Estimating functionals under local differential privacy

Angelika Rohde
(University of Freiburg)
with Cristina Butucea and Lukas Steinberger

Luminy
December 2020

Research supported by German Research Foundation (DFG): RO 3766/4-1

CLASSICAL VS. PRIVATE ESTIMATION

$(\mathcal{X}, \mathcal{F})$	the sample space
$\mathcal{P} = \mathcal{P}(\mathcal{X}, \mathcal{F})$	a model of probability measures P
$\theta : \mathcal{P} \to \mathbb{R}$	a functional of interest.

- **classical estimation problem:**
  - Given data $X_1, \ldots, X_n \overset{iid}{\sim} P \in \mathcal{P}$, estimate $\theta(P)$.

CLASSICAL VS. PRIVATE ESTIMATION

$(\mathcal{X}, \mathcal{F})$	the sample space
$\mathcal{P} = \mathcal{P}(\mathcal{X}, \mathcal{F})$	a model of probability measures P
$\theta : \mathcal{P} \to \mathbb{R}$	a functional of interest.

- **private estimation problem:**
  - Original data $X = (X_1, \ldots, X_n) \sim P^n$ are **not** observed.
  - Instead, a **privatized** sample Z on $(\mathcal{Z}, \mathcal{G})$ is given.
  - The **channel**

$$Q(A|x) = Pr(Z \in A | X = x), \quad A \in \mathcal{G}, x \in \mathcal{X}^n,$$

is restricted to be α**-locally differentially private**.

CLASSICAL VS. PRIVATE ESTIMATION

$(\mathcal{X}, \mathcal{F})$	sample space of the original data
$\mathcal{P} = \mathcal{P}(\mathcal{X}, \mathcal{F})$	model of probability measures P
$\theta : \mathcal{P} \to \mathbb{R}$	functional of interest
$(\mathcal{Z}, \mathcal{G})$	sample space of the privatized data
$Q : \mathcal{P}(\underbrace{X^n, \mathcal{F}^n}_{\text{source space}}) \to \mathcal{P}'(\underbrace{\mathcal{Z}, \mathcal{G}}_{\text{target space}})$	channel (Markov kernel)

- **The observed sample follows the distribution**

$$Z \sim QP^n,$$

where

$$QP^n(dz) := \int_{\mathcal{X}^n} Q(dz|x)\, P^n(dx).$$

α-Differential privacy

Definition 1 (Dwork et al., 2006)

Fix $\alpha \in (0, \infty)$. A channel $Q : \mathcal{P}(\mathcal{X}^n) \to \mathcal{P}'(\mathcal{Z})$ provides **α-differential privacy (α-DP)**, iff

$$Q(A|x) \;\leq\; e^{\alpha}\, Q(A|x'), \quad \forall A \in \mathcal{G}, \forall x, x' \in \mathcal{X}^n : d_0(x, x') = 1.$$

Here $d_0(x, x') := \#\{i : x_i \neq x'_i\}$ is the number of distinct components.

α-Differential privacy

$$\forall A, \forall x, x' : d_0(x, x') = 1 :$$
$$e^{-\alpha} \leq \frac{Q(A|x)}{Q(A|x')} \leq e^{\alpha}$$

- **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.
- The smaller $\alpha \in (0, \infty)$, the stronger is the privacy protection.

α-Differential privacy

$$\forall A, \forall x, x' : d_0(x, x') = 1 :$$
$$e^{-\alpha} \leq \frac{Q(A|x)}{Q(A|x')} \leq e^{\alpha}$$

- **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.
- The smaller $\alpha \in (0, \infty)$, the stronger is the privacy protection.

LOCAL DIFFERENTIAL PRIVACY

We say that an α-DP channel $Q : \mathcal{P}(\mathcal{X}^n) \to \mathcal{P}'(\mathcal{Z}^n)$ provides **local privacy**, if individual i can generate its privatized data Z_i on its 'local machine', without ever giving away its original data X_i.

