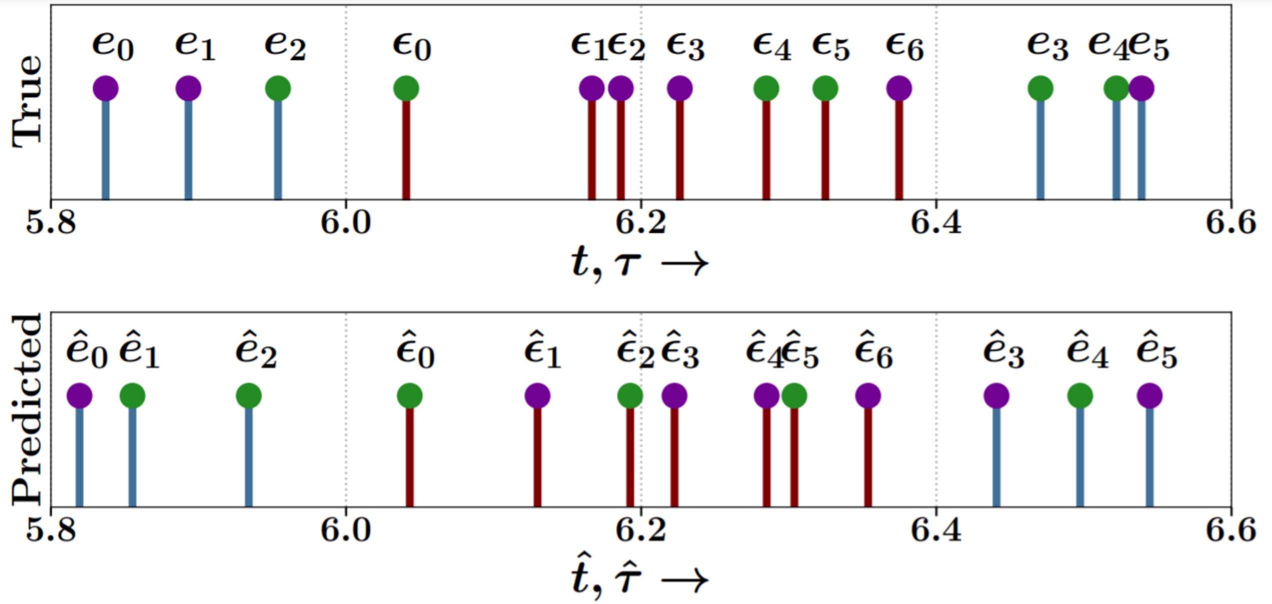
Gupta et al. Learning Temporal Point Processes with Intermittent Observations.

Qualitative Assessment





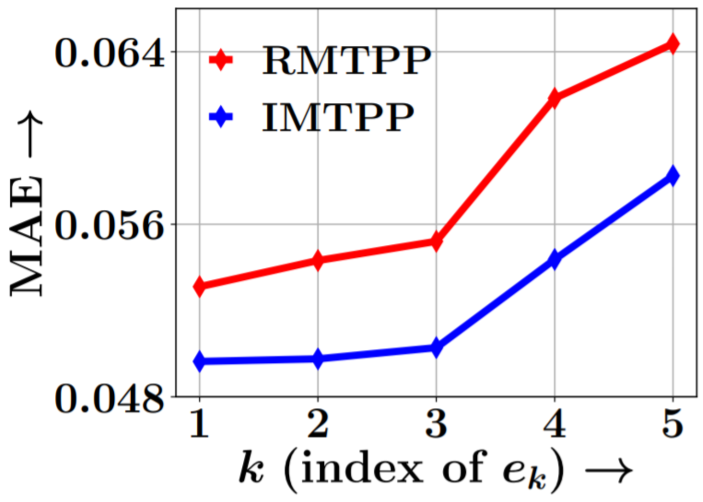
Difference between the actual time and the time predicted by IMTPP for Movies and Toys dataset.

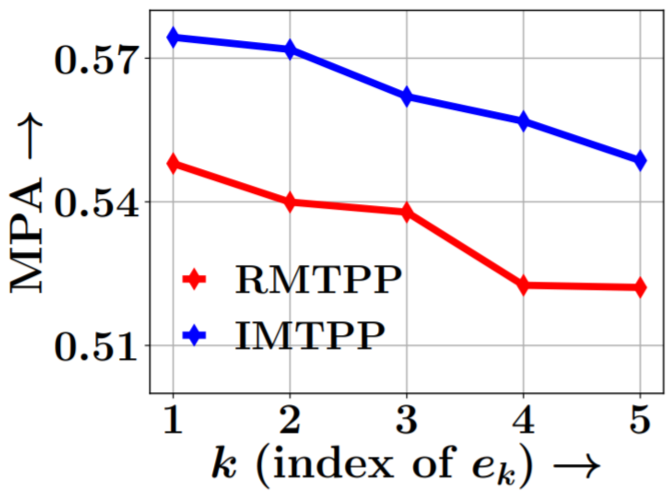


Actual and predicted events by IMTPP for an observed sequence.

Gupta et al. Learning Temporal Point Processes with Intermittent Observations.

