Privacy constrained functional estimation

Botond Szabó (Bocconi University)

Advances in High/Infinite-dimensional Inference Workshop
Verona, Italy, 08. 11. 2024.

Co-authors

Thibault Randrianarisoa
(Toronto)

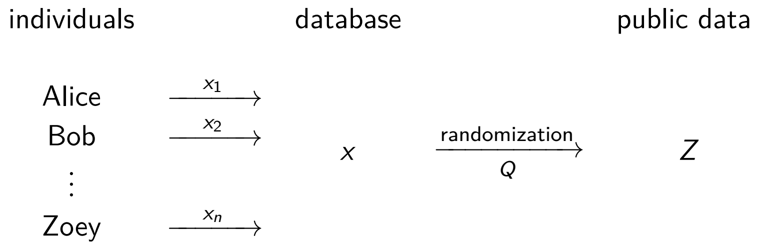
Lukas Steinberger
(Vienna)

Outline

- Introduction to privacy constrained inference
- Functional estimation: smooth vs atomic case
- Privacy constrained non-parametric inference
- Privacy constrained plug-in estimator
- Adaptation
- Summary

Introduction to α-differential privacy

Idea behind Differential Privacy



Distribution of Z should not depend too much on any individual contribution x_i.

Definition: α-DP

Definition: Let $X = (X_i)_{i=1\ldots,n}$ denote the original data and $Z = (Z_i)_{i=1,\ldots,n}$ denote its sanitized version. This data I obeys the local α-differential privacy constraint if

$$\sup_{A} \sup_{x,x':d_0(x,x')=1} \frac{Pr(Z \in A|X = x)}{Pr(Z \in A|X = x')} \leq e^{\alpha},$$

where $d_0(x, x') = |\{i : x_i \neq x_i'\}|$ denotes the Hamming distance.

Definition: α-DP

Definition: Let $X = (X_i)_{i=1\dots,n}$ denote the original data and $Z = (Z_i)_{i=1,\dots,n}$ denote its sanitized version. This data I obeys the local α-differential privacy constraint if

$$\sup_{A} \sup_{x,x':d_0(x,x')=1} \frac{Pr(Z \in A|X = x)}{Pr(Z \in A|X = x')} \leq e^{\alpha},$$

where $d_0(x, x') = |\{i : x_i \neq x'_i\}|$ denotes the Hamming distance.

Idea: The conditional distribution of Z given $X = x$ does not depend too much on the data of the i-th individual in the database, thereby protecting its privacy.

Strength: Smaller α denotes stronger privacy protection.

Relaxed version: (α, δ) differential privacy: for all A and $d_0(x, x') = 1$

$$Pr(Z \in A|X = x) \leq e^{\alpha} Pr(Z \in A|X = x') + \delta.$$

Properties

- "local" means that there is no trusted third party available for data collection and processing, see . Evfimievski (2003)
- **Protocols:**
  - non-interactive: Z_i is generated from X_i independently.
  - sequentially interactive: ith person has access to $Z_1, ..., Z_{i-1}$ when generating Z_i.
- **Random perturbation:**
  - Laplace: α-differentiable private mechanism
  - Gauss: (α, δ)-differentiable private mechanism
- **Applications:** Apple ($2 \leq \alpha \leq 8$), Google ($0.6 \leq \alpha \leq 10$, $0 \leq \delta \leq 10^{-10}$), Microsoft ($1.67 \leq \alpha \leq 4.7$, $0 \leq \delta \leq 10^{-5}$), US Census Bureau (ounty Business Patterns: $\alpha = 34.9$, $\delta = 10^{-5}$; 2020 Decennial Census: $13.64 \leq \alpha \leq 49.2$, $\delta = 10^{-5}$).

Literature review

Parametric models: Dwork et al (2006), Smith (2008), Duchi et al (2014), Kairouz et al. (2016), Kamath et al (2018), Cai et al (2020)

Nonparametric models:

- **density estimation:** global privacy Wasserman and Zhou (2010), Hall et al (2013); local Duchi et al (2013, 2018), Butucea (2020)
- **regression:** methodology Smith (2021), Golowich (2021) theory Gyorfi and Kroll (2023).

Semi-parametric problems:

- Linear functionals Rohde and Steinberger (2018)
- Integrated square $\int f^2(x)dx$, Butucea et al (2023)

BUT! No general approach, case-by-case studies.