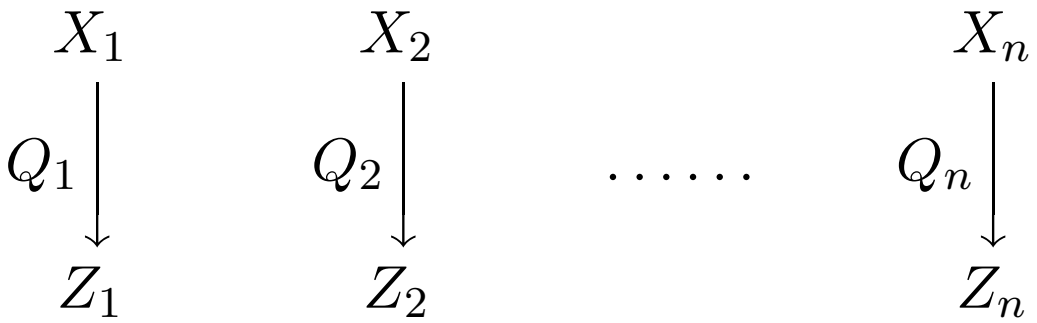
- No trusted third party needed

Local privacy - non-interactive case

Definition

We say that a channel $Q : \mathcal{P}(\mathcal{X}^n) \to \mathcal{P}'(\mathcal{Z}^n)$ is **non-interactive (NI)**, if there exist channels $Q_i : \mathcal{P}(\mathcal{X}) \to \mathcal{P}'(\mathcal{Z})$, such that

$$Q(dz|x) = \bigotimes_{i=1}^{n} Q_i(dz_i|x_i).$$

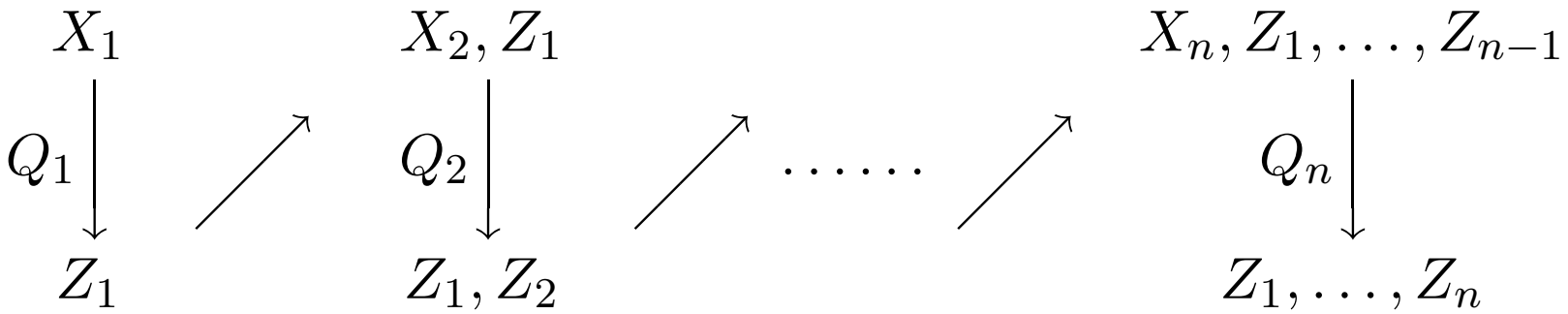


LOCAL PRIVACY - SEQUENTIALLY-INTERACTIVE CASE

Definition

We say that a channel $Q : \mathcal{P}(\mathcal{X}^n) \to \mathcal{P}'(\mathcal{Z}^n)$ is **sequentially-interactive (SI)**, if there exist channels $Q_i : \mathcal{P}(\mathcal{X} \times \mathcal{Z}^{i-1}) \to \mathcal{P}'(\mathcal{Z})$, $i = 1, \ldots, n$, such that

$$Q(dz|x) = Q_n(dz_n|x_n, z_1, \ldots, z_{n-1}) \cdots Q_2(dz_2|x_2, z_1) Q_1(dz_1|x_1).$$



LOCAL PRIVACY MECHANISMS

$$\mathcal{Q}_\alpha^{(NI)} := \bigcup_{(\mathcal{Z},\mathcal{G})} \{Q : Q \text{ is } \alpha\text{-non-interactive from } \mathcal{X}^n \text{ to } \mathcal{Z}^n\}$$

(α-non-interactive channels)

$$\mathcal{Q}_\alpha^{(SI)} := \bigcup_{(\mathcal{Z},\mathcal{G})} \{Q : Q \text{ is } \alpha\text{-sequentially interactive from } \mathcal{X}^n \text{ to } \mathcal{Z}^n\}$$

(α-sequentially-interactive channels)

Clearly, $\mathcal{Q}_\alpha^{(NI)} \subseteq \mathcal{Q}_\alpha^{(SI)}$.

