On robustness and local differential privacy

Tom Berrett (joint work with Mengchu Li and Yi Yu)
University of Warwick

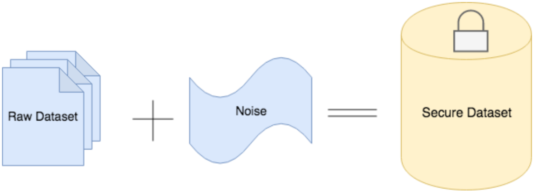
Meeting in Mathematical Statistics, CIRM

21st December 2023

Privacy mechanisms

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

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



Source: `medium.com`

How much noise should we add? What type of noise?

Differential privacy

Privacy mechanism Q is called *α-differentially private* (Dwork et al., 2006) if

$$\sup_A \frac{Q(A|\mathrm{x})}{Q(A|\mathrm{x}')} \leq e^{\alpha}$$

for all x, x' such that $d(\mathrm{x}, \mathrm{x}') := \sum_{i=1}^{n} \mathbb{1}_{x_i \neq x_i'} \leq 1$.

Differential privacy provides a rigorous framework to control the amount of personal information in published data. Large scale applications include

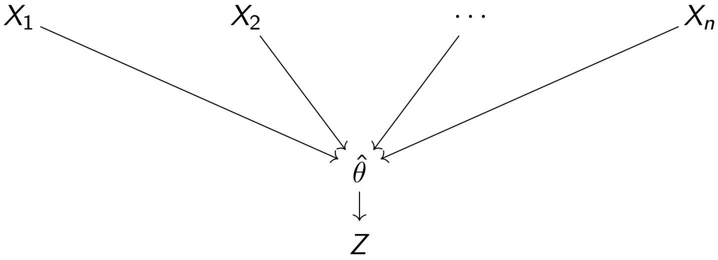
- Google Chrome (Erlingsson, Pihur and Korolova, 2014);
- Apple in iOS and macOS (Tang et al., 2017);
- Microsoft (Ding, Kulkarni and Yekhanin, 2017);
- Uber (Near, 2018);
- US Census (Dwork, 2019).

Can also be used to demonstrate GDPR compliance (Cohen and Nissim, 2020).

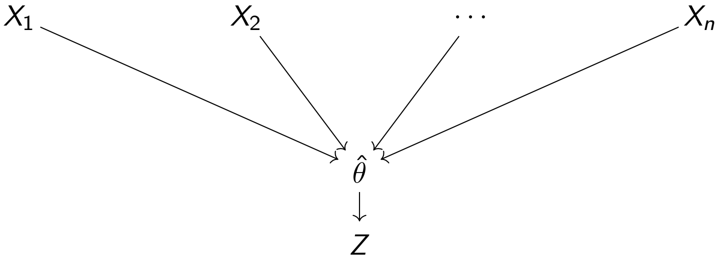
Differential privacy and robust statistics

Limiting the influence of any single input reminds us of robust statistics.

There has been interesting work on the link between these two areas (Dwork and Lei, 2009; Avella-Medina, 2021; Hopkins et al., 2023; Asi et al., 2023), focussed mainly on the *central model* of differential privacy.



Differential privacy and robust statistics



If $\hat{\theta} = f(X_1, \ldots, X_n)$ and

$$\Delta f := \sup_{x, x' : \sum_{i=1}^{n} \mathbb{1}_{\{x_i \neq x_i'\}} \leq 1} |f(x) - f(x')|$$

is the global sensitivity of f (Dwork and Lei, 2009), then we can take

$$Z = \hat{\theta} + \frac{\Delta f}{\alpha} W,$$

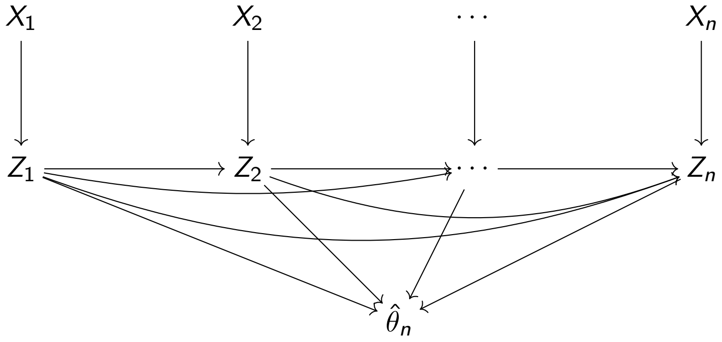
where $W \sim \mathrm{Laplace}(1)$.