Model and examples

Density estimation problem: $X_1, ..., X_n \overset{iid}{\sim} f$, with

$$f \in \mathcal{W}_p := \left\{ f \in C^p[0,1] : \; f \geq 0, \; \int_0^1 f = 1, \; \|f\|_{(\infty,p,\lambda)} < M \right\},$$

$p \in \mathbb{N}$, where for $1 \leq q \leq \infty$ and measures $\lambda = (\lambda_0, \ldots, \lambda_p)$ on $[0,1]$,

$$\|f\|_{(q,p,\lambda)} = \sum_{j=0}^{p} \left(\int_0^1 (f^{(j)})^q d\lambda_j \right)^{1/q}.$$

Semi-parametric model: Consider functionals $\Lambda : C^p \to \mathbb{R}$, s.t. for some $0 \leq m < p$,

$$\Lambda(f+h) = \Lambda(f) + T_f(h) + O(\|h\|^2_{(2,m,\lambda)}), \tag{1}$$

where for $f \in \mathcal{W}_p$, $h \in C^p[0,1]$ with $\|h\|_{(\infty,m)}$ small enough and T_f a bounded linear functional on $C^p[0,1]$, see Goldstein & Messer (1992).

Functional

In view of the Hahn-Banach and Riesz representation theorems

$$T_f(h) = \sum_{j=0}^{p} \int_0^1 h^{(j)} d\mu_j,$$

where μ_j is a finite signed Borel measures on $[0,1]$ (possibly depending on f).

Cases:

- **Smooth functionals:** $T_f(h) = \int h\omega_f$, $\forall f \in \mathcal{W}_p$, with $\sup_{f\in\mathcal{W}_p} \|\omega_f\|_\infty < \infty$.
- **Atomic functionals:** of index $s \in \{0,..,p\}$, where

$$T_f(h) = \sum_{j=0}^{s_f} \int_0^1 h^{(j)} d\mu_{j,f}$$

with $\mu_{s_f,f}$ having a discrete component $\delta_{s_f,f}$, and $s = \max_{f\in\mathcal{W}_p} s_f$.

Examples

Atomic:

- $\Lambda(f) = f^{(r)}(x_0)$. Rate: $n^{-(p-r)/(2p+1)}$
- $\Lambda(f) = \Lambda(f) = \int_0^1 |f^{(m)}|^2$ for $m \in \mathbb{N}_+$. Rate: $n^{-\frac{p-m+1}{2p+1}}$
- Fisher information: $\Lambda(f) = \int_0^1 (f')^2/f$. Rate: $n^{-p/(2p+1)}$.

Smooth:

- $\Lambda(f) = \Lambda(f) = \int_0^1 |f|^q$.
- Entropy: $\Lambda(f) = \int_0^1 f \log f$.

Privacy constrained estimation: Non-adaptive setting

Privatized data

$$Z_{ijk} = \begin{cases} \mathbf{B}_{k,d,\xi^{(j_0)}}(X_i) + \sigma_{j_0-1} Y_{i(j_0-1)k}, & \text{if } j = j_0 - 1, k \in \mathcal{M}_{j_0-1}, \\ \psi_{j,k}(X_i) + \sigma_j Y_{ijk}, & \text{if } j \geq j_0, k \in \mathcal{M}_j, \end{cases}$$

where $Y_{ijk} \overset{iid}{\sim} Lap(1)$, $\psi_{j,k}$ are the spline wavelet basis, $\mathbf{B}_{k,d,\xi^{(j_0)}}$ the B-Splines up to order d, and

$$\sigma_{\alpha,j_0-1} = \frac{C_d}{\alpha} 2^{j_0/2}, \quad \sigma_{\alpha,j} = \frac{C_d \|\psi\|_\infty}{\alpha} \frac{a}{a-1} j^a 2^{j/2}.$$

Lemma: The privacy mechanism defined above is locally α-differentially private.

Wavelet coefficients: privatized empirical wavelet coefficients $\bar{Z}_{jk} = n^{-1} \sum_{i=1}^{n} Z_{ijk}$

Density estimation:

$$\hat{f}_n = \hat{f}_n^{j_n} = \sum_{k \in \mathcal{M}_{j_0-1}} \bar{Z}_{(j_0-1)k} \tilde{\psi}_{j_0-1,k} + \sum_{j=j_0}^{j_n} \sum_{k \in \mathcal{M}_j} \bar{Z}_{jk} \tilde{\psi}_{j,k}.$$

Point-wise and L_2-convergence For $2^{j_n} \asymp (n\alpha^2 \log^{-2a} n)^{\frac{1}{2p+2}} \wedge n^{\frac{1}{2p+1}}$ we have

$$\max\left(\mathbb{E}_{Q\mathbb{P}_f} |\hat{f}_n^{(q)}(x_0) - f^{(q)}(x_0)|^2,\ \mathbb{E}_{Q\mathbb{P}_f} \|\hat{f}_n^{(q)} - f^{(q)}\|_{L_2(G)}^2\right)$$
$$\leq C_{d,q,M} (n\alpha^2 \log^{-2a} n)^{-\frac{2(p-q)}{2p+2}} \vee n^{-\frac{2(p-q)}{2p+1}}.$$

Convergence rate for atomic functionals

Theorem [estimation atomic]: Let $f \in \mathcal{W}_p$, $p \leq d+1$ and suppose Λ is an atomic functional of index s. Under some mild technical conditions, the plug-in estimator $\widehat{\Lambda(f)} = \Lambda(\hat{f}^{j_n})$ with $2^{j_n} \asymp (n\alpha^2 \log^{-2a} n)^{\frac{1}{2p+2}} \wedge n^{\frac{1}{2p+1}}$ converges towards $\Lambda(f)$ at rate

$$(n\alpha^2 \log^{-2a} n)^{-\frac{p-s}{2p+2}} \vee n^{-\frac{p-s}{2p+1}}.$$

Remark: Derived matching lower bound for $\alpha = O(1)$.

