Optimal Cox regression under federated differential privacy: Coefficients and cumulative hazards

Yi Yu
Department of Statistics, University of Warwick

The New York Times

The 2020 Census Suggests That People Live Underwater. There's a Reason.

Technology advances forced the Census Bureau to use sweeping measures to ensure privacy for respondents. The ensuing debate goes to the heart of what a census is.

The Census Bureau says that 14 people live in this bend in the Chicago River. It's one of thousands of bits of incorrect data in the 2020 census meant to protect the privacy of census respondents. Jamie Kelter Davis for The New York Times

https://www.theverge.com/2015/3/10/8177683/apple-research-kit-app-ethics-medical-research

A privacy mechanism is a randomised algorithm taking an input dataset $X = (X_1, \ldots, X_n) \in \mathcal{X}^n$ and producing publishable data Z. Formally, it is a collection of conditional distributions $\mathcal{Q} = \{Q(\cdot|x) : x \in \mathcal{X}^n\}$ such that

$$Z|\{X = x\} \sim Q(\cdot|x).$$

Privacy mechanism Q is called (ϵ, δ)-(central) differentially private (Dwork et al., 2006), with $\epsilon > 0$ and $\delta \geq 0$, if

$$Q(A|x) \leq e^{\epsilon} Q(A|x') + \delta,$$

for all measurable set A, any pair $x = (x_i)_{i=1}^n, x' = (x_i')_{i=1}^n \in \mathcal{X}^n$ such that $\sum_{i=1}^n \mathbf{1}\{x_i \neq x_i'\} \leq 1$. We focus on the regime $\epsilon \in (0, 1]$.

At a high level, this quantifies how similar the private outcomes are in terms of total variation distance, by changing one out of n samples.

A privacy mechanism is a randomised algorithm taking an input dataset $X = (X_1, \ldots, X_n) \in \mathcal{X}^n$ and producing publishable data Z. Formally, it is a collection of conditional distributions $\mathcal{Q} = \{Q(\cdot|x) : x \in \mathcal{X}^n\}$ such that

$$Z|\{X = x\} \sim Q(\cdot|x).$$

Privacy mechanism Q is called (ϵ, δ)-(central) differentially private (Dwork et al., 2006), with $\epsilon > 0$ and $\delta \geq 0$, if

$$Q(A|x) \leq e^{\epsilon} Q(A|x') + \delta,$$

for all measurable set A, any pair $x = (x_i)_{i=1}^n, x' = (x_i')_{i=1}^n \in \mathcal{X}^n$ such that $\sum_{i=1}^n \mathbf{1}\{x_i \neq x_i'\} \leq 1$. We focus on the regime $\epsilon \in (0, 1]$.

At a high level, this quantifies how similar the private outcomes are in terms of total variation distance, by changing one out of n samples.

For the central differential privacy (CDP), where there is a trusted central data curator having access to all the raw data. For example, when estimating a univariate mean, we can have

$$\widehat{\theta} = Z = \frac{1}{n}\sum_{i=1}^{n} X_i + \frac{1}{n\epsilon} W, \quad \text{with } W \sim \text{Lap}(1).$$

The variance of total added noise is of order $(n^2\epsilon^2)^{-1}$.

A stronger notion of differential privacy is the local differential privacy (LDP), where data are randomised before collection, that is

$$\mathbb{P}(Z_i \in A | X_i = x) \leq e^{\epsilon}\mathbb{P}(Z_i \in A | X_i = x') + \delta, \quad i \in \{1, \ldots, n\},$$

for all measurable set A and any pair $x, x' \in \mathcal{X}$. For example, when estimating a univariate mean, we can have

$$\widehat{\theta} = \frac{1}{n}\sum_{i=1}^{n} Z_i = \frac{1}{n}\sum_{i=1}^{n}\left(X_i + \frac{1}{\epsilon} W_i\right), \quad \text{with } \{W_i\}_{i=1}^{n} \overset{\text{i.i.d.}}{\sim} \text{Lap}(1).$$

The variance of total added noise is of order $(n\epsilon^2)^{-1}$.

For the central differential privacy (CDP), where there is a trusted central data curator having access to all the raw data. For example, when estimating a univariate mean, we can have

$$\widehat{\theta} = Z = \frac{1}{n}\sum_{i=1}^{n} X_i + \frac{1}{n\epsilon} W, \quad \text{with } W \sim \text{Lap}(1).$$

The variance of total added noise is of order $(n^2\epsilon^2)^{-1}$.

A stronger notion of differential privacy is the local differential privacy (LDP), where data are randomised before collection, that is

$$\mathbb{P}(Z_i \in A | X_i = x) \leq e^{\epsilon}\mathbb{P}(Z_i \in A | X_i = x') + \delta, \quad i \in \{1, \ldots, n\},$$

for all measurable set A and any pair $x, x' \in \mathcal{X}$. For example, when estimating a univariate mean, we can have

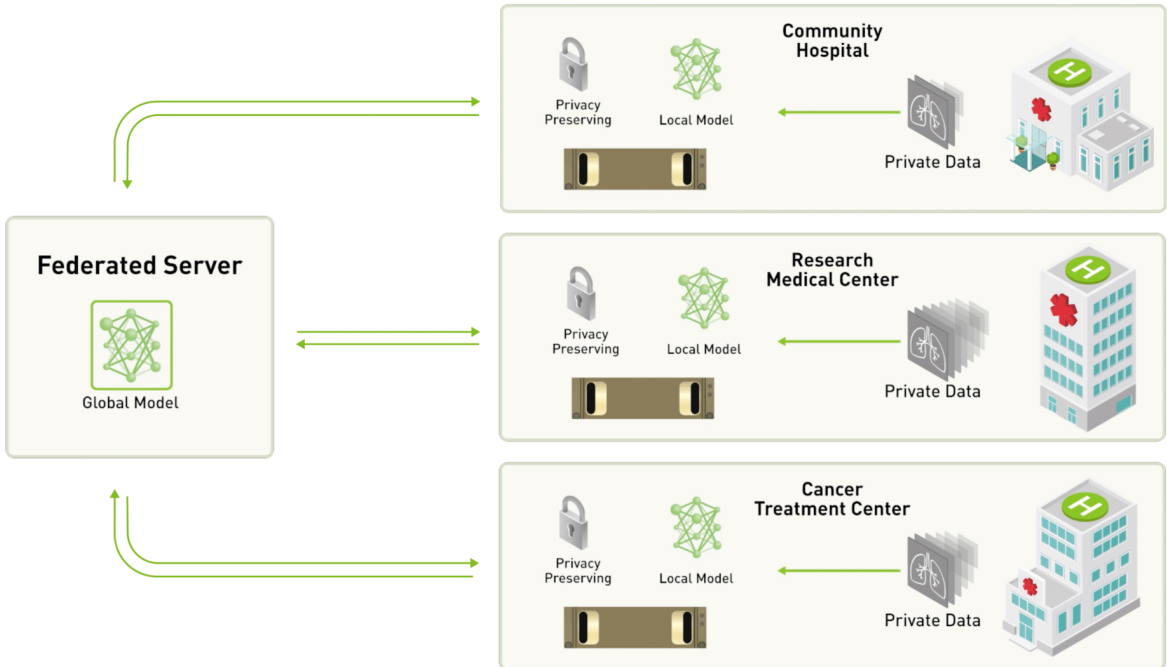
$$\widehat{\theta} = \frac{1}{n}\sum_{i=1}^{n} Z_i = \frac{1}{n}\sum_{i=1}^{n}\left(X_i + \frac{1}{\epsilon} W_i\right), \quad \text{with } \{W_i\}_{i=1}^{n} \overset{\text{i.i.d.}}{\sim} \text{Lap}(1).$$

The variance of total added noise is of order $(n\epsilon^2)^{-1}$.

Remarks

- Non-interactive, sequentially interactive and fully-interactive LDP mechanisms.
- Large ϵ regimes.

