

Local differentiable privacy: optimal estimation and adaptation

Cristina Butucea¹

joint work with Amandine Dubois^{1,2}, Martin Kroll¹ and Adrien Saumard²

1. ENSAE, Institut Polytechnique de Paris; 2. ENSAI Rennes

Statistics Conference - in honor of Aad van der Vaart

Leiden - June 17-21, 2019

1 Introduction

- Local differentiable privacy
- Bibliography

2 Results

- Optimality
- Privacy mechanism
- Linear estimator
- Adaptive results

1. Introduction - Statistical model

Probability density model. Given $X_1, \dots, X_n$ i.i.d. random variables on $(\mathcal{X} = [0, 1], \mathcal{B})$, with common probability density function (pdf) $f : [0, 1] \rightarrow \mathbb{R}_+$, the model is the collection of these probability distributions :

$$\mathcal{P} = (P_f, f \in \mathcal{F}).$$

We want to estimate f .

1. Introduction - Statistical model

Probability density model. Given $X_1, \dots, X_n$ i.i.d. random variables on $(\mathcal{X} = [0, 1], \mathcal{B})$, with common probability density function (pdf) $f : [0, 1] \rightarrow \mathbb{R}_+$, the model is the collection of these probability distributions :

$$\mathcal{P} = (P_f, f \in \mathcal{F}).$$

We want to estimate f .

Confidentiality prevents us from publishing/using r.v. $X_1, \dots, X_n$. Instead, we have to produce other random variables $Z_1, \dots, Z_n$ on $(\mathcal{Z}, \mathcal{A})$, such that:

1. the information in the data is sufficiently anonymized
2. we can still recover some information about f out of Z 's.

The initial sample $X = (X_1, \dots, X_n)$ with law P^n is not observed. Instead, we produce $Z = (Z_1, \dots, Z_n)$ with law

$$QP^n = \int Q(dz|X = x)dP^n(x), \quad x \in [0, 1]^n.$$

The initial sample $X = (X_1, \dots, X_n)$ with law P^n is not observed. Instead, we produce $Z = (Z_1, \dots, Z_n)$ with law

$$QP^n = \int Q(dz|X = x)dP^n(x), \quad x \in [0, 1]^n.$$

The Markov kernel defining the conditional distribution of $Z/X = x$ via

$$Q(A|x) = P(Z \in A|X = x), \quad \text{for all } A \in \mathcal{A}, x \in [0, 1]^n$$

is also known as **channel** or **privacy mechanism (PM)**.

The initial sample $X = (X_1, \dots, X_n)$ with law P^n is not observed. Instead, we produce $Z = (Z_1, \dots, Z_n)$ with law

$$QP^n = \int Q(dz|X = x)dP^n(x), \quad x \in [0, 1]^n.$$

The Markov kernel defining the conditional distribution of $Z/X = x$ via

$$Q(A|x) = P(Z \in A|X = x), \quad \text{for all } A \in \mathcal{A}, x \in [0, 1]^n$$

is also known as **channel** or **privacy mechanism (PM)**.

Dwork et al., 2006 For α in $(0, \infty)$ fixed, a PM Q provides α -differential privacy if, for all $A \in \mathcal{A}$ and x, x' that only differ at one coordinate:

$$Q(A|x) \leq e^\alpha Q(A|x').$$

Note that: the smallest α , the strongest the privacy.

We distinguish global vs. local privacy.

We distinguish global vs. local privacy.

In **global privacy**, all data are trusted to someone, who produces $Z_1, \dots, Z_n$ and publishes these r.v. - not realistic to trust one person/machine with all data.

We distinguish global vs. local privacy.

In **global privacy**, all data are trusted to someone, who produces $Z_1, \dots, Z_n$ and publishes these r.v. - not realistic to trust one person/machine with all data.

In **local privacy**, new r.v. Z_i are produced using only one X_i at a time. However, interactions with other randomized Z_k , $k \neq i$ are possible, and improving on the rates in some cases.

