Differentially Private Learning of Hawkes Processes

Mohsen Ghassemi* **Eleonora Kreačić*** **Niccolò Dalmasso**

Vamsi K. Potluru **Tucker Balch** **Manuela Veloso**

J.P. Morgan AI Research
{mohsen.ghassemi, eleonora.kreacic, niccolo.dalmasso}@jpmchase.com

2022-arkiv.org

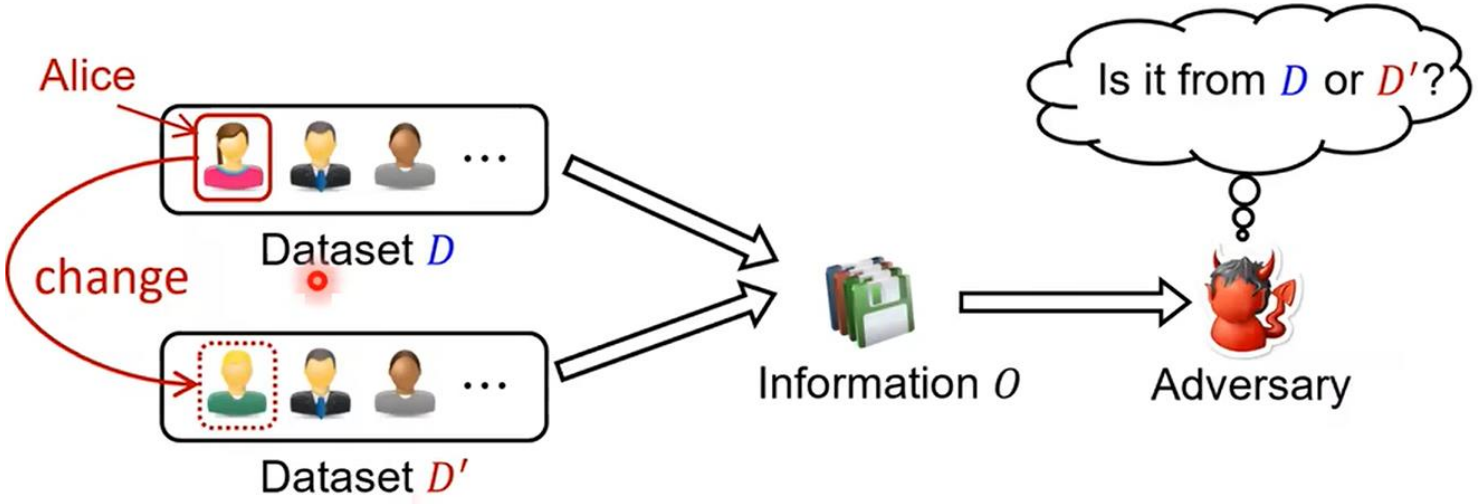
Reporter: Jiajia Sun

Outline

- DP modeling of sequential events data
- Private Estimation parameters of Hawkes process
- Sample complexity
- Experiments

DP modeling of sequential events data

- **Goal:** learn the parameters of the Hawkes process while **preserving the privacy** of the individuals whose data (events) are present in the stream



- ϵ − differentially private $$\frac{\Pr[A(D) = O]}{\Pr[A(D') = O]} \leq \exp(\varepsilon)$$
- randomly differentially private $$\mathbb{P}\left(\forall \mathcal{C} \subset \mathcal{Y}, \quad \mathbb{P}\left(\mathfrak{M}(\mathcal{D}) \in \mathcal{C}\right) \leq e^{\epsilon}\mathbb{P}\left(\mathfrak{M}(\mathcal{D}') \in \mathcal{C}\right)\right) \geq 1 - \gamma$$

DP modeling of sequential events data

- sequence of events $S(t) = \{e_i = (x_i, t_i) | t_i < t\}$

- parent-child relation

  We can define a cluster as a group of individuals whose events forms **a tree of parent-child relations**

- “neighboring” sequence

  $S(t)\ and\ S_{-j}(t)$

  they differ only in the presence of the events that **belong to the same cluster** as event j