α-PRIVATE MINIMAX RISKS

- For a class of
  probability measures $\mathcal{P}$ on $(\mathcal{X}, \mathcal{F})$
  and a parameter $\theta(P)$, $P \in \mathcal{P}$,
  we distinguish two α-private minimax risks:

$$\mathcal{M}_{n,\alpha}^{(NI)} = \inf_{Q \in \mathcal{Q}_\alpha^{(NI)}} \inf_{\hat{\theta}_n : \mathcal{Z}^n \to \mathbb{R}} \sup_{P \in \mathcal{P}} \mathbb{E}_{QP^n} \left[\ell\left(\hat{\theta}_n, \theta(P) \right) \right]$$

$$\mathcal{M}_{n,\alpha}^{(SI)} = \inf_{Q \in \mathcal{Q}_\alpha^{(SI)}} \inf_{\hat{\theta}_n : \mathcal{Z}^n \to \mathbb{R}} \sup_{P \in \mathcal{P}} \mathbb{E}_{QP^n} \left[\ell\left(\hat{\theta}_n, \theta(P) \right) \right].$$

- Note that the above infima include all possible measurable spaces $(\mathcal{Z}, \mathcal{G})$.

α-PRIVATE MINIMAX RISKS

- High-dimensional regression and mean estimation, density estimation in L_2-risk (Duchi et al., 2018):

$$\mathcal{M}_{n,\alpha}^{(NI)} \asymp \mathcal{M}_{n,\alpha}^{(SI)}$$

- Rohde and Steinberger (2020): Convex and dominated $\mathcal{P}$, linear and bounded $\theta : \mathcal{P} \to \mathbb{R}$:

$$\mathcal{M}_{n,\alpha}^{(NI)} \asymp \mathcal{M}_{n,\alpha}^{(SI)} \asymp \omega_{d_{\mathrm{TV}}}\left((n(e^{\alpha}-1)^2)^{-1/2}\right)$$

- (Adaptive) Density estimation (Butucea et al., 2019):

$$\mathcal{M}_{n,\alpha}^{(NI)} \asymp \mathcal{M}_{n,\alpha}^{(SI)}$$

- ...

From a statistical minimax point of view, do we actually have

$$\mathcal{M}_{\mathbf{n},\alpha}^{(\mathbf{NI})} \asymp \mathcal{M}_{\mathbf{n},\alpha}^{(\mathbf{SI})} \ ?$$

The quadratic functional

THE QUADRATIC FUNCTIONAL

- Bickel and Ritov (1988) were the first to discover the so-called **elbow phenomenon**:

While a $\sqrt{n}$-efficient estimator exists for Hölder smoothness to the exponent $s > 1/4$, the minimax rate of convergence over Hölder balls is