Local differential privacy

We consider the local model of differential privacy (e.g. Duchi et al., 2013), where data are randomised one-by-one.

$$\sup_{A} \sup_{x_i, x_i', z_1, \ldots, z_{i-1}} \frac{Q_i(A|x_i, z_1, \ldots, z_{i-1})}{Q_i(A|x_i', z_1, \ldots, z_{i-1})} \leq e^{\alpha}, \text{ for all } i = 1, ..., n.$$

No trusted third party: analyse $Z = (Z_1, \ldots, Z_n)$ with



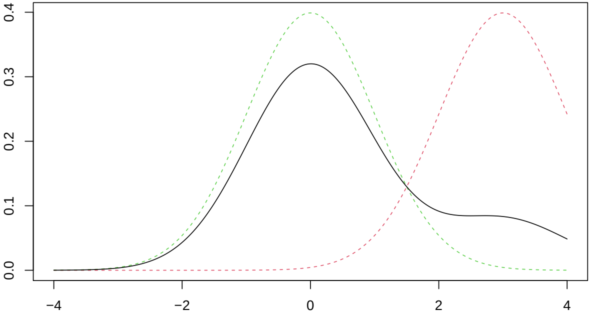
Local differential privacy and robustness

We study the relationship between privacy and robustness in this model.

For robustness we work with Huber contamination: instead of i.i.d. data from distribution of interest P, the raw data is i.i.d. from

$$(1-\varepsilon)P + \varepsilon G$$

for some $\varepsilon \in (0,1)$ and arbitrary distribution G.



Combination of contamination and privacy

We suppose that the raw data is contaminated, *before* being privatised:

$$X_1, \dots, X_n \sim (1-\epsilon)P + \epsilon G \quad \text{then} \quad Z_1, \dots, Z_n \sim Q(\cdot | X_1, \dots, X_n).$$

It is also possible to consider contamination after privatisation. The results can be very different (Cheu et al., 2021; Acharya et al., 2021; Chhor and Sentenac, 2023).

Minimax framework

Our object of interest is the minimax risk in this model:

$$\mathcal{R}_{n,\alpha}(\theta(\mathcal{P}), \Phi \circ \rho, \varepsilon) = \inf_{Q \in \mathcal{Q}_\alpha} \inf_{\hat{\theta}} \sup_{P_\varepsilon \in \mathcal{P}_\varepsilon(\mathcal{P})} \mathbb{E}_{P_\varepsilon, Q} \left[\Phi \circ \rho \left(\hat{\theta}, \theta(P) \right) \right],$$

where

- $\theta(P) \in \Theta$ is the quantity to be estimated;
- ρ is a semi-metric on Θ and Φ is non-decreasing with $\Phi(0) = 0$;
- $\mathcal{P}_\varepsilon = \{(1-\varepsilon)P + \varepsilon G : P \in \mathcal{P}, G \in \mathcal{G}\}$ with $\mathcal{P}$ a class of distributions of interest and $\mathcal{G}$ the class of all distributions on $\mathcal{X}$;
- the inner infimum is taken over all measurable functions $\hat{\theta}$ of the privatised data;
- $\mathcal{Q}_\alpha$ is the set of all α-LDP mechanisms.