Convergence rate for smooth functionals

Theorem [estimation smooth]: Let $f \in \mathcal{W}_p$ and suppose Λ is a smooth functional with $m \geq 0$, such that $\omega_f \in \mathcal{W}_1$ satisfy $\sup_f \|\omega_f\|_\infty < \infty$. Under some mild technical conditions, the plug-in estimator $\widehat{\Lambda(f)} = \Lambda(\hat{f}_n)$ with $a > 0$ and

$$\left(n \wedge (n\alpha^2)\right)^{1/2p} \leq 2^{j_n} \leq \left[\log^{-a/(m+1)}(n\alpha^2)(n\alpha^2)^{1/(4m+4)}\right] \wedge \left[\log^{-a}(n) n^{1/(4m+3)}\right]$$

converges towards $\Lambda(f)$ at rate

$$n^{-1/2} \vee (n\alpha^2)^{-1/2}.$$

Remark: Derived matching lower bound for $\alpha = O(1)$.

Adaptation

Lepski's type method

Grid: $x_t = t/M_n,\ t = 0, ..., M_n$, for $M_n \gtrsim n^{4/3}$

Data driven threshold:

$$\hat{j}_n = \min\{j \in \mathcal{J} : \|(\hat{f}_n^j)^{(q)} - (\hat{f}_n^l)^{(q)}\|^2_{L^2[0,1]} \vee \max_{x_t,\, t=0,\ldots,M_n} |(\hat{f}_n^j)^{(q)}(x_t) - (\hat{f}_n^l)^{(q)}(x_t)|^2$$
$$\leq \tau n^{-1} 2^{2lq} l(2^l + 2^{2l} l^{2a} \alpha^{-2}),$$
$$\forall l > j,\ l \in \mathcal{J}, \forall q \in \{0, 1, .., p\}\},$$

Estimator (Lepski): $\hat{f}_n(x) = \hat{f}_n^{\hat{j}_n}(x)$.

Adaptation: density estimation

Theorem [adaptation density]: The estimator $\hat{f}_n(x) = \hat{f}_n^{\hat{j}_n}(x)$ satisfies that for all $q+1 \leq p$ and $x \in [0,1]$

$$\sup_{f \in \mathcal{W}^p(L) \cap \|f\|_\infty \leq L} \mathbb{E}_{Q\mathbb{P}_f} \|\hat{f}_n^{(q)} - f^{(q)}\|_{L^2[0,1]} \vee \mathbb{E}_f |\hat{f}_n^{(q)}(x) - f(x)^{(q)}|$$

$$\lesssim (n\alpha^2 \log^{-(1+2a)} n)^{-\frac{p-q}{2p+2}} \vee (n/\log n)^{-\frac{p-q}{2p+1}}.$$

Adaptation: atomic functional

Theorem [adaptation atomic functional]: Let $f \in \mathcal{W}_p$ be such that $\|f\|_\infty \leq L$ and suppose that the operator Λ is atomic for $m, s \geq 0$ and $p \geq \max(s+1, m+1, 2m-s)$, where $T_f(h) = \sum_{j=1}^{s} \int h^{(j)} d\mu_j$, μ_s with discrete component. Then the plug in estimator $\Lambda(\hat{f}_n)$ with $\hat{f}_n = \hat{f}_n^{\hat{j}_n}$ satisfies that

$$\mathbb{E}_{Q\mathbb{P}_f} |\Lambda(\hat{f}_n) - \Lambda(f)| \lesssim (n\alpha^2 \log^{-(1+2a)} n)^{-\frac{p-s}{2p+2}} \vee (n/\log n)^{-\frac{p-s}{2p+1}}.$$

Adaptation to smooth functional:

- The plug-in estimator $\Lambda(\hat{f}^{\hat{j}_n})$ doesn't work (too smooth)
- One can consider an rougher estimator $\hat{f}_n$ with threshold not depending on p.

Summary

- Privacy constrained inference is becoming increasingly popular, in particular differential privacy.
- Methods are typically case-by-case. New privacy constrained estimator requires new mechanism.
- We consider α-differential private plug-in estimators in semi-parametric problems.
- Can be used for a wide range, including smooth and atomic functionals.
- Derived matching minimax lower bounds.
- Adaptive inference for atomic functionals (smooth functionals need over-fitting).