Federated learning

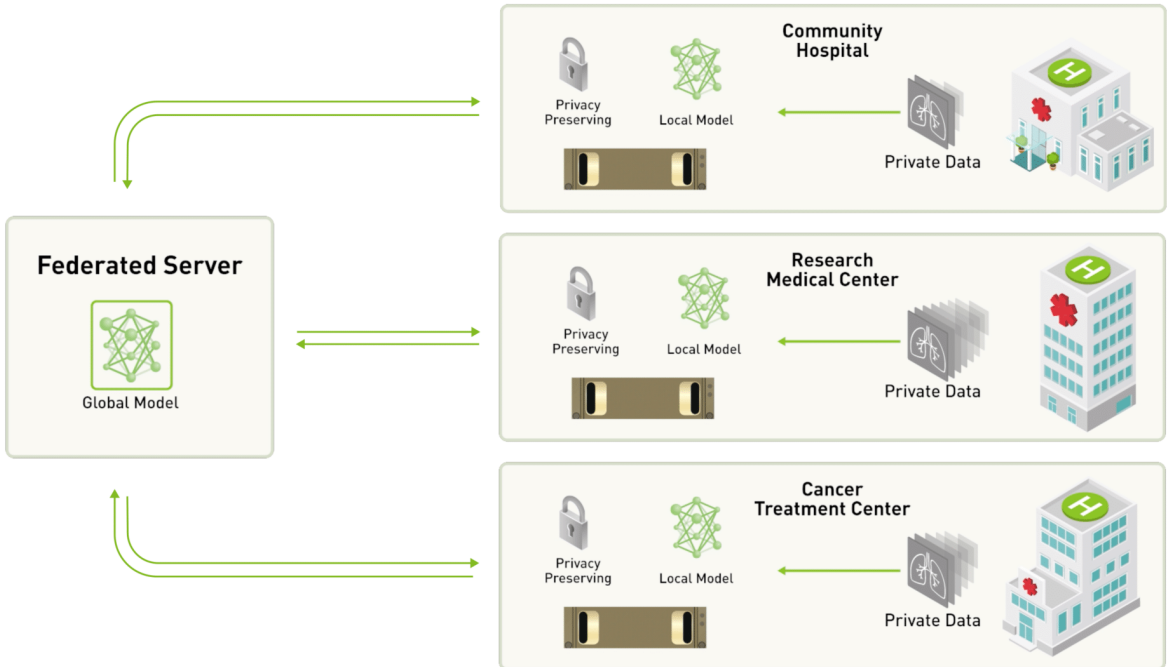


https://blogs.nvidia.com/blog/what-is-federated-learning/

Challenges

- Heterogeneity: distributions, privacy requirement types, privacy budgets.
- Communications: efficiency in aggregating and communicating siloed information.

Federated learning

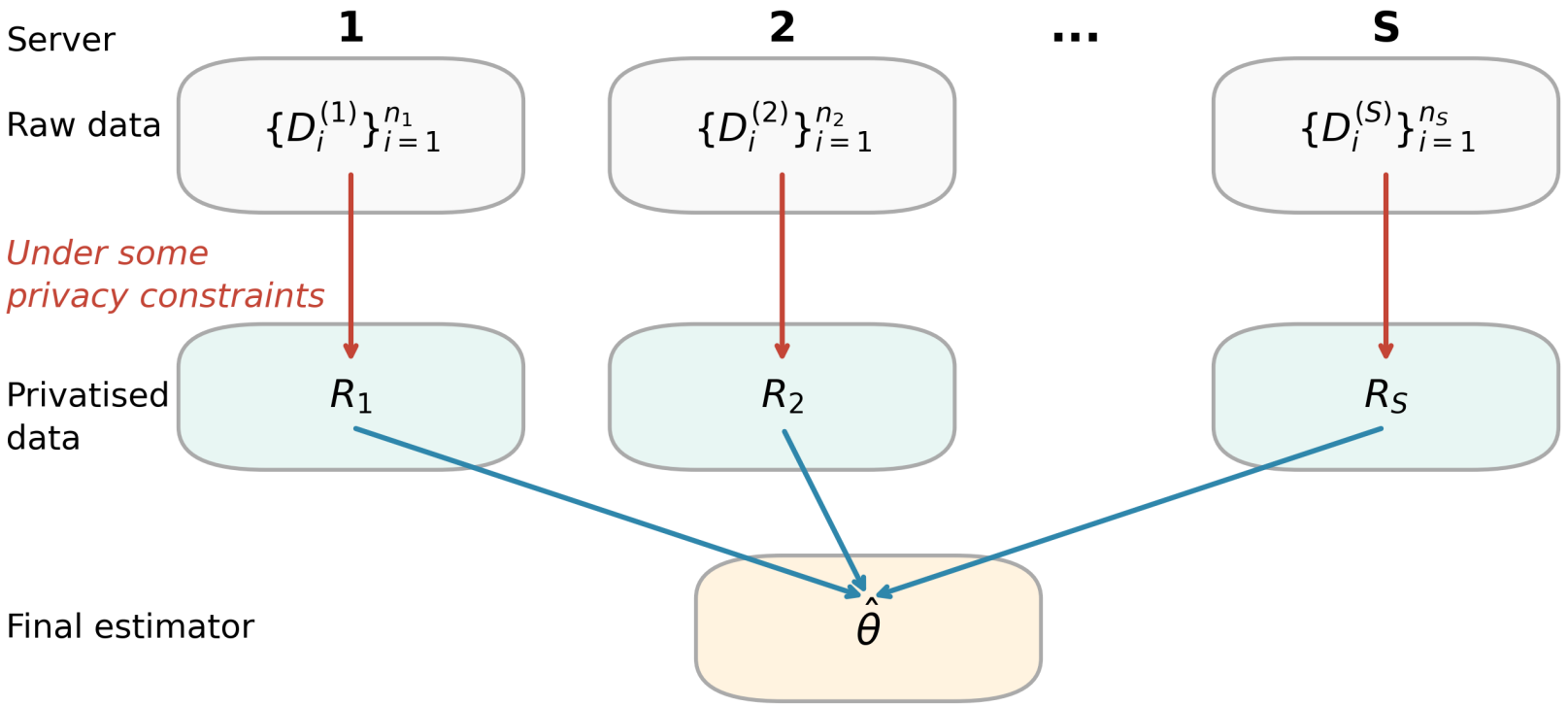


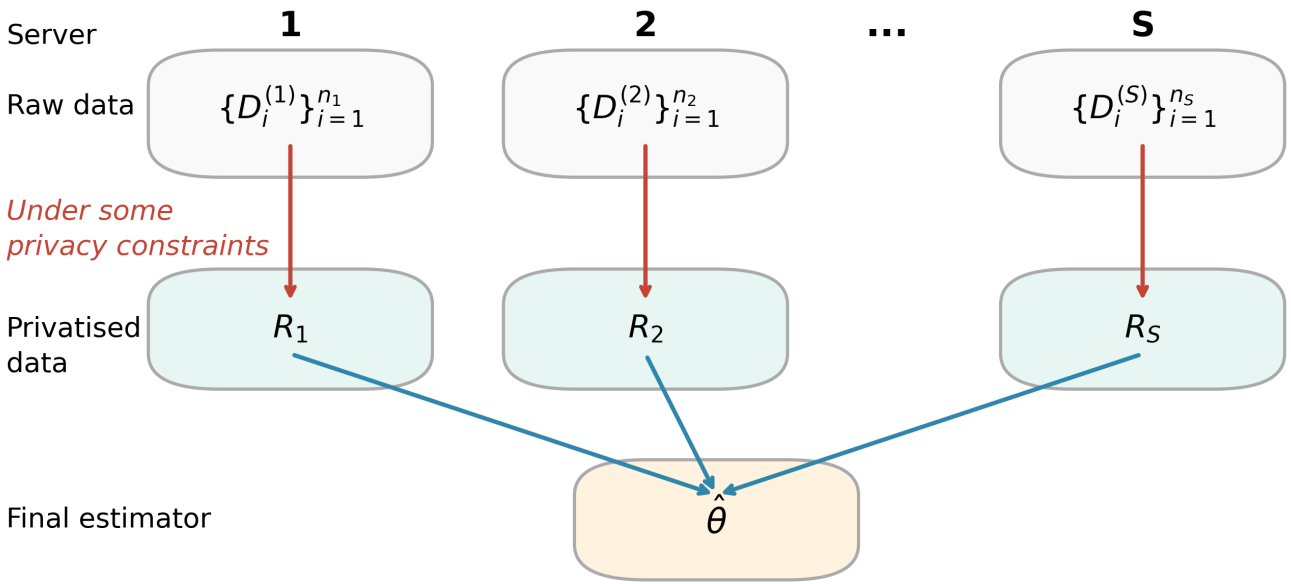
https://blogs.nvidia.com/blog/what-is-federated-learning/

Challenges

- Heterogeneity: distributions, privacy requirement types, privacy budgets.
- Communications: efficiency in aggregating and communicating siloed information.