TV modulus of continuity

In the classical i.i.d. model Donoho and Liu (1991) show that in a broad class of estimation problems the minimax risk is controlled by

$$\omega_{\mathrm{H}}(\varepsilon) := \sup\{\rho(\theta(R_0), \theta(R_1)) : \mathrm{H}(R_0, R_1) \leq \varepsilon/(1-\varepsilon), R_1, R_2 \in \mathcal{P}\}.$$

Chen et al. (2016) (cf. Devroye and Lugosi (2001)) develop general theory in Huber's model, and show that

$$\omega_{\mathrm{TV}}(\varepsilon) := \sup\{\rho(\theta(R_0), \theta(R_1)) : \mathrm{TV}(R_0, R_1) \leq \varepsilon/(1-\varepsilon), R_1, R_2 \in \mathcal{P}\}$$

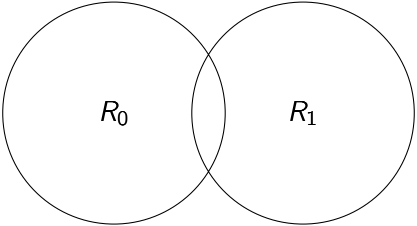
controls the statistical difficulty due to contamination in many problems.

Rohde and Steinberger (2020) show that ω_{TV} controls the statistical difficulty of a class of estimation problems under α-LDP.

General lower bound

Following Chen et al. (2016), if $\mathrm{TV}(R_0, R_1) \leq \varepsilon/(1-\varepsilon)$ then there exist G_0, G_1 with

$$(1-\varepsilon)R_0 + \varepsilon G_0 = (1-\varepsilon)R_1 + \varepsilon G_1.$$



Choose R_0 and R_1 to attain the supremum in

$$\omega_{\mathrm{TV}}(\varepsilon) = \sup\{\rho(\theta(R_0), \theta(R_1)) : \mathrm{TV}(R_0, R_1) \leq \varepsilon/(1-\varepsilon)\}.$$

By Le Cam's two point method we have $\mathcal{R}_{n,\alpha}(\theta(\mathcal{P}), \Phi \circ \rho, \varepsilon) \geq \frac{1}{2}\Phi(\frac{\omega(\varepsilon)}{2})$.

General lower bound

Combining with a trivial lower bound, we have the general simple result

$$\mathcal{R}_{n,\alpha}(\theta(\mathcal{P}), \Phi \circ \rho, \varepsilon) \geq \mathcal{R}_{n,\alpha}(\theta(\mathcal{P}), \Phi \circ \rho, 0) \vee \frac{1}{2}\Phi\Big(\frac{\omega(\varepsilon)}{2}\Big),$$

where the difficulty due to privacy and contamination separate.

We study a range of problems, showing that we can attain this lower bound in each case.

- (Simple hypothesis testing)
- Mean estimation
- Density estimation
- Median estimation

Simple hypothesis testing

Simple hypothesis testing

We start by considering the robust simple hypothesis testing problem

$$\begin{aligned} &\mathrm{H}_0 : P \in \mathcal{P}_\varepsilon(P_0) = \{P_\varepsilon : (1-\varepsilon)P_0 + \varepsilon G,\ G \in \mathcal{G}\} \\ \text{vs.}\quad &\mathrm{H}_1 : P \in \mathcal{P}_\varepsilon(P_1) = \{P_\varepsilon : (1-\varepsilon)P_1 + \varepsilon G,\ G \in \mathcal{G}\} \end{aligned}$$

for fixed P_0, P_1.

In the non-private setting, we can use the *Scheffé test* (Devroye and Lugosi, 2001; Chen et al., 2016): Reject H_0 if and only if

$$\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{X_i \in A^c\}} > \frac{1}{2}\{P_0(A) + P_1(A)\},$$

where A satisfies $P_0(A) - P_1(A) = \sup_S\{P_0(S) - P_1(S)\} = \mathrm{TV}(P_0, P_1)$.

Simple hypothesis testing

We apply this method to the output of the randomised response mechanism (Warner, 1965; Gopi et al., 2020)

$$Z_i = \begin{cases} \mathbb{1}_{\{X_i \in A^c\}}, & \text{w.pr. } e^\alpha/(1+e^\alpha), \\ 1 - \mathbb{1}_{\{X_i \in A^c\}}, & \text{otherwise.} \end{cases}$$

Reject if and only if[1] $\frac{1}{n}\sum_{i=1}^n Z_i > \frac{1}{2}\{P_0(A) + P_1(A)\}$.