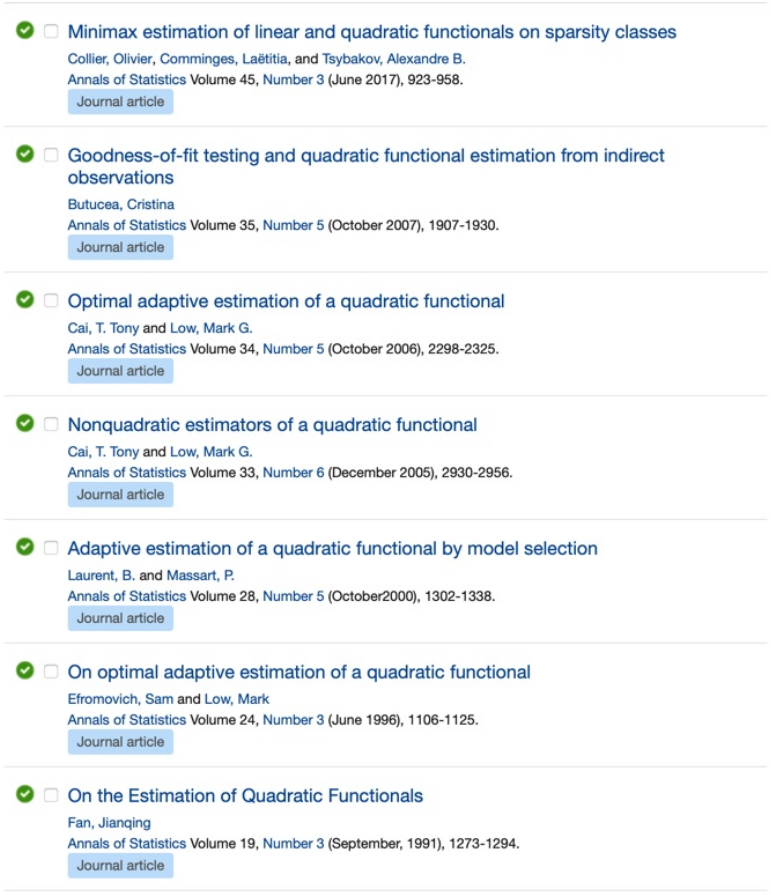
$$n^{-s/(s+1/4)}$$

whenever $s \leq 1/4$ although the standard information bound is strictly positive and finite, see Ritov and Bickel (1990).

Elbow at $s = 1/4$!

THE RUN ON QUADRATIC FUNCTIONAL

With their discovery, they triggered an avalanche of research activity, see



Minimax estimation of linear and quadratic functionals on sparsity classes
Collier, Olivier, Comminges, Laëtitia, and Tsybakov, Alexandre B.
Annals of Statistics Volume 45, Number 3 (June 2017), 923-958.
Journal article

Goodness-of-fit testing and quadratic functional estimation from indirect observations
Butucea, Cristina
Annals of Statistics Volume 35, Number 5 (October 2007), 1907-1930.
Journal article

Optimal adaptive estimation of a quadratic functional
Cai, T. Tony and Low, Mark G.
Annals of Statistics Volume 34, Number 5 (October 2006), 2298-2325.
Journal article

Nonquadratic estimators of a quadratic functional
Cai, T. Tony and Low, Mark G.
Annals of Statistics Volume 33, Number 6 (December 2005), 2930-2956.
Journal article

Adaptive estimation of a quadratic functional by model selection
Laurent, B. and Massart, P.
Annals of Statistics Volume 28, Number 5 (October2000), 1302-1338.
Journal article

On optimal adaptive estimation of a quadratic functional
Efromovich, Sam and Low, Mark
Annals of Statistics Volume 24, Number 3 (June 1996), 1106-1125.
Journal article

On the Estimation of Quadratic Functionals
Fan, Jianqing
Annals of Statistics Volume 19, Number 3 (September, 1991), 1273-1294.
Journal article

or Klemelä (2006), Giné and Nickl (2010), to mention just a few.

THE QUADRATIC FUNCTIONAL UNDER PRIVACY

Starting point:

- Where is the elbow?
- Do we see a difference between $\mathcal{M}_{n,\alpha}^{(NI)}$ and $\mathcal{M}_{n,\alpha}^{(SI)}$?

PRELIMINARIES

- $X_1, \ldots, X_n \overset{i.i.d.}{\sim} \mathcal{P}_f$ on $[0,1]$
- Lebesgue density $f : [0,1] \to \mathbb{R}_+$, $f \in L^2[0,1]$
- $D(f) = \int_0^1 f^2(x)dx$
- The conditional distribution of the privatized observations given the original sample

$$\mathcal{L}(Z_1, \ldots, Z_n \,|\, X_1, \ldots, X_n)$$

is described by the *channel distribution* Q

$$\uparrow$$

Markov kernel from $([0,1]^n, \mathcal{B}([0,1])^{\otimes n})$ to $(\mathcal{Z}^n, \mathcal{G}^{\otimes n})$

- Joint distribution of $Z_1, \ldots, Z_n$ on $\mathcal{Z}^n$ is given by

$$Q_f = Q\mathbb{P}_f^n$$

Besov spaces

- For any $h > 0$, and any real-valued function g on $[0, 1]$, Δ_h is defined by

$$\Delta_h g(t) = \begin{cases} g(t+h) - g(t) & \text{if } 0 \leq t \leq 1-h \\ 0 & \text{otherwise.} \end{cases}$$

- For any $2 \leq r \in \mathbb{N}$, $\Delta_h^r = \Delta_h \circ \Delta_h^{r-1}$ inductively defines its r-fold composition
- If $|g|^p$ is Lebesgue integrable, $p \geq 1$,

$$\omega_r(g, t, p) = \sup_{h \in (0,t]} \|\Delta_h^r g\|_{L_p}$$

denotes the rth *modulus of smoothness* in the Lebesgue space L_p.