Federated differential privacy





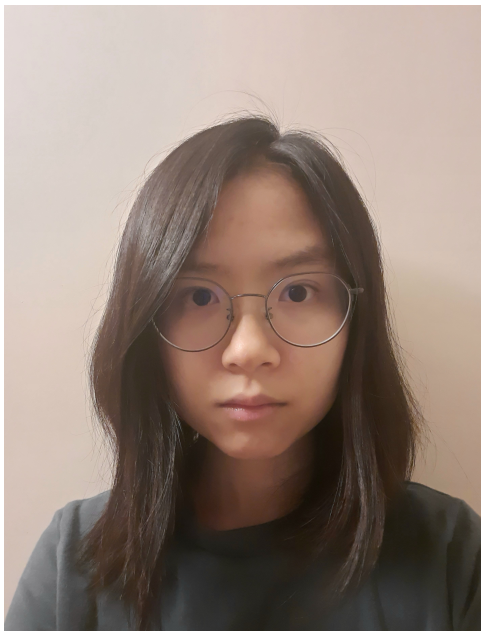
- User-level DP: Rate optimality and phase transition for user-level local differential privacy (arXiv: 2405.11923, Alexander Kent, Thomas B. Berrett and **Y.**)
- CDP: Federated transfer learning with differential privacy (arXiv: 2403.11343, Mengchu Li, Ye Tian, Yang Feng and **Y.**)
- A mixture of both: Private distributed learning in functional data (arXiv:2412.06582, Gengyu Xue, Zhenhua Lin and **Y.**)

A simple example: univariate mean estimation measured in squared loss, with S users/sites and n units of data per user.

Setting	Minimax rates	References
No privacy	$1/(Sn)$	Very easy to show
Local item-level	$1/(Sn\varepsilon^2)$	Duchi et al. (2018)
Local user-level (small n)	$1/(Sn\varepsilon^2)$	Our result
Local user-level (large n)	$e^{-S\varepsilon^2}$	Our result
Central item-level	$1/(Sn) \vee 1/(S^2n^2\varepsilon^2)$	Levy et al. (2021)
Central user-level (small n)	$1/(Sn) \vee 1/(S^2n\varepsilon^2)$	Levy et al. (2021)
Federated	$1/(Sn) \vee 1/(Sn^2\varepsilon^2)$	Our result

- Optimal Cox regression under federated differential privacy: coefficients and cumulative hazards (arXiv: 2508.196401)

- R package FDPCox available at `https://github.com/EKHung/FDPCox`.



Elly K. H. Hung
(Univ. of Warwick)

Cox regression model

At time $t \in [0, 1]$, conditional on the covariate $Z(t) \in \mathbb{R}^d$, the conditional hazard rate of the survival time $\widetilde{T}$ is

$$\lambda(t) = \frac{f_{\widetilde{T}}(t)}{S_{\widetilde{T}}(t)} = \lambda_0(t) \exp\{\beta_0^\top Z(t)\},$$

where

- $\lambda_0(\cdot)$ is an unknown baseline hazard function, with its cumulative version denoted by $\Lambda_0(\cdot)$, and
- $\beta_0 \in \mathbb{R}^d$ is an unknown regression coefficient.

Right censoring

Let $C \in \mathbb{R}_+$ be a random variable conditionally independent of $\widetilde{T}$ given $\{Z(t) : t \in [0, 1]\}$.

Cox regression model

At time $t \in [0, 1]$, conditional on the covariate $Z(t) \in \mathbb{R}^d$, the conditional hazard rate of the survival time $\widetilde{T}$ is

$$\lambda(t) = \frac{f_{\widetilde{T}}(t)}{S_{\widetilde{T}}(t)} = \lambda_0(t) \exp\{\beta_0^\top Z(t)\},$$

where

- $\lambda_0(\cdot)$ is an unknown baseline hazard function, with its cumulative version denoted by $\Lambda_0(\cdot)$, and
- $\beta_0 \in \mathbb{R}^d$ is an unknown regression coefficient.

Right censoring

Let $C \in \mathbb{R}_+$ be a random variable conditionally independent of $\widetilde{T}$ given $\{Z(t) : t \in [0, 1]\}$.

Data

Let $T = \min\{\widetilde{T},\ C\}$ be the observed time and $\Delta = \mathbb{1}\{\widetilde{T} \leq C\}$ be the (not) censoring indicator.

The observed data

$$\{(T_{s,i}, \Delta_{s,i}, \{Z_{s,i}(t), t \in [0,1]\})\}_{s,i=1}^{S,n_s}$$

are i.i.d. copies of the generic triplet $(T, \Delta, \{Z(\cdot),\ t \in [0,1]\})$.

Tasks

Estimating the regression coefficients β_0 and the cumulative hazard function $\Lambda_0(\cdot)$, subject to the federated differential privacy constraints.

Data

Let $T = \min\{\widetilde{T},\ C\}$ be the observed time and $\Delta = \mathbb{1}\{\widetilde{T} \leq C\}$ be the (not) censoring indicator.

The observed data

$$\{(T_{s,i}, \Delta_{s,i}, \{Z_{s,i}(t), t \in [0,1]\})\}_{s,i=1}^{S,n_s}$$

are i.i.d. copies of the generic triplet $(T, \Delta, \{Z(\cdot),\ t \in [0,1]\})$.

Tasks

Estimating the regression coefficients β_0 and the cumulative hazard function $\Lambda_0(\cdot)$, subject to the federated differential privacy constraints.