Generative Process and Scalability

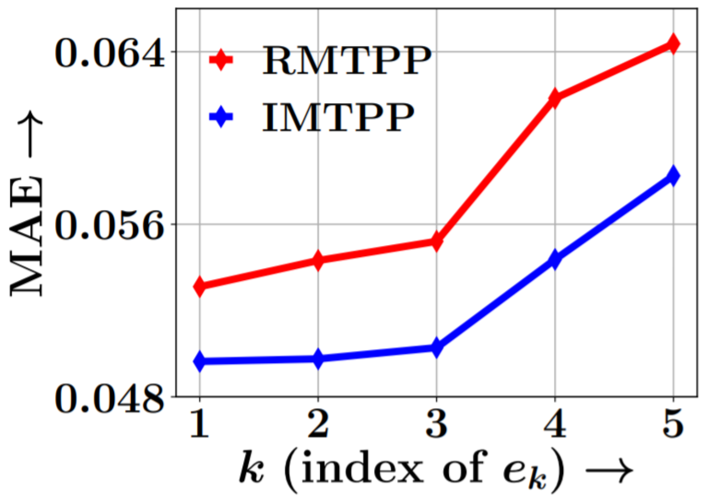


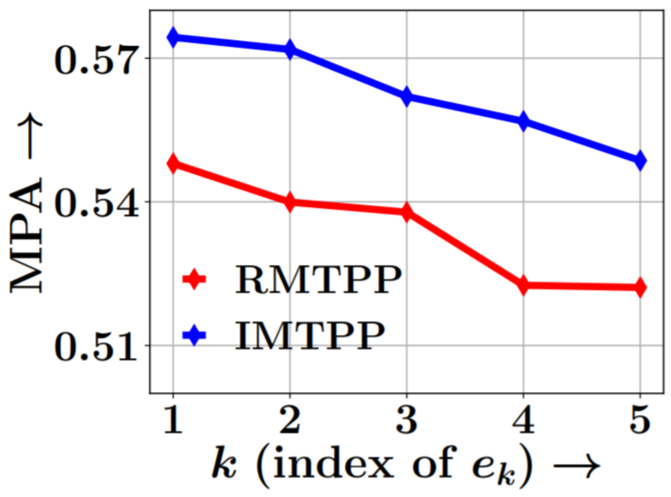


Evaluating the generative process by next *N* event prediction given an input.

Gupta et al. Learning Temporal Point Processes with Intermittent Observations.

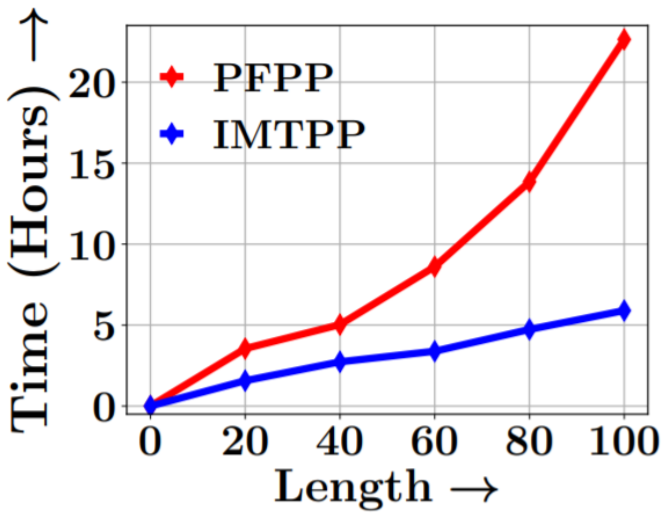
Generative Process and Scalability

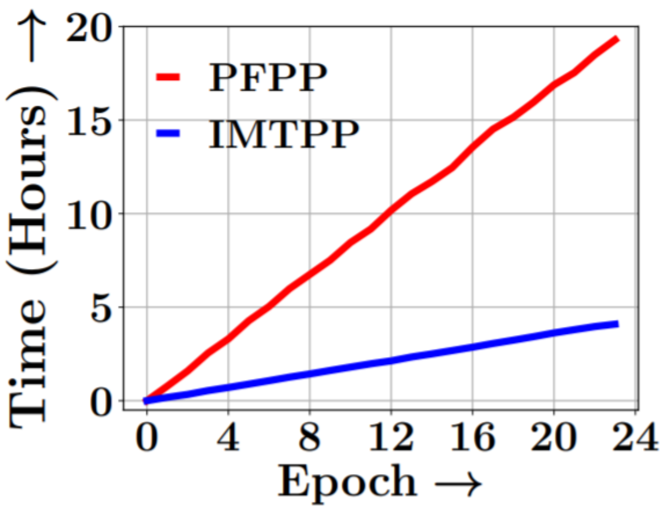




Evaluating the generative process by next *N* event prediction given an input.

Scalability of IMTPP and PFPP in terms of sequence length and training epochs.





Gupta et al. Learning Temporal Point Processes with Intermittent Observations.

References

- Du, N., Dai, H., Trivedi, R., Upadhyay, U., Rodriguez, M. G., and Song, L. (2016). Recurrent marked temporal point processes: Embedding event history to vector. In KDD.
- Hawkes, A. G. (1971). Spectra of some self-exciting and mutually exciting point processes. Biometrika, 58(1).
- Liniger, T. J. (2009). Multivariate hawkes processes. PhD thesis, ETH Zurich.
- Mei, H., Qin, G., and Eisner, J. (2019). Imputing missing events in continuous-time event streams. In ICML.
- Shchur, O., Biloš, M., and Günnemann, S. (2020). Intensity-free learning of temporal point processes. In ICLR.
- Shelton, C. R., Qin, Z., and Shetty, C. (2018). Hawkes process inference with missing data. In AAAI.
- Zhang, Q., Lipani, A., Kirnap, O., and Yilmaz, E. (2020). Self-attentive hawkes processes. ICML.
- Zuo, S., Jiang, H., Li, Z., Zhao, T., and Zha, H. (2020). Transformer hawkes process. In ICML.

THANK YOU,
Stay Safe!