Examples of local PM:

- a) non interactive mechanism: each Z_i is produced given only X_i ;
- b) sequentially interactive privacy mechanisms: Z_i is produced given $Z_1, \dots, Z_{i-1}$ and X_i .

Local differentiable privacy

We write, e.g. for the sequentially interactive privacy mechanism that

$$Z_i \sim Q_i(\cdot | X_i, Z_1, \dots, Z_{i-1}), \quad Q_i \text{ is a Markov kernel.}$$

A PM $\{Q_i, i = 1, \dots, n\}$ provides α -local differential privacy (α -LDP) if

$$\sup_{A \in \mathcal{A}} \frac{Q_i(A \mid X_i = x, Z_1 = z_1, \dots, Z_{i-1} = z_{i-1})}{Q_i(A \mid X_i = x', Z_1 = z_1, \dots, Z_{i-1} = z_{i-1})} \leq \exp(\alpha), \quad \text{for all } x \neq x'$$

Local differentiable privacy

We write, e.g. for the sequentially interactive privacy mechanism that

$$Z_i \sim Q_i(\cdot | X_i, Z_1, \dots, Z_{i-1}), \quad Q_i \text{ is a Markov kernel.}$$

A PM $\{Q_i, i = 1, \dots, n\}$ provides α -local differential privacy (α -LDP) if

$$\sup_{A \in \mathcal{A}} \frac{Q_i(A | X_i = x, Z_1 = z_1, \dots, Z_{i-1} = z_{i-1})}{Q_i(A | X_i = x', Z_1 = z_1, \dots, Z_{i-1} = z_{i-1})} \leq \exp(\alpha), \quad \text{for all } x \neq x'$$

In the non-interactive case, the condition becomes

$$\sup_{A \in \mathcal{A}} \frac{Q(A | X_i = x)}{Q(A | X_i = x')} \leq \exp(\alpha), \quad \text{for all } x \neq x'.$$

Denote $\mathcal{Q}_\alpha$ the set of Markov kernels providing α -differential privacy

Goals

We are interested in:

1. providing mechanism Q or $\{Q_i\}_i$ in $\mathcal{Q}_\alpha$ (sufficient confidentiality if α is small),
2. estimate f in an optimal way.

We are interested in:

1. providing mechanism Q or $\{Q_i\}_i$ in $\mathcal{Q}_\alpha$ (sufficient confidentiality if α is small),
2. estimate f in an optimal way.

Questions:

- is the estimation rate deteriorated?

We are interested in:

1. providing mechanism Q or $\{Q_i\}_i$ in $\mathcal{Q}_\alpha$ (sufficient confidentiality if α is small),
2. estimate f in an optimal way.

Questions:

- is the estimation rate deteriorated?
- if so, quantify the loss in the rate?

We are interested in:

1. providing mechanism Q or $\{Q_i\}_i$ in $\mathcal{Q}_\alpha$ (sufficient confidentiality if α is small),
2. estimate f in an optimal way.

Questions:

- is the estimation rate deteriorated?
- if so, quantify the loss in the rate?
- check optimality of the rates over all possible estimators
and
over all possible mechanisms in $\mathcal{Q}_\alpha$?

When X_i 's are observed, we denote by

$$R_n^*(\mathcal{F}_s) := \inf_{\hat{f}_n} \sup_{f \in \mathcal{F}_s} E_f[\|\hat{f}_n - f\|_r^r], \quad r \geq 1.$$

When X_i 's are observed, we denote by

$$R_n^*(\mathcal{F}_s) := \inf_{\hat{f}_n} \sup_{f \in \mathcal{F}_s} E_f[\|\hat{f}_n - f\|_r^r], \quad r \geq 1.$$