Federated differential privacy

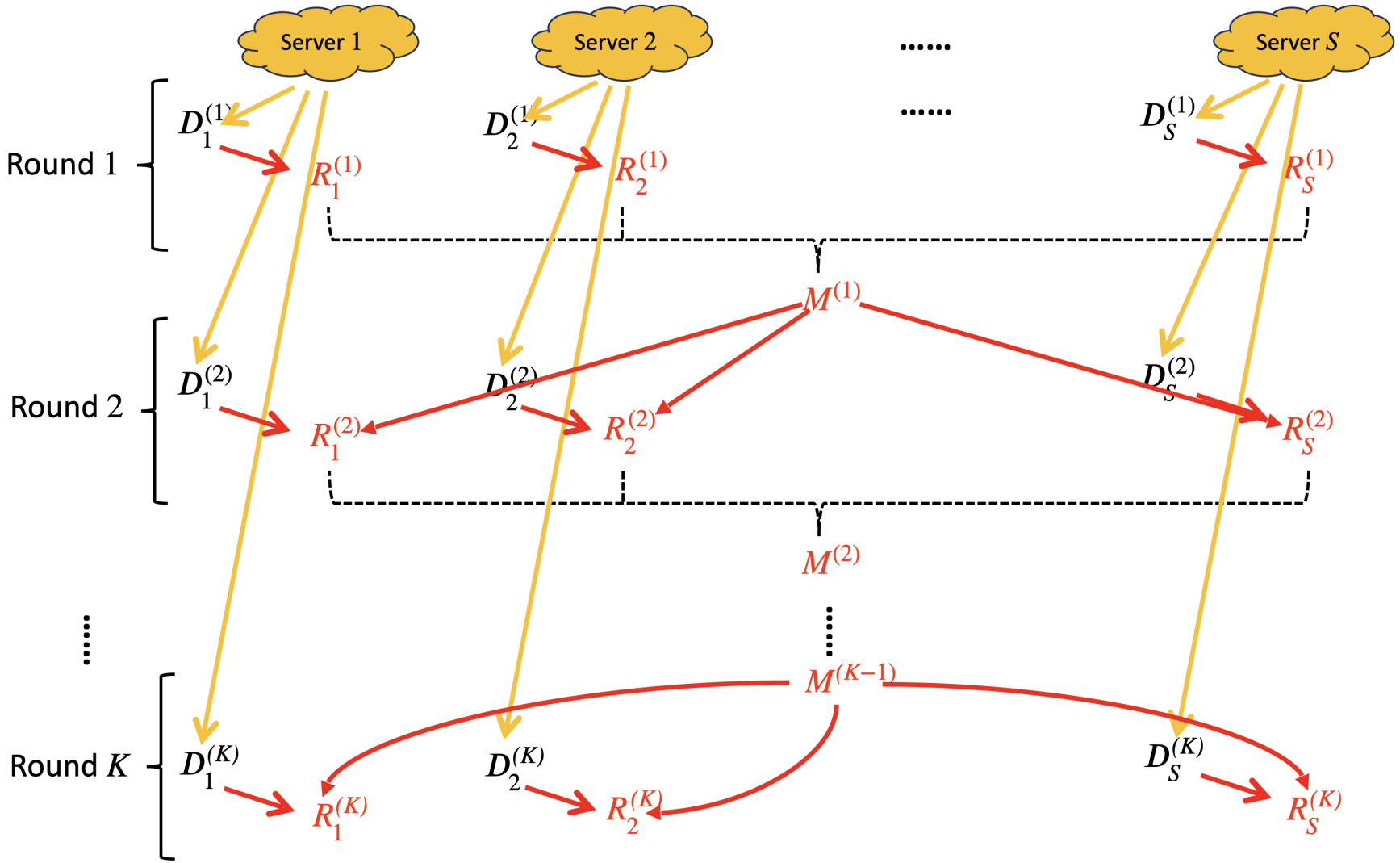
In this paper, we consider a class of K-round mechanisms.

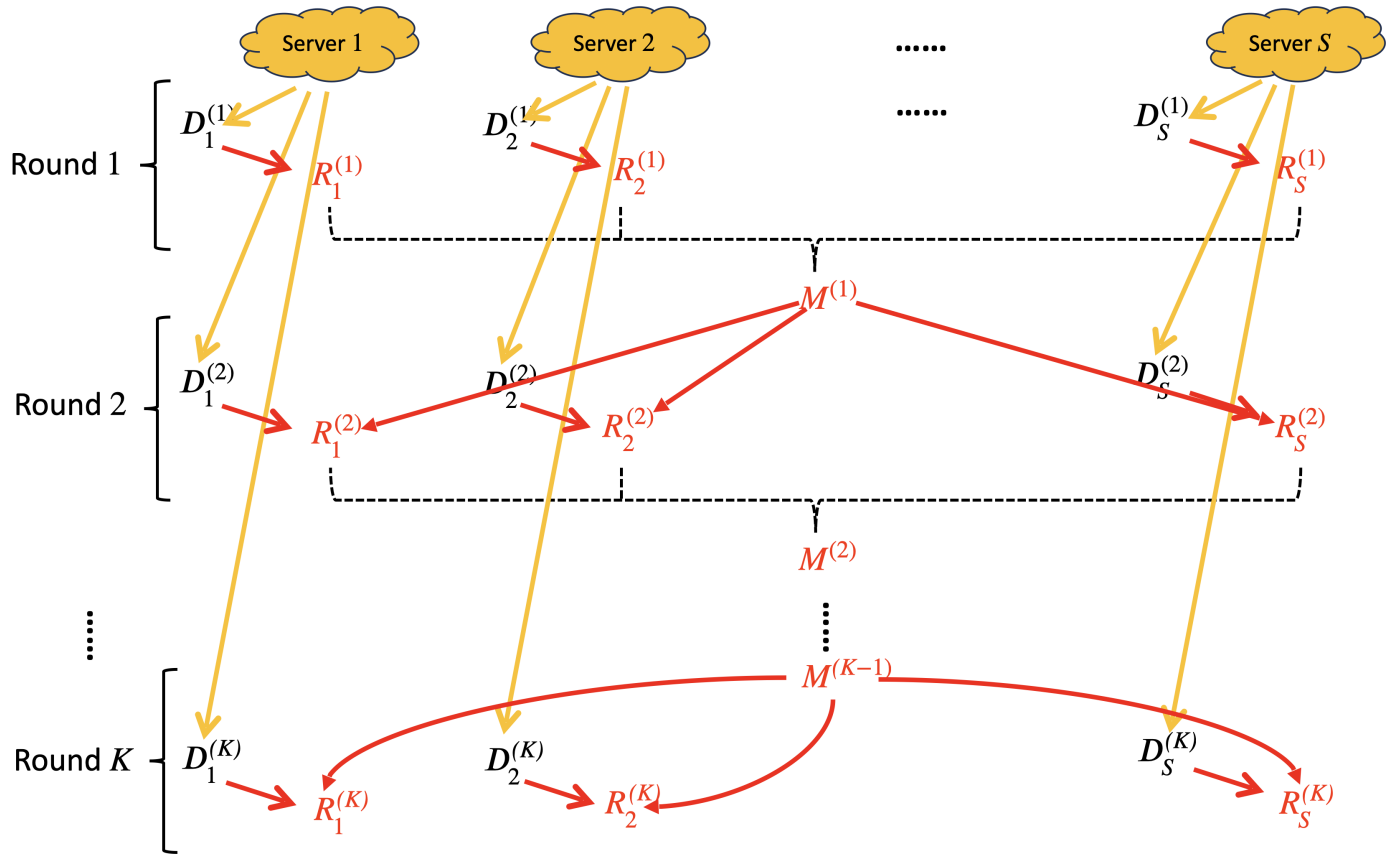
Definition [($\{\epsilon_s, \delta_s\}_{s=1}^S$, K)-FDP]

For $S \in \mathbb{N}_+$, let $\epsilon_s > 0$ and $\delta_s \geq 0$, $s \in [S]$, be privacy parameters. For $K \in \mathbb{N}_+$, we say that a privacy mechanism $Q = \{Q_s^{(k)}\}_{s,k=1}^{S,K}$ satisfies $(\{\epsilon_s, \delta_s\}_{s=1}^S, K)$-FDP, if for any $s \in [S]$ and $k \in [K]$, the data $R_s^{(k)}$ shared by the server s satisfies (ϵ_s, δ_s)-CDP, i.e.

$$Q_s^{(k)}(R_s^{(k)} \in A_s^{(k)} | M^{(k-1)}, D_s^{(k)}) \leq e^{\epsilon_s} Q_s^{(k)}(R_s^{(k)} \in A_s^{(k)} | M^{(k-1)}, (D_s^{(k)})') + \delta_s,$$

for any measurable set $A_s^{(k)}$, $M^{(k-1)} = \cup_{l=1}^{k-1} \cup_{s=1}^{S} R_s^{(l)}$, and any pairs $D_s^{(k)}, (D_s^{(k)})'$ that differing by at most one entry, where $\cup_{k=1}^{K} D_s^{(k)}$ forms a partition of the dataset at server s, $s \in [S]$.


