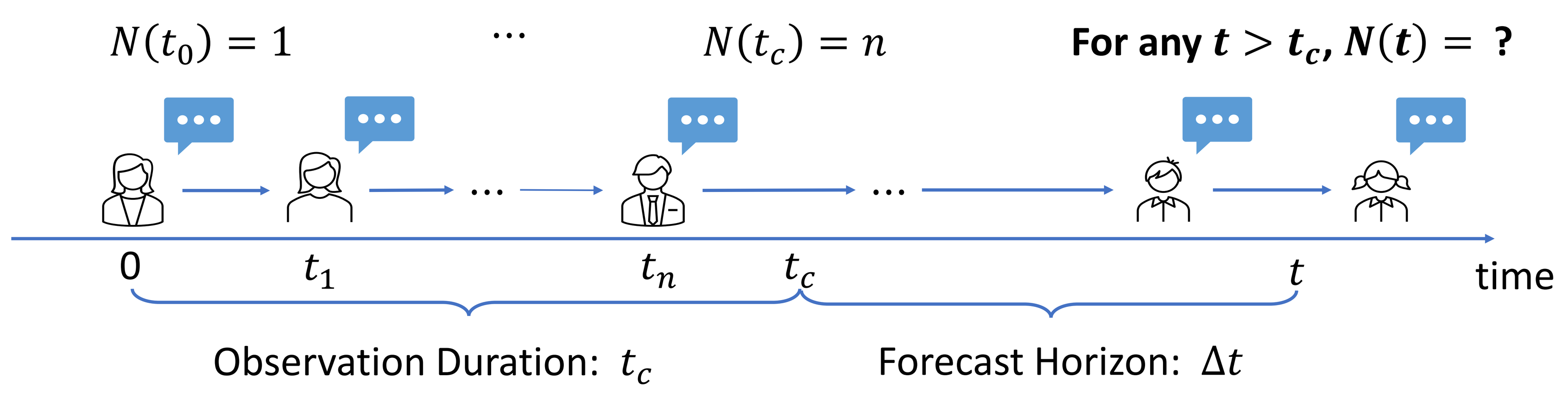
Anytime Information Cascade Popularity Prediction via Self-Exciting Processes

Xi Zhang[1], Akshay Aravamudan[1], Georgios C. Anagnostopoulos[1]
[1]Florida Institute of Technology

Motivation

- The Hawkes point process has been widely adopted in modeling the diffusion of cascades due to its self-exciting property.
- The process's conditional (on the observed events) mean counts are generally adopted for prediction.



However,

- There are no closed-form expressions to calculate such means.
- Trained generative models are not optimized for prediction.

Marked Hawkes Point Process (MHPP)

Consider an MHPP $N(t)$ with ground intensity

$$\lambda^*(t) \triangleq b(t) + \sum_{i:t_i<t} \phi_{m_i}(t-t_i)$$

with the following **assumptions**

- Both $b(\cdot)$ and $\{\phi_m(\cdot)\}_{m\in\mathcal{M}}$ are Lebesgue-integrable on $\mathbb{R}^+$.
- Marks are unpredictable and follow some distribution $g(m)$.

Contributions

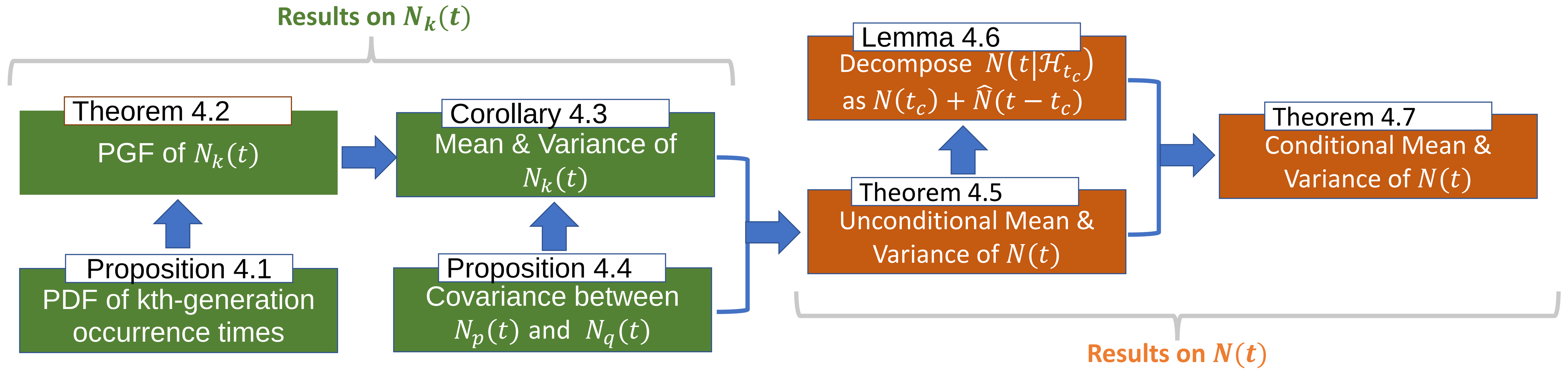
- For general MHPPs, we derived **closed-form** expressions for the **conditional** (on the observed history $\mathcal{H}_{t_c}$) mean and variance of its counting process $N(t)$ at $t \geq t_c$.
- For anytime popularity prediction, we propose **CASPER**, a Hawkes process based **predictive** model, which directly minimizes the prediction error.