Analysing the risk of this test shows that

$$\begin{aligned} \mathcal{R}_{n,\alpha}(\varepsilon) &= \inf_{Q \in \mathcal{Q}_\alpha} \inf_{\phi \in \Phi_Q} \left\{ \sup_{P \in \mathcal{P}_\varepsilon(P_0)} \mathbb{E}_{P,Q}(\phi) + \sup_{P' \in \mathcal{P}_\varepsilon(P_1)} \mathbb{E}_{P',Q}(1-\phi) \right\} \\ &\leq 2\exp[-Cn\alpha^2\{\mathrm{TV}(P_0, P_1) - 2\varepsilon\}_+^2]. \end{aligned}$$

[1] Using $\frac{1-\varepsilon}{2}\{P_0(A) + P_1(A)\} + \frac{\varepsilon}{2}$ gives $2\exp[-Cn\alpha^2\{\mathrm{TV}(P_0, P_1) - \varepsilon/(1-\varepsilon)\}_+^2]$

Simple hypothesis testing

We have a lower bound to match the previous upper bound. For M_0, M_1 corrupted versions of P_0, P_1 we have

$$\begin{aligned}\mathcal{R}_{n,\alpha}(\varepsilon) &\geq \inf_{Q \in \mathcal{Q}_\alpha} \{1 - \mathrm{TV}(QM_0^n, QM_1^n)\} \\ &\geq \inf_{Q \in \mathcal{Q}_\alpha} \frac{1}{2} \exp\bigl(-\mathrm{KL}(QM_0^n, QM_1^n)\bigr) \\ &\geq \frac{1}{2} \exp\bigl(-4n(e^\alpha - 1)^2 \mathrm{TV}(M_0, M_1)^2\bigr).\end{aligned}$$

Choosing the corruption distributions appropriately we have

$$\mathrm{TV}(M_0, M_1) = (1 - \varepsilon)\mathrm{TV}(P_0, P_1) - \varepsilon.$$

For $\alpha \in (0, 1]$ this leads to

$$\mathcal{R}_{n,\alpha}(\varepsilon) \geq \frac{1}{2} \exp[-16n\alpha^2 \{\mathrm{TV}(P_0, P_1) - \varepsilon/(1 - \varepsilon)\}_+^2].$$

Simple hypothesis testing

For combined error rate ≤ 0.1 we require:

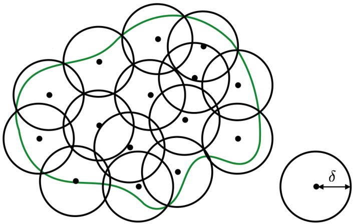
- Classical model: $H(P_0, P_1) \gtrsim 1/\sqrt{n}$;
- ε-Huber with $n = \infty$: $\mathrm{TV}(P_0, P_1) > \varepsilon/(1-\varepsilon)$;
- α-LDP: $\mathrm{TV}(P_0, P_1) \gtrsim 1/\sqrt{n\alpha^2}$;
- α-LDP and ε-Huber: $\mathrm{TV}(P_0, P_1) \gtrsim \varepsilon/(1-\varepsilon) + 1/\sqrt{n\alpha^2}$

With a suitably-chosen privacy mechanism, an existing robust method is minimax rate optimal.

Scheffé tournament

Optimal robust procedures can often be found by a Scheffé tournament approach (Devroye and Lugosi, 2001; Chen et al., 2016).

Find a δ-covering set $\{\theta_1, \ldots, \theta_m\}$ of Θ and select the hypothesis θ_j that is rejected least often in pairwise tests.