Besov spaces

Besov norm:

$$\|f\|_{B_s^{pq}} = \begin{cases} \|f\|_{L_p} + \left(\sum_{j=0}^{\infty} \left[2^{js}\omega_r(f, 2^{-j}, p)\right]^q\right)^{1/q} & \text{if } 1 \le q < \infty \\ \|f\|_{L_p} + \sup_{j\ge 0} \left[2^{js}\omega_r(f, 2^{-j}, p)\right] & \text{if } q = \infty \end{cases}$$

Corresponding Besov spaces:

$$B_s^{pq} = \left\{ f \in L_p([0,1]) : \|f\|_{B_s^{pq}} < \infty \right\}, \quad \text{for } 1 \le p < \infty,$$
$$B_s^{\infty q} = \left\{ f \in \mathcal{C}([0,1]) : \|f\|_{B_s^{pq}} < \infty \right\}$$

- Note: The relation $s > (1/p - 1/2)_+$ reveals that $B_s^{pq} \subset L_2$

HAAR WAVELETS

- scaling function
$$\phi = \psi_{-10} = \mathbb{1}_{(0,1]}$$
- wavelet
$$\psi = \mathbb{1}_{(0,1/2]} - \mathbb{1}_{(1/2,1]}$$
- Shifted and rescaled version

$$\psi_{jk} = 2^{j/2}\psi(2^j \cdot -k) \text{ for } j \in \mathbb{N} \cup \{0\}, k \in \{0, 1, \ldots, 2^j - 1\}$$

Then

$$\left\{\psi_{-10}, \psi_{jk} : j \in \mathbb{N} \cup \{-1, 0\}, k \in \{0, 1, \ldots, 2^j - 1\}\right\}$$

defines the orthonormal *Haar wavelet basis* of L_2.

The quadratic functional

Non-Interactive (NI) privacy mechanisms

REWRITING THE QUADRATIC FUNCTIONAL

Parseval's identity:

$$\int_0^1 f(x)^2 dx = \sum_{j\geq -1} \sum_{k=0}^{(1\vee 2^j)-1} \langle f, \psi_{jk}\rangle_{L_2}^2$$

$$= \sum_{j\geq -1} \sum_{k=0}^{(1\vee 2^j)-1} \beta_{jk}^2$$

with wavelet coefficients $\beta_{jk} = \beta_{jk}(f) = \langle f, \psi_{jk}\rangle_{L_2}$

A SUITABLE NON-INTERACTIVE MECHANISM AND A CORRESPONDING ESTIMATOR

- **Privacy mechanism:** (Butucea et al., 2019)
  Given its original data X_i, individual i generates $(Z_{ijk})_{jk}$ with
  $$Z_{ijk} = \psi_{jk}(X_i) + \sigma_j(\alpha) \cdot W_{ijk},$$
  $k = 0, ..., \lceil 2^j - 1 \rceil$, $j = -1, ..., J-1$
  W_{ijk} are i.i.d. Laplace distributed with density
  $$f^W(x) = \frac{1}{2}\exp(-|x|), \quad x \in \mathbb{R}$$

Proposition

For a suitable choice of the constants $\sigma_j(\alpha)$, the privacy mechanism $Q^{(NI)}$ described above is α-non-interactive for any $J \in \mathbb{N}$.