Conditional Moments of MHPP

By viewing the MHPP as its equivalent branching process, we can express its counting process as

$$N(t) = \sum_{k\geq 0} N_k(t)$$

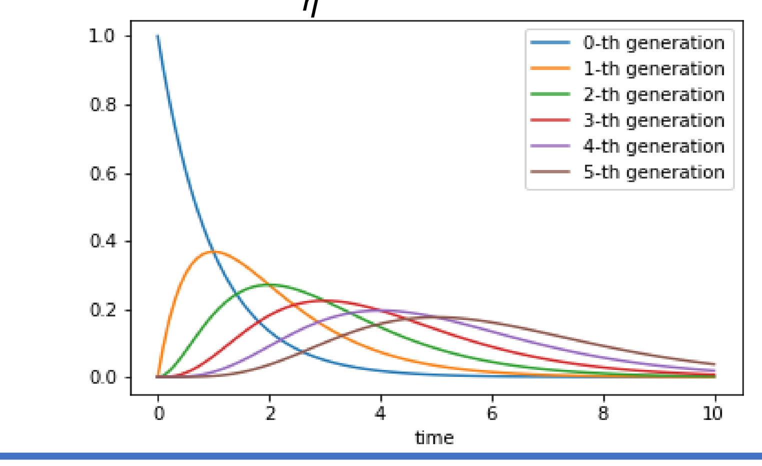
where $N_k(t)$ is the k-th generation counting process.



Proposition 4.1 *Let $\eta \triangleq \int_{s=0}^{\infty} b(s)ds$, $\gamma_m \triangleq \int_{s=0}^{\infty} \phi_m(s)ds$ and $\xi(t) \triangleq E_m\left\{\frac{\phi_m(t)}{\gamma_m}\right\}$, the occurrence times of k-th generation events are i.i.d. with pdf*

$$f_k(t) \triangleq \frac{1}{\eta}\left(b * \xi^{*k}\right)(t), \qquad t \geq 0$$

Pdf of k-th generation event occurrence time given $\frac{1}{\eta}b(t) = \xi(t) = e^{-t}$



Theorem 4.2 *For $k \geq 1$, the pgf of $N_k(t)$ is given as*

$$G_{N_k(t)}(w) = G_0\left(G^{\circ k}\left(G_k(w)\right)\right)$$

where $G_0(w) \triangleq e^{\eta(w-1)}$, $G^{\circ k}(w)$ is the k-fold composition of $G(w) \triangleq E_m\{e^{\gamma_m(w-1)}\}$, and $G_k(w) \triangleq 1 + F_k(t)(w-1)$, with $F_k(t)$ be the cdf of the k-th generation event occurrence times.

Theorem 4.7 [Partial] Given $\mathcal{H}_{t_c} = \{(t_i, m_i)\}_{i:\,t_i\leq t_c}$. Let $\Delta t \triangleq t - t_c$, then for $t \geq t_c$,

$$\mathrm{E}\{N(t|\mathcal{H}_{t_c})\} = N(t_c) + \sum_{k\geq 0} u * \hat{b} * (\gamma\xi)^{*k}(\Delta t)$$

where $\gamma \triangleq E_m\{\gamma_m\}$, and $\hat{b}(\Delta t) \triangleq b(\Delta t + t_c) + \sum_{(t_i,m_i)\in\mathcal{H}_{t_c}} \phi_{m_i}(\Delta t + t_c - t_i)$.