Server 1
Server 2
.....
.....
Server S
Round 1
Round 2
Round K
$D_1^{(1)}$
$D_2^{(1)}$
$D_S^{(1)}$
$R_1^{(1)}$
$R_2^{(1)}$
$R_S^{(1)}$
$M^{(1)}$
$D_1^{(2)}$
$D_2^{(2)}$
$D_S^{(2)}$
$R_1^{(2)}$
$R_2^{(2)}$
$R_S^{(2)}$
$M^{(2)}$
$M^{(K-1)}$
$D_1^{(K)}$
$D_2^{(K)}$
$D_S^{(K)}$
$R_1^{(K)}$
$R_2^{(K)}$
$R_S^{(K)}$



Private minimax rate

$$\inf_{Q \in \mathcal{Q}} \inf_{\widehat{\theta}} \sup_{P \in \mathcal{P}} \mathbb{E}_{P,Q}\{L(\widehat{\theta}, \theta(P))\}.$$

Recall the Cox model that

$$\lambda(t) = \lambda_0(t) \exp\{\beta_0^\top Z(t)\}$$

and the data triple $(T, \Delta, \{Z(\cdot)\})$. Using the counting process representation, denote $N(t) = \mathbb{1}\{T < t, \Delta = 1\}$ and $Y(t) = \mathbb{1}\{T \geq t\}$.

Assumptions

- Compact time horizon $[0, 1]$.
- Boundedness conditions on the covariates and coefficients.
- Eigenvalues of the Hessian are homogeneous of order $1/d$.
- Baseline hazard functions are regulated.

Eigenvalues of the Hessian

For any $t \in [0, 1]$ and $\beta \in \mathbb{R}^p$, let

$$G(t, \beta) = Y(t) \exp\{\beta^\top Z(t)\}\{Z(t) - \mu(t, \beta)\}^{\otimes 2},$$

with

$$\mu(t, \beta) = \frac{\mathbb{E}[Z(t) Y(t) \exp\{\beta^\top Z(t)\}]}{\mathbb{E}[Y(t) \exp\{\beta^\top Z(t)\}]}$$

Assume that

$$\frac{\rho_-}{d} \leq \lambda_{\min}\left(\mathbb{E}\left[\int_0^1 G(s, \beta_0)\, \mathrm{d}\Lambda_0(s)\right]\right) \leq \lambda_{\max}\left(\mathbb{E}\left[\int_0^1 G(s, \beta_0)\, \mathrm{d}\Lambda_0(s)\right]\right) \leq \frac{\rho_+}{d},$$

for some constants $\rho_+ \geq \rho_- > 0$.

Baseline hazards

Assume that the hazard rate $\lambda_0(\cdot)$ exists on $[0, 1]$. For any $t \in [0, 1]$, one of the following two holds.

a. The cumulative hazard $\Lambda_0(t) = \int_0^t \lambda_0(s)\,\mathrm{d}s < \infty$.

b. There exist absolute constants $C_\lambda, p_0 > 0$, such that $\lambda_0(t) < C_\lambda$ and $\mathbb{P}\{Y(1) = 1\} = p_0$.

Remark. a. is used for estimating β_0 and b. is used for estimating $\Lambda_0(\cdot)$.

Baseline hazards

Assume that the hazard rate $\lambda_0(\cdot)$ exists on $[0, 1]$. For any $t \in [0, 1]$, one of the following two holds.

a. The cumulative hazard $\Lambda_0(t) = \int_0^t \lambda_0(s)\,\mathrm{d}s < \infty$.

b. There exist absolute constants $C_\lambda, p_0 > 0$, such that $\lambda_0(t) < C_\lambda$ and $\mathbb{P}\{Y(1) = 1\} = p_0$.

Remark. a. is used for estimating β_0 and b. is used for estimating $\Lambda_0(\cdot)$.

Optimal estimation of the regression coefficients

Theorem Denote by $\mathcal{P}$ the class of distributions satisfying the assumptions, and denote $\mathcal{Q}$ the class of $(\{(\epsilon_s, \delta_s)\}_{s\in[S]}, K)$-FDP mechanisms for all $K \in \mathbb{N}_+$. Suppose that $\delta_s \log(1/\delta_s) \lesssim \epsilon_s^2/d$, for $s \in [S]$. We have that

$$\inf_{Q\in\mathcal{Q}} \inf_{\widehat{\beta}} \sup_{P\in\mathcal{P}} \mathbb{E}_{Q,P}\{\|\widehat{\beta} - \beta_0\|_2^2\} \asymp L \frac{d^2}{\sum_{s=1}^S \min\{n_s,\ n_s^2\epsilon_s^2/d\}},$$

for some $L \in [1, \log^2(\sum_{s=1}^S n_s) \max_{s\in[S]} \log(1/\delta_s) \log^2(n_s)]$.

$$\inf_{Q\in\mathcal{Q}} \inf_{\widehat{\beta}} \sup_{P\in\mathcal{P}} \mathbb{E}_{Q,P}\{\|\widehat{\beta}-\beta_0\|_2^2\} \asymp L\frac{d^2}{\sum_{s=1}^{S}\min\{n_s,\ n_s^2\epsilon_s^2/d\}}$$

Remarks

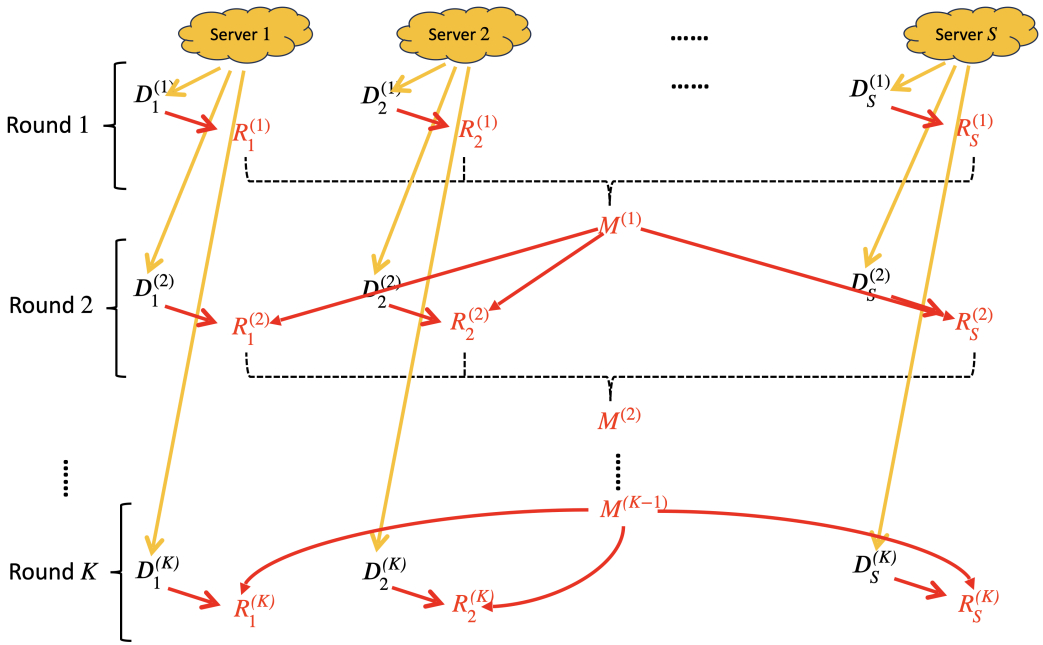
- The upper bound is achieved by choosing K to be a logarithmic factor and the lower bound is achieved by setting $K=1$.
- The lower bound is proved via the van Trees inequality and score attack arguments.
- In a homogeneous setting, $n_s = n$ and $\epsilon_s = \epsilon$, for $s \in [S]$, we have the rate

$$\max\left\{\frac{d^2}{Sn},\ \frac{d^3}{Sn^2\epsilon^2}\right\}.$$

- When $S=1$, we have the central DP rate

$$\max\left\{\frac{d^2}{n},\ \frac{d^3}{n^2\epsilon^2}\right\}.$$

FDP Cox algorithm



1. Obtain gradients from each batch of each server.
2. Privatise the gradients by adding appropriate Gaussian noise.
3. Aggregate the gradients from all servers in each step by appropriate weights.
4. Update the estimator and truncate it.
5. Output the final estimator.

Question: Does having public covariates lead to an accuracy improvement in estimating the Cox regression coefficients?