A SUITABLE NON-INTERACTIVE MECHANISM AND A CORRESPONDING ESTIMATOR

- **A corresponding estimator:**

$$\hat{D}_n = \frac{1}{n(n-1)} \sum_{i \neq h}^{n} \sum_{j=-1}^{J-1} \sum_{k=0}^{(1 \vee 2^j)-1} Z_{ijk} \cdot Z_{hjk}.$$

(standard U-statistic of order 2)

- Note: The estimator is a bias-corrected version of

$$\sum_{j=-1}^{J-1} \sum_{k=0}^{(1 \vee 2^j)-1} \hat{\beta}_{jk}^2$$

with $\hat{\beta}_{jk} = \frac{1}{n} \sum_{i=1}^{n} Z_{ijk}$

NI-α-PRIVATIZED MINIMAX RATE

$$\mathcal{P}_s^{pq}(L) = \left\{ f : [0,1] \to \mathbb{R} : f \geq 0, \int_0^1 f(x)\,dx = 1, \|f\|_{B_s^{pq}} \leq L \right\}$$

Theorem (Butucea, R. and Steinberger, 2020)

Let $\alpha \in (0,1]$, $s > 0$, $p \geq 2$, $q \geq 1$. With

$$\bar{\mathcal{P}}_s^{pq}(L,M) = \mathcal{P}_s^{pq}(L) \cap L_3(M),$$

$$\inf_{Q \in \mathcal{Q}_\alpha^{(NI)}} \inf_{\hat{D}_n} \sup_{f \in \bar{\mathcal{P}}_s^{pq}(L,M)} \mathbb{E}_{Q\mathbb{P}_f^n}\left[\left|\hat{D}_n - D(f)\right|^2\right] \asymp \frac{1}{n\alpha^2} + \left(n\alpha^2\right)^{-\frac{2s}{s+3/4}}.$$

(up to logarithmic factors in the nonparametric part)

[The proof of the lower bound extends a beautiful approach of Lam-Weil et al. (2020)]

NI-α-PRIVATIZED MINIMAX RATE

Notice the elbow at

$$s = 3/4!$$

The quadratic functional

Sequentially Interactive (SI) privacy mechanisms

A SUITABLE SEQUENTIALLY INTERACTIVE MECHANISM AND A CORRESPONDING ESTIMATOR

We allow for sequential interaction between data owners

- **The privacy mechanism:**
  - Sample size $2n$
  - Split the data providing individuals into two groups: The first group holds data $X^{(1)} = (X_1^{(1)}, \ldots, X_n^{(1)})$ and the second group holds the data $X^{(2)} = (X_1^{(2)}, \ldots, X_n^{(2)})$

A SUITABLE SEQUENTIALLY INTERACTIVE MECHANISM AND A CORRESPONDING ESTIMATOR

- **First group and intermediate estimation step:**
  - They use the non-interactive privacy mechanism as just described to generate arrays $(Z_{ijk})_{jk}$, $i = 1, \ldots, n$.
  - These sanitized data are now used to estimate the unknown density (Butucea et al., 2019)

$$\hat{f}_J^{(1)}(x) := \sum_{j=-1}^{J-1} \sum_{k=0}^{(1 \vee 2^j)-1} \hat{\beta}_{jk} \psi_{jk}(x).$$

Note: The choice of J is crucial. Even for the same Besov class, its choice is different for the density estimator above and the U-statistics estimator for the quadratic functional in the non-interactive case.

A SUITABLE SEQUENTIALLY INTERACTIVE MECHANISM AND A CORRESPONDING ESTIMATOR

The second group now proceeds with estimating the (random) linear functional

$$f \mapsto \int_0^1 \hat{f}_J^{(1)}(x) f(x)\, dx$$

with the methodology of Rohde and Steinberger (2020).

A SUITABLE SEQUENTIALLY INTERACTIVE MECHANISM AND A CORRESPONDING ESTIMATOR

- **Second group – privacy mechanism**
  - Each individual i from the second group independently generates $Z_i^{(2)}$ by

$$Z_i^{(2)} = \begin{cases} \tau \frac{e^\alpha+1}{e^\alpha-1}, & \text{with probability } \frac{1}{2}\left(1 + \frac{\hat{f}_J^{(1)}(X_i^{(2)})}{\tau \frac{e^\alpha+1}{e^\alpha-1}}\right) \\ -\tau \frac{e^\alpha+1}{e^\alpha-1}, & \text{with probability } \frac{1}{2}\left(1 - \frac{\hat{f}_J^{(1)}(X_i^{(2)})}{\tau \frac{e^\alpha+1}{e^\alpha-1}}\right), \end{cases}$$

  with $\tau = \| \hat{f}_J^{(1)} \|_\infty$ (slightly simplified).