Definition 4. *Let $\mathcal{S}_{\mathbf{P}}^{T}$ be the set of all possible realizations of a temporal point process $\mathbf{P}(t)$ until time T. A randomized mechanism $\mathfrak{M} : \mathcal{S}_{\mathbf{P}}^{T} \rightarrow \mathcal{Y}$ is (ϵ, γ)-randomly differentially private if*

$$\mathbb{P}\Big(\forall \mathcal{C} \subset \mathcal{Y}, \quad \mathbb{P}\left(\mathfrak{M}(S(T)) \in \mathcal{C}\right) \leq e^{\epsilon}\mathbb{P}\left(\mathfrak{M}(S_{-i}(T)) \in \mathcal{C}\right)\Big) \geq 1 - \gamma$$

where the inner probability is over the randomness of the mechanism, and the outer probability is over neighboring streams $S(T), S_{-i}(T) \in \mathcal{S}_{\mathbf{P}}^{t}$ drawn from point process $\mathbf{P}(t)$ until time T.

Estimating parameters of Hawkes process

- Hawkes process $\lambda_t^* = \mu + \sum_{t_i<t} \alpha e^{-\beta(t_i-t)}$, $\beta = 1$

$$\lambda_\infty := \lim_{t\to\infty} \mathbb{E}[\lambda_t^*] = \frac{\mu}{1-\alpha}$$

- The count series with interval size Δ $Y_i(\Delta) = \mathbf{N}(i\Delta) - \mathbf{N}((i-1)\Delta)$

$$Y_1(\Delta), Y_2(\Delta), \cdots, Y_K(\Delta)$$

- Standard stationarity assumption [1]

$$\eta := \mathbb{E}[Y_i(\Delta)] = \frac{\mu\Delta}{(1-\alpha)}$$

$$\sigma^2 := \mathrm{Var}[Y_i(\Delta)] = \frac{\mu\Delta}{(1-\alpha)^3} + \frac{\alpha^2\mu\left(1-e^{-2(1-\alpha)\Delta}\right)}{2(1-\alpha)^4} - \frac{2\alpha\mu\left(1-e^{-(1-\alpha)\Delta}\right)}{(1-\alpha)^4}$$

One can compute μ and α given the values of η and σ^2

[1] Gao X, Zhu L. Functional central limit theorems for stationary Hawkes processes and application to infinite-server queues[J]. Queueing Systems, 2018, 90(1): 161-206.

Sensitivity

- Suppose that the maximum number of correlated events is B
- maximal amount of change in the sample mean $\hat{\eta}$ is $\frac{\mathrm{B}}{\mathrm{K}}$
- maximal amount of change in the sample variance $\widehat{\sigma^2}$ is upper bounded by

$$\frac{B^2}{K} + \frac{2B^{3/2}\sqrt{\Delta}C_1}{(K-1)} \text{ } where \text{ } C_1 = \sqrt{\frac{1.1 \cdot \mu_{upper}}{(1-\alpha_{upper})^3} \cdot \frac{1}{\gamma}}.$$ with probability at least $1-\gamma$

Lemma 1. *Consider a Hawkes process $H(t)$ defined by intensity function 1 observed until time T. For any $0 < \gamma \leq 1$ and $T \geq \left(\frac{\mu \cdot e^2}{\gamma}\right)^{5/2}$, with probability at least $1-\gamma$, all existing trees contain at most $\frac{3\log T}{(1-\alpha)^2}$ individuals.*

$$\mathrm{B} = C_2 \log T, where \ C_2 = \frac{3}{\left(1-\alpha_{upper}\right)^2}$$

DP sample mean and sample variance

- Laplace mechanism

$M_{Lap}(\mathcal{D}, f(.), \epsilon) = f(\mathcal{D}) + \Lambda(0, \Delta_f/\epsilon)$ where $\Delta_f = \max_{\mathcal{D},\mathcal{D}'} \|f(\mathcal{D}) - f(\mathcal{D}')\|_1$

$$\hat{\eta}_{\text{private}} = \hat{\eta} + \Lambda\left(\frac{C_2 \log T}{K \cdot \epsilon}\right)$$

$$\hat{\sigma}^2_{\text{private}} = \hat{\sigma}^2 + \Lambda\left(\frac{C_2^2(\log T)^2 + 2C_2^{3/2}C_1 \cdot \frac{K}{K-1}(\log T)^{3/2}\sqrt{\Delta}}{K \cdot \epsilon}\right)$$

$\hat{\mu}_{private}$ and $\hat{\alpha}_{private}$ can be solved with $\hat{\eta}_{private}$ and $\hat{\sigma}_{private}$

Sample complexity

The minimum length of sequence required for the **non-private** estimator

Theorem 1.