In the private setup, we want to study the asymptotic behaviour of the risk:

$$R_n(\mathcal{F}) := \inf_{\hat{f}_n, Q \in \mathcal{Q}_\alpha} \sup_{f \in \mathcal{F}_s} E_{f,Q}[\|\hat{f}_n - f\|_r^r], \quad r \geq 1.$$

Classical vs. private minimax risk

When X_i 's are observed, we denote by

$$R_n^*(\mathcal{F}_s) := \inf_{\hat{f}_n} \sup_{f \in \mathcal{F}_s} E_f[\|\hat{f}_n - f\|_r^r], \quad r \geq 1.$$

In the private setup, we want to study the asymptotic behaviour of the risk:

$$R_n(\mathcal{F}) := \inf_{\hat{f}_n, Q \in \mathcal{Q}_\alpha} \sup_{f \in \mathcal{F}_s} E_{f,Q}[\|\hat{f}_n - f\|_r^r], \quad r \geq 1.$$

Here, we consider classes of **Besov smooth functions**, $\mathcal{F}_s = B_{pq}^s$, $s > \frac{1}{p}$, $1 \leq p, q \leq \infty$.

These classes can also be sparse for some values of s, p, q .

Short bibliography

- Wasserman and Zhou (2010) global privacy can be achieved with similar rates in classical and private setups;
- Duchi, Jordan, Wainwright (2018): $r = 2$ (MISE); Sobolev smooth functions:

$$R_n(W_s(L)) \asymp n^{-\frac{s}{s+1}}, \quad s > 0,$$

instead of $n^{-\frac{2s}{2s+1}}$ when $X_1, \dots, X_n$ are available.

- Rohde, Steinberger (2018) - linear functionals, include $f(x_0)$ for some fixed point x_0 , i.e. the pointwise risk (MSE).
- B., Rohde, Steinberger (soon) - quadratic functionals.

- Wasserman and Zhou (2010) global privacy can be achieved with similar rates in classical and private setups;
- Duchi, Jordan, Wainwright (2018): $r = 2$ (MISE); Sobolev smooth functions:

$$R_n(W_s(L)) \asymp n^{-\frac{s}{s+1}}, \quad s > 0,$$

instead of $n^{-\frac{2s}{2s+1}}$ when $X_1, \dots, X_n$ are available.

- Rohde, Steinberger (2018) - linear functionals, include $f(x_0)$ for some fixed point x_0 , i.e. the pointwise risk (MSE).
- B., Rohde, Steinberger (soon) - quadratic functionals.

Here, we provide first **adaptive results**: both the privacy mechanism and the estimator are free of all the parameters of the class.

We show an elbow effect too, and differences to the classical setup.

2. Results - Optimality

Theorem

Let $\alpha \in (0, \infty)$. We have

$$R_n(B_{p,q}^s) \geq (n(e^\alpha - 1)^2)^{-r \frac{s}{2s+2}}$$

and for $s \geq 1/p$

$$R_n(B_{p,q}^s) \geq \left(\frac{\log(n(e^\alpha - 1)^2)}{n(e^\alpha - 1)^2} \right)^{r \frac{s-1/p+1/r}{2(s-1/p)+2}}.$$

Also, obviously,

$$R_n(B_{p,q}^s) \geq R_n^*(B_{p,q}^s).$$

Theorem (Donoho, Johnstone, Kerkycharian, Picard 1996, Härdle, Kerkycharian, Picard, Tsybakov 2000)

Wavelet estimators with sufficiently regular wavelets attain the minimax rates (log – factors):

$$R_n^*(B_{pq}^s) \asymp \begin{cases} n^{-\frac{rs}{2s+1}}, & \text{if } p > \frac{r}{2s+1} \text{ dense case,} \\ \left(\frac{n}{\log n}\right)^{-r \cdot \frac{(s-1/p+1/r)}{2(s-1/p)+1}}, & \text{if } p \leq \frac{r}{2s+1} \text{ sparse case.} \end{cases}$$

Theorem (Donoho, Johnstone, Kerkycharian, Picard 1996, Härdle, Kerkycharian, Picard, Tsybakov 2000)