Definition (Label-CDP) For $\epsilon > 0$ and $\delta \geq 0$, a privacy mechanism M is an (ϵ, δ)-label-CDP mechanism for survival data, if it is a conditional distribution satisfying

$$\begin{aligned}&\mathbb{P}\{M(\{T_i, \Delta_i, Z_i\}_{i\in[n]}) \in A | (T_i, \Delta_i, Z_i)_{i\in[n]}\}\\ \leq& e^{\epsilon}\mathbb{P}\{M(\{T_i', \Delta_i', Z_i\}_{i\in[n]}) \in A | (T_i', \Delta_i', Z_i)_{i\in[n]}\} + \delta,\end{aligned}$$

for all measurable set A, all possible $\{Z_i\}_{i\in[n]}$ and all possible $\{(T_i, \Delta_i, T_i', \Delta_i')\}_{i\in[n]}$ such that $\sum_{i=1}^n \mathbb{1}\{(T_i, \Delta_i) \neq (T_i', \Delta_i')\} \leq 1$.

A label-FDP definition can be made correspondingly.

Question: Does having public covariates lead to an accuracy improvement in estimating the Cox regression coefficients?

Definition (Label-CDP) For $\epsilon > 0$ and $\delta \geq 0$, a privacy mechanism M is an (ϵ, δ)-label-CDP mechanism for survival data, if it is a conditional distribution satisfying

$$\begin{aligned}&\mathbb{P}\{M(\{T_i, \Delta_i, Z_i\}_{i\in[n]}) \in A | (T_i, \Delta_i, Z_i)_{i\in[n]}\}\\ \leq& e^{\epsilon}\mathbb{P}\{M(\{T_i', \Delta_i', Z_i\}_{i\in[n]}) \in A | (T_i', \Delta_i', Z_i)_{i\in[n]}\} + \delta,\end{aligned}$$

for all measurable set A, all possible $\{Z_i\}_{i\in[n]}$ and all possible $\{(T_i, \Delta_i, T_i', \Delta_i')\}_{i\in[n]}$ such that $\sum_{i=1}^n \mathbb{1}\{(T_i, \Delta_i) \neq (T_i', \Delta_i')\} \leq 1$.

A label-FDP definition can be made correspondingly.

Question: Does having public covariates lead to an accuracy improvement in estimating the Cox regression coefficients?

Definition (Label-CDP) For $\epsilon > 0$ and $\delta \geq 0$, a privacy mechanism M is an (ϵ, δ)-label-CDP mechanism for survival data, if it is a conditional distribution satisfying

$$\begin{aligned}&\mathbb{P}\{M(\{T_i, \Delta_i, Z_i\}_{i\in[n]}) \in A | (T_i, \Delta_i, Z_i)_{i\in[n]}\}\\ \leq& e^{\epsilon}\mathbb{P}\{M(\{T_i', \Delta_i', Z_i\}_{i\in[n]}) \in A | (T_i', \Delta_i', Z_i)_{i\in[n]}\} + \delta,\end{aligned}$$

for all measurable set A, all possible $\{Z_i\}_{i\in[n]}$ and all possible $\{(T_i, \Delta_i, T_i', \Delta_i')\}_{i\in[n]}$ such that $\sum_{i=1}^{n} \mathbb{1}\{(T_i, \Delta_i) \neq (T_i', \Delta_i')\} \leq 1$.

A label-FDP definition can be made correspondingly.

Does having public covariates lead to an accuracy improvement in estimating the Cox regression coefficients?

Theorem Denote by $\mathcal{P}$ the class of distributions satisfying all assumptions and denote $\mathcal{Q}$ the class of $(\{(\epsilon_s, \delta_s)\}_{s\in[S]}, K)$-FDP mechanisms for all $K \in \mathbb{N}_+$. Suppose that $\delta_s \log(1/\delta_s) \lesssim \epsilon_s^2/d$, for $s \in [S]$. We have that

$$\inf_{Q\in\mathcal{Q}} \inf_{\widehat{\beta}} \sup_{P\in\mathcal{P}} \mathbb{E}_{Q,P}\{\|\widehat{\beta} - \beta_0\|_2^2\} \asymp L \frac{d^2}{\sum_{s=1}^S \min\{n_s,\ n_s^2\epsilon_s^2/d\}},$$

for some $L \in [1, \log^2(\sum_{s=1}^S n_s) \max_{s\in[S]} \log(1/\delta_s) \log^2(n_s)]$.

Remarks

- Sensitivity.
- Beyond public covariates.
- Lower bound proofs.

Does having public covariates lead to an accuracy improvement in estimating the Cox regression coefficients?

Theorem Denote by $\mathcal{P}$ the class of distributions satisfying all assumptions and denote $\mathcal{Q}$ the class of $(\{(\epsilon_s, \delta_s)\}_{s\in[S]}, K)$-FDP mechanisms for all $K \in \mathbb{N}_+$. Suppose that $\delta_s \log(1/\delta_s) \lesssim \epsilon_s^2/d$, for $s \in [S]$. We have that

$$\inf_{Q\in\mathcal{Q}} \inf_{\widehat{\beta}} \sup_{P\in\mathcal{P}} \mathbb{E}_{Q,P}\{\|\widehat{\beta} - \beta_0\|_2^2\} \asymp L \frac{d^2}{\sum_{s=1}^S \min\{n_s,\ n_s^2\epsilon_s^2/d\}},$$

for some $L \in [1, \log^2(\sum_{s=1}^S n_s) \max_{s\in[S]} \log(1/\delta_s) \log^2(n_s)]$.

Remarks

- Sensitivity.
- Beyond public covariates.
- Lower bound proofs.

Optimal estimation of the cumulative hazard function

An important question in the survival analysis is to understand the survival function of the event. In the Cox model, it can be written as

$$S_{\widetilde{T}}(t) = \exp\left\{-\int_0^t \exp\{\beta_0^\top Z(s)\}\, \mathrm{d}\Lambda_0(s)\right\}, \quad t \in [0, 1].$$

We have obtained an estimator of β_0. To estimate $S_{\widetilde{T}}(\cdot)$, a renowned estimator of $\Lambda_0(\cdot)$ is the Breslow estimator, which written in counting process representation is,

$$\widehat{\Lambda}(t) = \sum_{i=1}^n \int_0^t \frac{\mathrm{d}N_i(s)}{\sum_{j=1}^n Y_j(s) \exp\{\widehat{\beta}^\top Z_j(s)\}}, \quad t \in [0, 1].$$

In the non-private case, $\sqrt{n}(\widehat{\Lambda} - \Lambda_0)$ converges to a zero-mean Gaussian process.

An important question in the survival analysis is to understand the survival function of the event. In the Cox model, it can be written as