If $$T \geq \frac{\sigma^2}{\xi} \max\left\{\frac{C_9^2\Psi(1-\frac{\delta}{8})^2}{\xi\Delta}, \frac{9C_9^2\Psi(1-\frac{\delta}{16})^2(\eta_4-\sigma^2)}{\xi\Delta}, 3C_9\Psi(1-\frac{\delta}{16})^2, \frac{24C_9}{\delta}\right\} \quad (11)$$

$\textit{for some } 0 < \delta \leq 1 \quad 0 < \xi < \frac{C_9\mu_{lower}}{6}$

$\Psi(\cdot)$ denote the inverse CDF of the standard normal distribution.

$$C_9 = \max\{\frac{8}{\mu_{lower}(1-\alpha_{upper})}, 1 + \frac{8\mu_{upper}}{\mu_{lower}(1-\alpha_{upper})^2} + \frac{4}{3(1-\alpha_{upper})}\}$$

Then $\mathbb{P}\left(|\alpha - \hat{\alpha}| > \xi\right) \leq \delta \textit{ and } \mathbb{P}\left(|\mu - \hat{\mu}| > \xi\right) \leq \delta$

Proof sketch. $\mathbb{P}\left(|\hat{\eta} - \eta| > \frac{\xi\Delta}{C_9}\right) < \delta/2$ and $\mathbb{P}\left(|\hat{\sigma}^2 - \sigma^2| > \frac{\xi\Delta}{C_9}\right) < \delta/2$

And employing Berry-Essen theorem

Sample complexity

The minimum length of sequence required for the **private** estimator

Theorem 2.

If

$$T \geq \frac{C_9^2 \mu_{upper}}{(1-\alpha_{upper})^3 \xi^2} \max\left\{\Psi(1-\frac{\delta}{16})^2,\ 9C_9^2\Psi(1-\frac{\delta}{32})^2(\eta_4-\sigma^2)\right\} \quad and \tag{12}$$

$$\frac{T}{\log T} \geq \frac{c\mu_{upper}}{(1-\alpha_{upper})^3 \xi} \max\left\{3C_9\Psi(1-\frac{\delta}{32})^2,\ \frac{48C_9}{\delta}\right\} \quad and \tag{13}$$

$$\frac{T}{(\log T)^{5/2}} > \frac{4\sqrt{c}C_1C_2^2C_9}{\epsilon\xi}\log\left(\frac{4}{\delta}\right) \tag{14}$$

Then

$$\mathbb{P}\left(|\hat{\mu}_{private}-\mu| > \xi\right) \leq \delta \ and \ \mathbb{P}\left(|\hat{\alpha}_{private}-\alpha| > \xi\right) \leq \delta$$

condition (14) is required to bound the tail of Laplace distribution

Cost of privacy

For the inverse CDF function $\Psi(\cdot)$ $\lim_{x \to 0} \Psi(1-x) = \sqrt{2 \log \frac{1}{x}}$ $0 < \xi < \frac{C_9 \mu_{lower}}{6}$

T can be simplified to

- Non-private estimate $T = O(\frac{1}{\delta \xi})$
- Private estimate $T = O(\frac{\log(1/\delta\xi)}{\delta\xi})$