Wavelet estimators with sufficiently regular wavelets attain the minimax rates (log – factors):

$$R_n^*(B_{pq}^s) \asymp \begin{cases} n^{-\frac{rs}{2s+1}}, & \text{if } p > \frac{r}{2s+1} \text{ dense case,} \\ \left(\frac{n}{\log n}\right)^{-r \cdot \frac{(s-1/p+1/r)}{2(s-1/p)+1}}, & \text{if } p \leq \frac{r}{2s+1} \text{ sparse case.} \end{cases}$$

Note an "elbow effect" at $p = \frac{r}{2s+1}$.

Moreover, the dense case can be separated in

-the homogeneous case $p \geq r$, where the **linear estimator**

$$\hat{f}_n = \sum_k \hat{\alpha}_{j_0 k} \phi_{j_0 k} + \sum_{j=j_0}^{j_1} \sum_k \hat{\beta}_{jk} \psi_{jk}$$

attains the optimal rate, and

-the non-homogeneous case $\frac{r}{2s+1} < p < r$, where any linear estimator is suboptimal.

Instead, whenever $p < r$ non-linear (thresholded) estimators with $\hat{\beta}_{jk}$ replaced by

$$\tilde{\beta}_{jk} = \hat{\beta}_{jk} \cdot I(|\hat{\beta}_{jk}| > t_j).$$

with the right choice of j_1 and t_j are optimal.

Instead, whenever $p < r$ non-linear (thresholded) estimators with $\hat{\beta}_{jk}$ replaced by

$$\tilde{\beta}_{jk} = \hat{\beta}_{jk} \cdot I(|\hat{\beta}_{jk}| > t_j).$$

with the right choice of j_1 and t_j are optimal.

Adaptive results, i.e. estimators free of $s, \dots$, are provided by nonlinear estimators as well.

Privacy mechanism

We assume that

ϕ and ψ are compactly supported on an interval $[-A, A]$.

Let us construct the first **noninteractive mechanism** which is α -LDP:
for $i = 1, \dots, n$, $j \in \{-1, \dots, j_1\}$, define

$$Z_{ijk} = \begin{cases} \phi_{j_0 k}(X_i) + c_\phi \frac{2^{j_0/2}}{\alpha} \cdot W_{i, j_0-1, k}, & \text{if } j = j_0 - 1, k \in \mathcal{N}_{j_0-1}, \\ \psi_{jk}(X_i) + c_\psi \frac{2^{j_1/2}}{\alpha} \cdot W_{ijk}, & \text{if } j = j_0, \dots, j_1, k \in \mathcal{N}_j, \end{cases}$$

where W_{ijk} are independent Laplace distributed random variables.

Note that $|\mathcal{N}_j| \sim 2^j$.

Privacy mechanism

We assume that

ϕ and ψ are compactly supported on an interval $[-A, A]$.

Let us construct the first **noninteractive mechanism** which is α -LDP:
for $i = 1, \dots, n$, $j \in \{-1, \dots, j_1\}$, define

$$Z_{ijk} = \begin{cases} \phi_{j_0 k}(X_i) + c_\phi \frac{2^{j_0/2}}{\alpha} \cdot W_{i, j_0-1, k}, & \text{if } j = j_0 - 1, k \in \mathcal{N}_{j_0-1}, \\ \psi_{jk}(X_i) + c_\psi \frac{2^{j_1/2}}{\alpha} \cdot W_{ijk}, & \text{if } j = j_0, \dots, j_1, k \in \mathcal{N}_j, \end{cases}$$

where W_{ijk} are independent Laplace distributed random variables.
Note that $|\mathcal{N}_j| \sim 2^j$.

Indeed,

$$\prod_{j=j_0-1}^{j_1} \prod_{k \in \mathcal{N}_j} \frac{q^{Z/X=x}(z_{jk})}{q^{Z/X=x'}(z_{jk})} \leq e^\alpha,$$

for all $x \neq x'$ and z_{jk} .