  Note:

$$\mathbb{E}\left(Z_i^{(2)} \,\middle|\, Z_1, \ldots, Z_n\right) = \int_0^1 \hat{f}_J^{(1)}(x) f(x)\, dx.$$

A SUITABLE SEQUENTIALLY INTERACTIVE MECHANISM AND A CORRESPONDING ESTIMATOR

- **The corresponding final estimator**

$$\tilde{D}_n = \frac{1}{n} \sum_{i=1}^{n} Z_i^{(2)}.$$

SI-α-PRIVATIZED MINIMAX RATE

Theorem (Butucea, R. and Steinberger, 2020)

Let $\alpha \in (0,1]$, $s > 0$, $p \geq 2$, $q \geq 1$. With

$$\bar{\mathcal{P}}_s^{pq}(L,M) = \mathcal{P}_s^{pq}(L) \cap L_\infty(M),$$

$$\inf_{Q \in \mathcal{Q}_\alpha^{(SI)}} \inf_{\hat{D}_n} \sup_{f \in \bar{\mathcal{P}}_s^{pq}(L,M)} \mathbb{E}_{Q\mathbb{P}_f^n}\left[\left|\hat{D}_n - D(f)\right|^2\right] \asymp \frac{1}{n\alpha^2} + \left(n\alpha^2\right)^{-\frac{2s}{s+1/2}}.$$

(up to logarithmic factors in the nonparametric part)

SI-α-PRIVATIZED MINIMAX RATE

Notice the elbow at

$$s = 1/2!$$

SUMMARY

- In the non-interactive case we construct an α-differentially private data release mechanism and estimator for $\int f^2 d\lambda$ based on U-statistics and sanitized empirical wavelet coefficients. For Besov classes B_s^{pq} with $p \geq 2$, $q \geq 1$, the optimal convergence rate in quadratic mean is given by

$$(n\alpha^2)^{-\frac{s}{s+3/4}} \vee (n\alpha^2)^{-1/2}, \quad \text{elbow at } s = 3/4.$$

- We improve the U-statistics approach by considering a two-step procedure that requires sequential interaction between data owners. The achieved minimax optimal rate is given by

$$(n\alpha^2)^{-\frac{s}{s+1/2}} \vee (n\alpha^2)^{-1/2}, \quad \text{elbow at } s = 1/2.$$

Thank you!

Bickel, P. J. and Ritov, Y. (1988). Estimating integrated squared density derivatives: sharp best order of convergence estimates. *Sankhya A*, 50(3):381–393.

Butucea, C., Dubois, A., Kroll, M., and Saumard, A. (2019). Local differential privacy: Elbow effect in optimal density estimation and adaptation over besov ellipsoids. *Bernoulli*, forthcoming.

Duchi, J. C., Jordan, M. I., and Wainwright, M. J. (2018). Minimax optimal procedures for locally private estimation. *J. Amer. Statist. Assoc.*, 113(521):182–201.

Dwork, C., McSherry, F., Nissim, K., and Smith, A. (2006). Calibrating noise to sensitivity in private data analysis. In Halevi, S. and Rabin, T., editors, *Theory of Cryptography*, Lecture Notes in Computer Science, pages 265–284. Springer.

Giné, E. and Nickl, R. (2010). Adaptive estimation of a distribution function and its density in sup-norm loss by wavelet and spline projections. *Bernoulli*, 16(4):1137–1163.

Klemelä, J. (2006). Sharp adaptive estimation of quadratic functionals. *Probab. Theory Relat. Fields*, 134:539–564.

Lam-Weil, J., Laurent, B., and Loubes, J.-M. (2020). Minimax optimal goodness-of-fit testing for densities and multinomials under a local differential privacy constraint.

Ritov, Y. and Bickel, P. (1990). Achieving information bounds in non and semiparametric models. *Ann. Statist.*, 18:925–938.

Rohde, A. and Steinberger, L. (2020). Geometrizing rates of convergence under local differential privacy constraints. *Ann. Statist.*, 48(6):2646–2670.