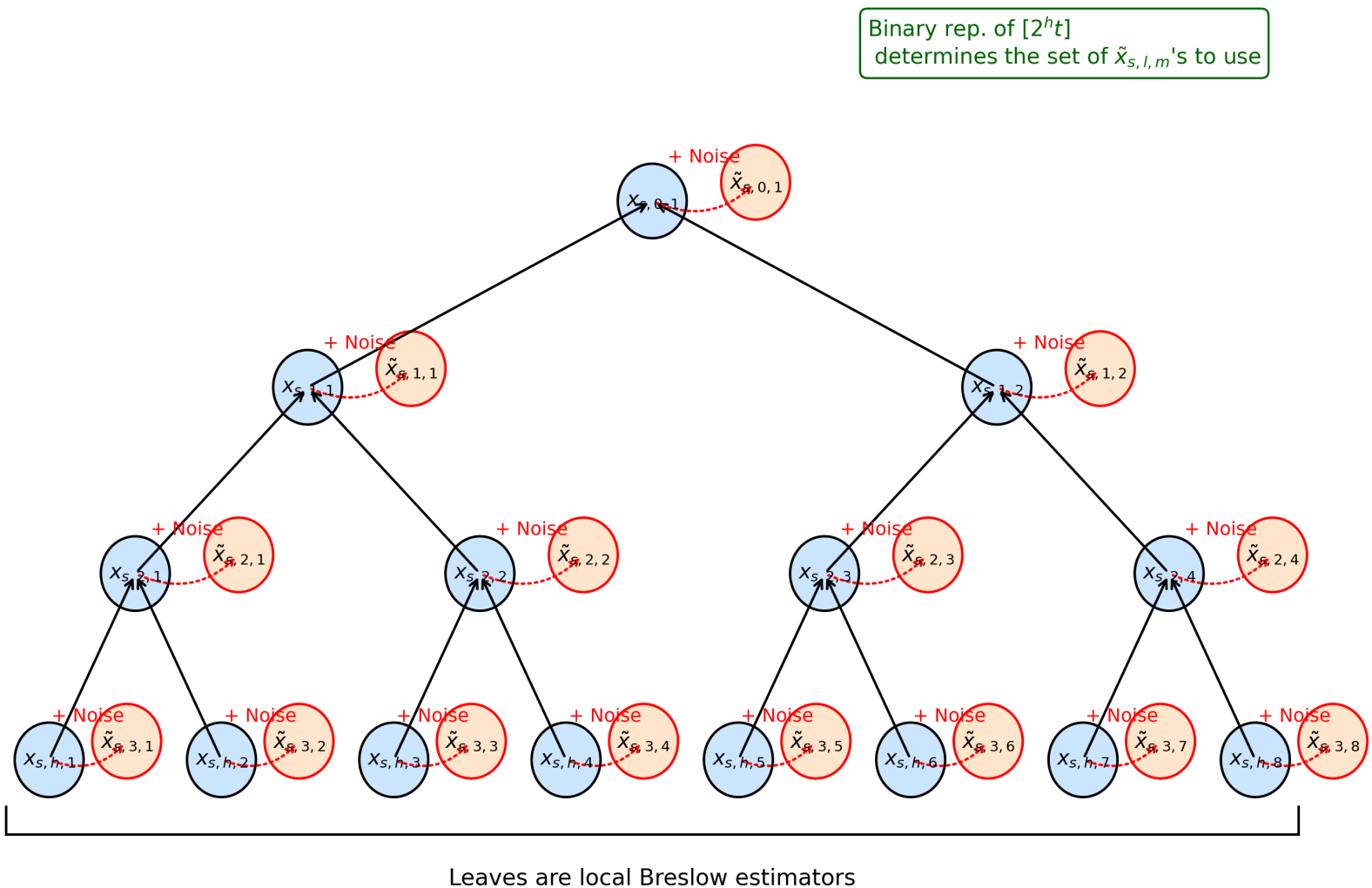
$$S_{\widetilde{T}}(t) = \exp\left\{-\int_0^t \exp\{\beta_0^\top Z(s)\}\, \mathrm{d}\Lambda_0(s)\right\}, \quad t \in [0, 1].$$

We have obtained an estimator of β_0. To estimate $S_{\widetilde{T}}(\cdot)$, a renowned estimator of $\Lambda_0(\cdot)$ is the Breslow estimator, which written in counting process representation is,

$$\widehat{\Lambda}(t) = \sum_{i=1}^n \int_0^t \frac{\mathrm{d}N_i(s)}{\sum_{j=1}^n Y_j(s) \exp\{\widehat{\beta}^\top Z_j(s)\}}, \quad t \in [0, 1].$$

In the non-private case, $\sqrt{n}(\widehat{\Lambda} - \Lambda_0)$ converges to a zero-mean Gaussian process.

FDP-Breslow algorithm



Theorem (informal) Under regularity assumptions and with suitable inputs, we have that

- the output of the `FDP-Breslow` algorithm $\widehat{\Lambda}$ satisfies $(\{\epsilon_s, \delta_s\}_{s\in S}, 1)$-FDP, and
- that

$$\mathbb{E}\left[\sup_{t\in[0,1]} |\widehat{\Lambda}(t) - \Lambda_0(t)|\right] \lesssim_{\log} \frac{1}{\sqrt{\sum_{s=1}^{S} \min\{n_s,\ n_s^2\epsilon_s^2\}}} + \mathbb{E}\{\|\widehat{\beta} - \beta_0\|\}$$

Remarks

- Inputs: $\widehat{\beta}$ and $\hat{p}$.
- Non-interactive.

Theorem (informal) Under regularity assumptions and with suitable inputs, we have that

- the output of the `FDP-Breslow` algorithm $\widehat{\Lambda}$ satisfies $(\{\epsilon_s, \delta_s\}_{s\in S}, 1)$-FDP, and
- that

$$\mathbb{E}\left[\sup_{t\in[0,1]} |\widehat{\Lambda}(t) - \Lambda_0(t)|\right] \lesssim_{\log} \frac{1}{\sqrt{\sum_{s=1}^{S} \min\{n_s,\ n_s^2\epsilon_s^2\}}} + \mathbb{E}\{\|\widehat{\beta} - \beta_0\|\}$$

Remarks

- Inputs: $\widehat{\beta}$ and $\hat{p}$.
- Non-interactive.

Proposition Let $\mathcal{H}$ be the set of cumulative hazard functions that satisfy the assumptions. Fix privacy parameters $\{\epsilon_s, \delta_s\}_{s\in[S]}$ such that $\delta_s \le n_s\epsilon_s^2$ for all $s \in [S]$. Let $\mathcal{Q}$ be the class of $(\{(\epsilon_s, \delta_s)\}_{s\in[S]}, 1)$-FDP estimators. It then holds that

$$\inf_{\widehat{\Lambda}\in\mathcal{Q}} \sup_{\Lambda_0\in\mathcal{H}} \mathbb{E}\left[\sup_{t\in[0,1]} |\widehat{\Lambda}(t) - \Lambda_0(t)|\right] \gtrsim \frac{1}{\sqrt{\sum_{s=1}^{S} \min\{n_s,\ n_s^2\epsilon_s^2\}}}.$$

Remarks

- Homogeneous minimax rate

$$\max\left\{\frac{1}{\sqrt{nS}}, \frac{1}{\sqrt{n^2S\epsilon^2}}\right\}.$$

- Lower bound proof: 1) coupling methods, and 2) challenges in the cumulative hazard functions.
- Other loss functions, including the survival function.
- Interactive mechanisms: we can show a lower bound for general K-round interactive mechanism

$$\frac{1}{\sqrt{\min\{\sum_{s=1}^{S} n_s^2\epsilon_s^2,\ \sum_{s=1}^{S} n_s\}}}.$$

Proposition Let $\mathcal{H}$ be the set of cumulative hazard functions that satisfy the assumptions. Fix privacy parameters $\{\epsilon_s, \delta_s\}_{s\in[S]}$ such that $\delta_s \le n_s\epsilon_s^2$ for all $s \in [S]$. Let $\mathcal{Q}$ be the class of $(\{(\epsilon_s, \delta_s)\}_{s\in[S]}, 1)$-FDP estimators. It then holds that

$$\inf_{\widehat{\Lambda}\in\mathcal{Q}} \sup_{\Lambda_0\in\mathcal{H}} \mathbb{E}\left[\sup_{t\in[0,1]} |\widehat{\Lambda}(t) - \Lambda_0(t)|\right] \gtrsim \frac{1}{\sqrt{\sum_{s=1}^S \min\{n_s,\ n_s^2\epsilon_s^2\}}}.$$