Dense homogeneous case: $p \geq r$

Thus,

$$\bar{Z}_{\cdot,jk} = \begin{cases} \hat{\alpha}_{j_0k} + c_\phi \frac{2^{j_0/2}}{\alpha} \cdot \bar{W}_{\cdot,j_0-1,k}, & j = j_0 - 1, k \in \mathcal{N}_{j_0-1} \\ \hat{\beta}_{jk} + c_\psi \frac{2^{j_1/2}}{\alpha} \cdot \bar{W}_{\cdot,j,k}, & j = j_0, \dots, j_1, k \in \mathcal{N}_j \end{cases}$$

are unbiased estimates of the true coefficients.

Theorem

The linear estimator

$$\tilde{f}_n = \sum_k \bar{Z}_{\cdot,j_0-1,k} \phi_{j_0k} + \sum_{j=j_0}^{j_1} \sum_k \bar{Z}_{\cdot,j,k} \psi_{jk}$$

with j_0 fixed and j_1 such that

$$2^{j_1} \asymp (n\alpha^2)^{\frac{1}{2s+2}},$$

yields

$$R_n(B_{pq}^s) \leq (n\alpha^2)^{-r \frac{s}{2s+2}} \vee n^{-r \frac{s}{2s+1}}.$$

Note that for $\alpha = \alpha_n \geq n^{1/(2s+1)}$ the privacy is very low and the classical rates can be attained.

Note that for $\alpha = \alpha_n \geq n^{1/(2s+1)}$ the privacy is very low and the classical rates can be attained.

As in the classical setup, the linear estimators cannot be optimal for $p < r$.

Adaptation

Let us construct the second **noninteractive mechanism** which is α -LDP:
for $i = 1, \dots, n$, $j \in \{-1, \dots, j_1\}$, define

$$Z_{ijk} = \begin{cases} \phi_{j_0 k}(X_i) + c_\phi \frac{2^{j_0/2}}{\alpha} \cdot W_{i, j_0-1, k}, & \text{if } j = j_0 - 1, k \in \mathcal{N}_{j_0-1}, \\ \psi_{jk}(X_i) + C_\psi j^\nu \frac{2^{j/2}}{\alpha} \cdot W_{ijk}, & \text{if } j = j_0, \dots, j_1, k \in \mathcal{N}_j, \end{cases}$$

where $\nu > 1$ is fixed and W_{ijk} are independent Laplace distributed random variables.

Let us construct the second **noninteractive mechanism** which is α -LDP: for $i = 1, \dots, n$, $j \in \{-1, \dots, j_1\}$, define

$$Z_{ijk} = \begin{cases} \phi_{j_0 k}(X_i) + c_\phi \frac{2^{j_0/2}}{\alpha} \cdot W_{i, j_0-1, k}, & \text{if } j = j_0 - 1, k \in \mathcal{N}_{j_0-1}, \\ \psi_{jk}(X_i) + C_\psi j^\nu \frac{2^{j/2}}{\alpha} \cdot W_{ijk}, & \text{if } j = j_0, \dots, j_1, k \in \mathcal{N}_j, \end{cases}$$

where $\nu > 1$ is fixed and W_{ijk} are independent Laplace distributed random variables.

The nonlinear estimator $\tilde{f}_n$ is defined for j_0, j_1, t_j :

$$\tilde{f}_n = \sum_k \bar{Z}_{\cdot, j_0-1, k} \phi_{j_0 k} + \sum_{j \geq j_0} \sum_k \bar{Z}_{\cdot, jk} \cdot I(|\bar{Z}_{\cdot, jk}| \geq t_j) \psi_{jk},$$

automatically adapts to the parameters of the class within some intervals.

For some integer $N \geq 0$, let $s \in (1/p, N)$, $1 \leq p, q \leq \infty$, $L \in [\underline{L}, \overline{L}]$.