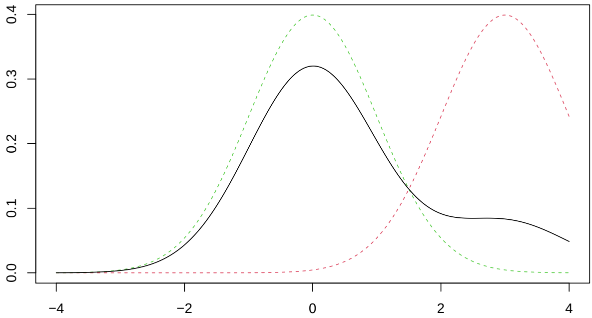


Credit: Han Bao

Gopi et al. (2020) shows that hypothesis selection is exponentially more difficult under α-LDP. Non-private procedures require $n \gtrsim \log(m)/\delta^2$ while α-LDP procedures require $n\alpha^2 \gtrsim m/\delta^2$.

Scheffé tournament

Consider $\mathcal{P} = \{\mathcal{N}(\mu, 1) : \mu \in [-1, 1]\}$. Here $m \asymp 1/\delta$ so selection of the closest hypothesis requires $n\alpha^2 \gtrsim \delta^{-3}$. Thus, tournament estimators have convergence rate bounded below by $(n\alpha^2)^{-1/3}$.



However, estimation at the rate $(n\alpha^2)^{-1/2}$ is possible here.

We therefore take problem-specific approaches.

Mean estimation

Robust mean estimation

Here we take

$$\mathcal{P} = \mathcal{P}_k(D,k) := \left\{ P : \mu = \mathbb{E}_P X \in [-D, D],\, \mathbb{E}_P(|X-\mu|^k) \leq 1 \right\}$$

and aim to estimate μ under squared error loss.

Theorem

We find α-LDP $\hat{\mu}$ with

$$\mathbb{E}\{(\hat{\mu}-\mu)^2\} \lesssim (n\alpha^2)^{-\frac{k-1}{k}} \vee \varepsilon^{2-2/k}$$

whenever $\max(\varepsilon, \log(D)/(n\alpha^2)) \leq c$. *This is minimax rate optimal.*

In the classical model with $D = \infty$ we have rate $n^{-\min(2\frac{k-1}{k},1)}$.

In the ε-Huber model with $D = \infty$ and $k \geq 2$ the rate is $\max(1/n, \varepsilon^{2-2/k})$.

Under α-LDP the rate is $(n\alpha^2)^{-\frac{k-1}{k}}$ when $\log(D)/(n\alpha^2) \leq c$.

Large parameter spaces

Previous literature has discussed issues with large D / unbounded parameter spaces in local and central models (Duchi et al., 2013; Brunel and Avella-Medina, 2020; Kamath et al., 2021) .

For constant D, one approach is to use the Laplace mechanism

$$Z_i = [X_i]_M + \frac{2M}{\alpha} W_i, \quad i = 1, \ldots, n,$$

where $[\cdot]_M = \max\{-M, \min(\cdot, M)\}$ and $W_1, \ldots, W_n \sim \mathrm{Laplace}(1)$, and take $\hat{\mu} = \bar{Z}_n$.

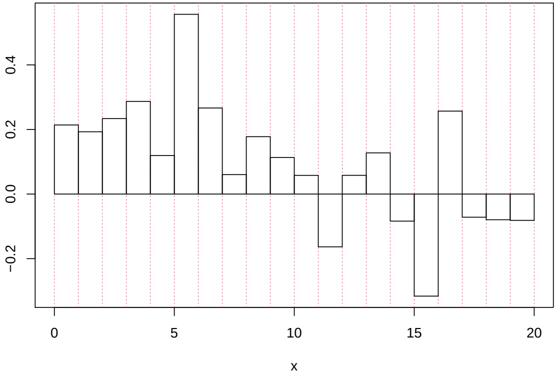
With $M \asymp D^{1/k} \min\{\varepsilon^{-1/k}, (n\alpha^2)^{1/(2k)}\}$ we have

$$\mathbb{E}\{(\hat{\mu} - \mu)^2\} \lesssim D^{2/k} \max\{(n\alpha^2)^{-\frac{k-1}{k}}, \varepsilon^{2-2/k}\},$$

which is optimal when D is constant.