Remarks

- Homogeneous minimax rate

$$\max\left\{\frac{1}{\sqrt{nS}},\ \frac{1}{\sqrt{n^2S\epsilon^2}}\right\}.$$

- Lower bound proof: 1) coupling methods, and 2) challenges in the cumulative hazard functions.
- Other loss functions, including the survival function.
- Interactive mechanisms: we can show a lower bound for general K-round interactive mechanism

$$\frac{1}{\sqrt{\min\{\sum_{s=1}^S n_s^2\epsilon_s^2,\ \sum_{s=1}^S n_s\}}}.$$

Numerical experiments

Rényi differential privacy in CDP

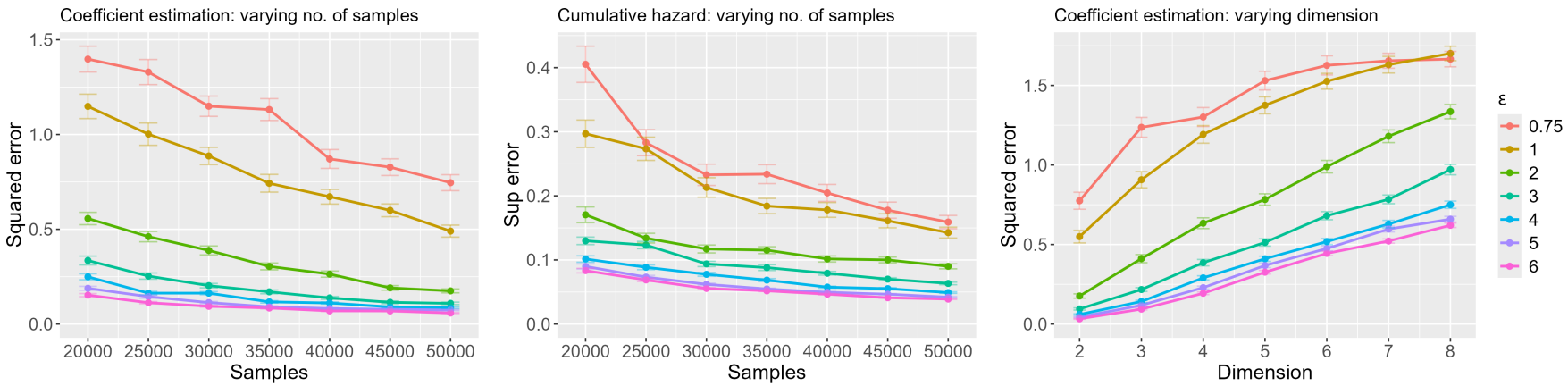


Figure: Simulation results for CDP Cox regression coefficients (left panel) and cumulative hazard (middle panel) estimation, with varying sample sizes and privacy budgets; and for CDP Cox regression coefficients estimation with varying dimensions (right panel).

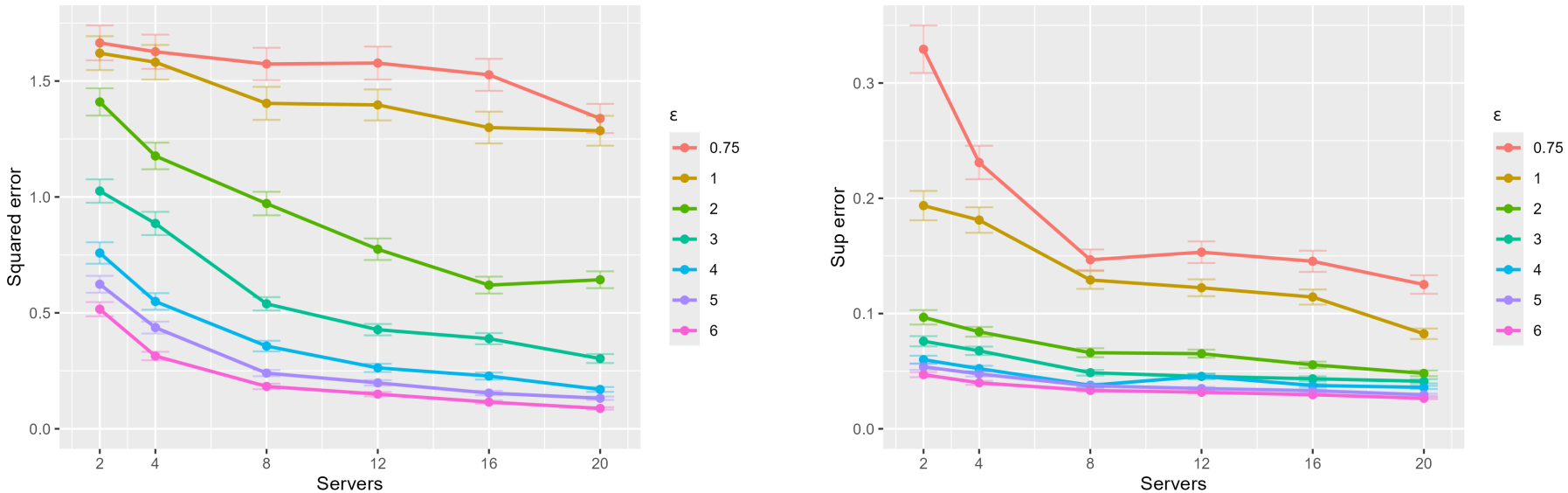


Figure: Simulations results from `FDP-Cox` (left) and `FDP-Breslow` (right), varying the number of servers and the ϵ privacy budget.

Censoring rate analysis

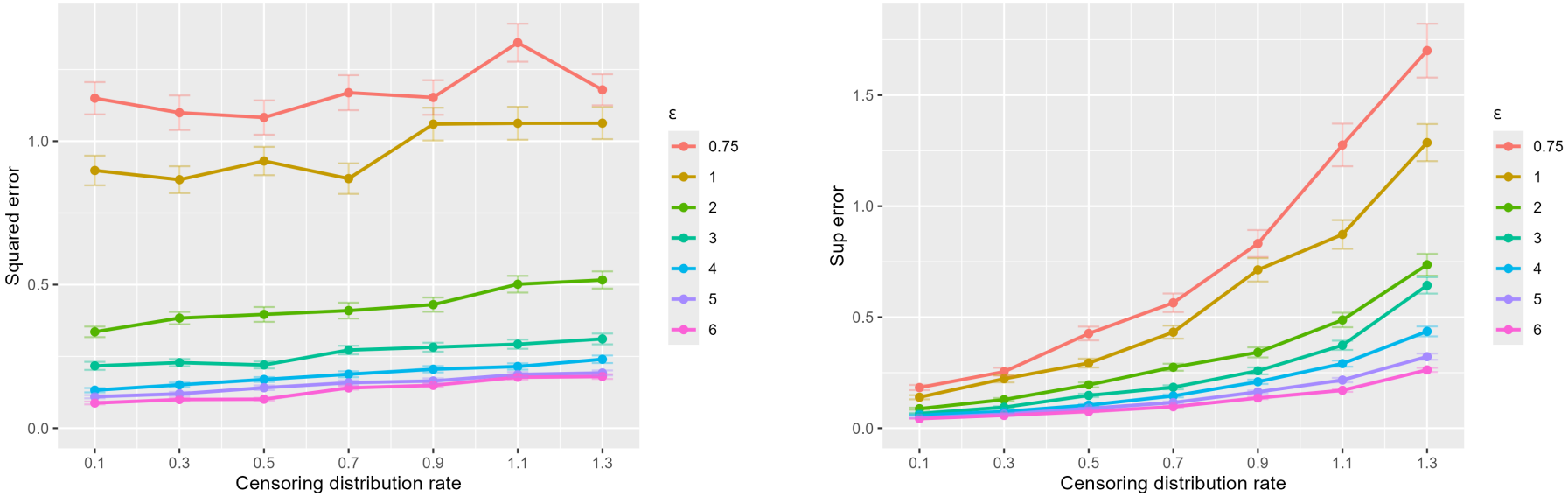


Figure: Effect of censoring on β_0 (left) and Λ_0 estimation (right), by varying the censoring distribution as $\mathrm{Exp}(\alpha)$, where $\alpha \in \{0.1, 0.3, \ldots, 1.3\}$.

α	0.1	0.3	0.5	0.7	0.9	1.1	1.3
$\mathbb{P}(\Delta = 0 \mid T < 1)$	0.090	0.229	0.330	0.410	0.471	0.520	0.561
$\mathbb{P}(Y(1) = 1)$	0.33	0.273	0.223	0.183	0.150	0.123	0.100

Table: Monte Carlo estimates (from 10^6 samples) of $\mathbb{P}(\Delta = 0 \mid T < 1)$ and $\mathbb{P}(Y(1) = 1)$ under different rates for the censoring distribution $\mathrm{Exp}(\alpha)$.

- Fully-interactive lower bounds.
- Transfer learning.
- Recurrent events and user-level differential privacy.
- ...