Theorem

The nonlinear estimator $\tilde{f}_n$ with j_0 such that

$$2^{j_0} \asymp (n\alpha^2)^{\frac{1}{2(N+1)+2}} \wedge n^{\frac{1}{2(N+1)+1}}$$

and $j_1 = j_1' \wedge j_1''$

$$2^{j_1'} \asymp \frac{n}{\log n}, \quad 2^{2j_1''} \asymp \frac{n\alpha^2}{\log(n\alpha^2)},$$

and the threshold

$$t_{j,n} = \gamma \cdot \frac{j^{\nu+\frac{1}{2}}}{\sqrt{n}} \left(1 \vee \frac{2^{j/2}}{\alpha} \right), \quad \gamma > r(N+1),$$

attains the following rates for $s \geq \frac{1}{p}$ up to $\log$ -factors,

$$\left\{ \begin{array}{lll} n^{-r \frac{s}{2s+1}} & \vee & (n\alpha^2)^{-r \frac{s}{2s+2}}, & \text{if } p > \frac{r}{s+1}, \\ n^{-r \frac{s}{2s+1}} & \vee & \left(\frac{n\alpha^2}{\log(n\alpha^2)} \right)^{-\frac{r(s-1/p+1/r)}{2(s-1/p)+2}}, & \text{if } \frac{r}{2s+1} < p \leq \frac{r}{s+1}, \\ \left(\frac{n}{\log n} \right)^{-\frac{r(s-1/p+1/r)}{2(s-1/p)+1}} & \vee & \left(\frac{n\alpha^2}{\log(n\alpha^2)} \right)^{-\frac{r(s-1/p+1/r)}{2(s-1/p)+2}}, & \text{if } p \leq \frac{r}{2s+1}. \end{array} \right.$$

Connection to inverse problems

In the case of non interactive PM, each observation Z_i is obtained from X_i . Assume the PM has pdf $q^{Z|X=x}$, then the pdf of Z_i 's is

$$f^Z(z) = \int q^{Z|X=x}(z)f(x)dx =: Kf(z),$$

for some linear operator K .

In the particular case where $q^{Z|X=x}(z) = q(z - x)$ we have a convolution operator.

Connection to inverse problems

In the case of non interactive PM, each observation Z_i is obtained from X_i . Assume the PM has pdf $q^{Z|X=x}$, then the pdf of Z_i 's is

$$f^Z(z) = \int q^{Z|X=x}(z) f(x) dx =: K f(z),$$

for some linear operator K .

In the particular case where $q^{Z|X=x}(z) = q(z - x)$ we have a convolution operator.

If K^*K admits the eigenvalues $\lambda_k^2 \sim k^{-2\rho}$ as $k \rightarrow \infty$, we call ρ the 'difficulty' of the inverse problem.

Note that the convolution with the Laplace distribution has 'difficulty' $\rho = 2$.

Kerkyacharian, Petrushev, Picard, Willer, 2007

considered a Gaussian white noise framework

$$dy(t) = Kf(t) + \varepsilon dW_t, \quad t \in [0, 1],$$

with the linear operator K having 'difficulty' ρ .

Theorem

The minimax rates are (up to log – factors)

$$R_n^*(B_{pq}^s) \preceq \begin{cases} n^{-\frac{rs}{2(s+\rho)+1}}, & \text{if } s > (\rho + \frac{1}{2})(\frac{r}{p} - 1) \text{ dense case,} \\ n^{-r \cdot \frac{(s-1/p+1/r)}{2(s-1/p+\rho)+1}}, & \text{if } \frac{1}{p} \leq s \leq (\rho + \frac{1}{2})(\frac{r}{p} - 1) \text{ sparse case.} \end{cases}$$

Local differentiable privacy acts as an inverse problems of 'difficulty'
 $\rho = \frac{1}{2}$.

A globally and locally privatized

Happy Birthday Aad !