CASPER's Predictive Training

- Let $S(t_c) \triangleq \{(i,j): 0 < t_i < t_j \leq t_c\}$, CASPER's **loss function** is defined as

$$L(\boldsymbol{\theta}|\mathcal{H}_{t_c}) \triangleq \frac{1}{|S(t_c)|}\sum_{(i,j)\in S(t_c)}\left(\mathrm{E}\{N(t_j|\mathcal{H}_{t_i};\boldsymbol{\theta})\} - j\right)^2$$

True count at time t_j

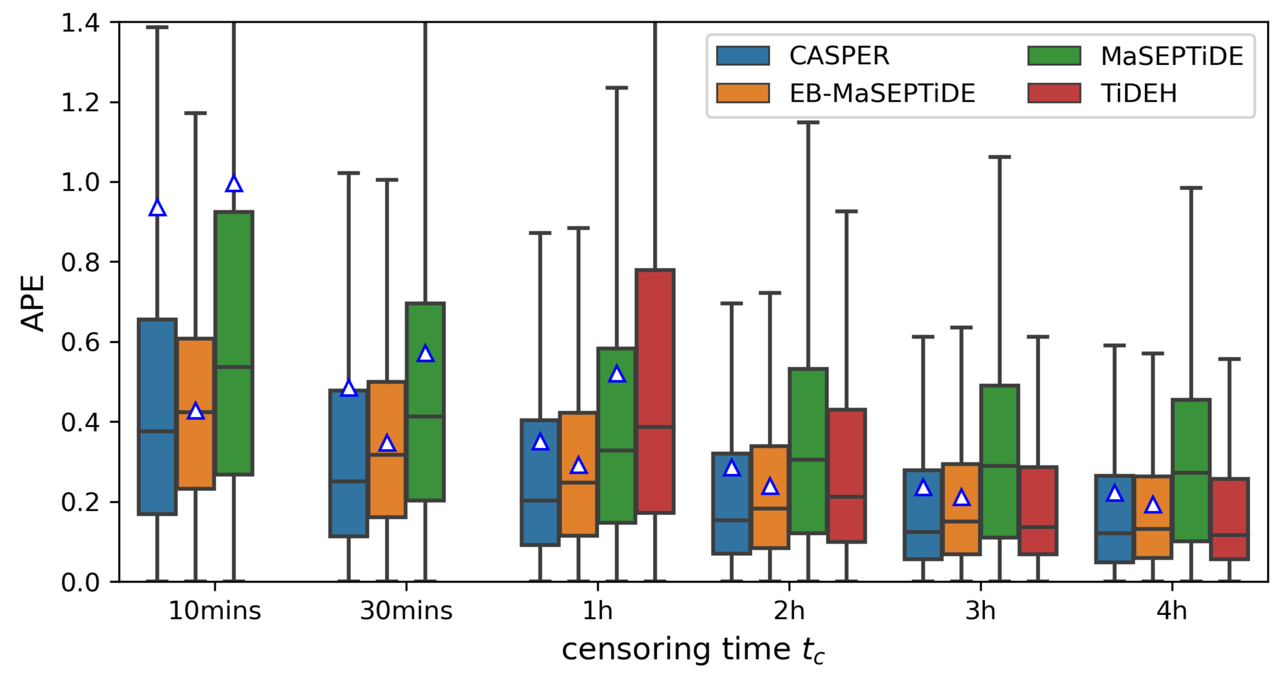
Given observation up to t_i, predicted count at time t_j

- The model parameters are learned as

$$\boldsymbol{\theta}^* \in \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\operatorname{argmin}}\, L(\boldsymbol{\theta}|\mathcal{H}_{t_c})$$

Tweets Popularity Prediction

CASPER exhibits competitive, if not the best performance, especially for early-stage prediction.



Boxplots of APE values for prediction in 4 days (Δt = 4 days) between our and other models on real-world SEISMIC dataset, with various censoring times. Horizontal bars indicate medians, and the white triangles indicate means.

CASPER Python Code: https://github.com/xizhang-cc/casper

Correspondence to: Xi Zhang <zhang2012@my.fit.edu>