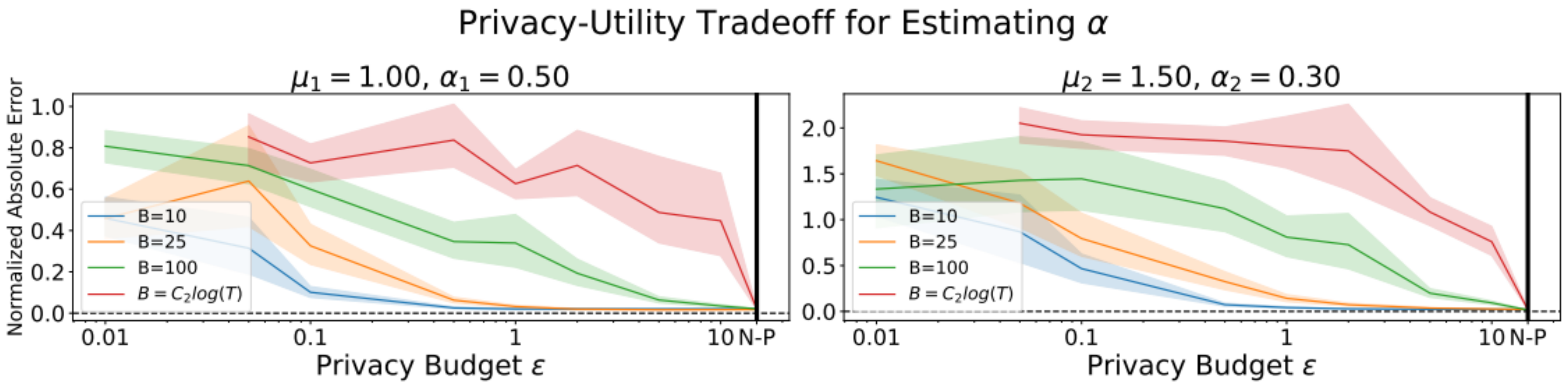
non-private estimates $(\hat{\mu}, \hat{\alpha})$ have a convergence rate of $O(\frac{1}{T})$

private estimates $(\hat{\mu}_{\text{private}}, \hat{\alpha}_{\text{private}})$ is $O(\frac{\boxed{\log T}}{T})$

Experiments

set $\alpha_{\text{upper}} = 0.75$, $\mu_{\text{upper}} = 2.0$, $\gamma = 0.05$ and the Hawkes process decay $\beta = 1.0$

- Synthetic Data



A larger privacy budget corresponds to a lower estimation error

For a given privacy budget, the smaller estimation error the smaller the maximum tree length B

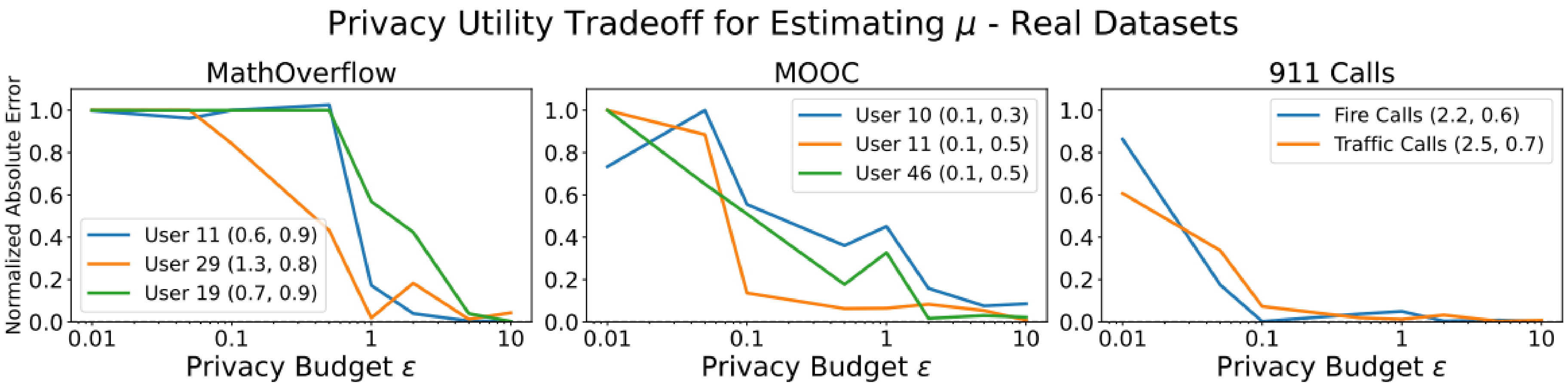
Experiments

set $\alpha_{\text{upper}} = 0.75$, $\mu_{\text{upper}} = 2.0$, $\gamma = 0.05$ and the Hawkes process decay $\beta = 1.0$

- Real Data

  1. **MathOverflow**: user interactions in a question-answering website
  2. **MOOC**: user interactions in open online course
  3. **911 Calls**: medical emergency calls, fire and traffic-related emergency calls

  ground truth: non-private estimates



Conclusion

- differentially private version for estimating the parameters of a Hawkes process
- provide sample complexity results for estimating the parameters of a Hawkes process

Thank you!