Optimal mean estimator

Find $J = \operatorname{argmax}_{j=1,\dots,2D/r} \sum_{i=1}^{n/2} \left(\mathbb{1}_{\{-D+(j-1)r \le X_i < -D+jr\}} + \frac{2}{\alpha} W_{ij} \right)$ with $r = 100^{1/k}$.



Set[2] $\hat{\mu} = J + \frac{2}{n} \sum_{i=n/2+1}^{n} \left([X_i - J]_M + \frac{2M}{\alpha} W_i \right)$ with $M \asymp \varepsilon^{-\frac{1}{k}} \wedge (n\alpha^2)^{\frac{1}{2k}}$.

[2] There exists a non-interactive procedure with the same guarantees.

Density estimation

Density estimation

We consider density estimation problems with L_2 and L_∞ loss. We show that basis expansion estimators due to Duchi et al. (2018); Butucea et al. (2020) are robust against contamination.

We consider Sobolev-smooth densities

$$\mathcal{F}_\beta = \left\{ f : [0,1] \to \mathbb{R}_+ : \int_0^1 f = 1, \ \sum_{j=1}^{\infty} j^{2\beta} \left(\int_0^1 f \gamma_j \right)^2 \leq r^2 \right\}$$

where (γ_j) is an orthonormal basis for $L^2[0,1]$.

Density estimation

For L_∞ loss we consider the wavelet estimator of Butucea et al. (2020)

$$\hat{f} = \sum_{j=-1}^{J} \sum_{k} \hat{\beta}_{jk}\psi_{jk}, \quad \hat{\beta}_{jk} = \frac{1}{n}\sum_{i=1}^{n}\left\{\psi_{jk}(X_i) + \frac{C2^{J/2}}{\alpha}W_{ijk}\right\}$$

with J chosen such that $2^J \asymp \{\frac{\log(n\alpha^2)}{n\alpha^2}\}^{-\frac{1}{2\beta+1}} \wedge \varepsilon^{-\frac{2}{2\beta+1}}$.

Lower bounds follow from a combination of Butucea et al. (2020) and Uppal et al. (2020).

We find that

$$\mathcal{R}_{n,\alpha}(\mathcal{F}_\beta, \|\cdot\|_\infty, \varepsilon) \asymp \left\{\frac{\log(n\alpha^2)}{n\alpha^2}\right\}^{\frac{2\beta-1}{4\beta+2}} \vee \varepsilon^{\frac{2\beta-1}{2\beta+1}}.$$

Density estimation

By considering the density estimator of Duchi et al. (2018) based on the trigonometric basis, we show that

$$\mathcal{R}_{n,\alpha}(\mathcal{F}_\beta, \|\cdot\|_2^2, \varepsilon) \asymp (n\alpha^2)^{-\frac{2\beta}{2\beta+2}} \vee \varepsilon^{\frac{4\beta}{2\beta+1}}.$$

Lower bounds follow from a combination of Duchi et al. (2018) and Uppal et al. (2020).

Here, existing LDP methods are automatically robust.

Median estimation

Median estimation

Want to estimate $\theta(P) = \mathrm{med}(P)$ over

$$\mathcal{P}_r = \{P : |\theta(P)| \leq r, \mathbb{E}_P|X| < \infty\}$$

with loss function the excess risk $R(\hat{\theta}) - R(\theta(P))$ where $R(\cdot) = \mathbb{E}_P|X - \cdot|$.

We show the optimal rate to be

$$\mathcal{R}_{n,\alpha}(\varepsilon) \asymp \frac{r}{\sqrt{n\alpha^2}} \vee (r\varepsilon).$$

Stochastic gradient descent

This rate is attained by a general private stochastic gradient descent algorithm (Duchi et al., 2018).

Let $W_1, \ldots, W_n$ be i.i.d. in $\{-1, 1\}$ with $\mathbb{P}(W_1 = 1) = e^\alpha/(1 + e^\alpha)$ and let $\eta_1 \geq \ldots \geq \eta_n$ be step sizes. Iterate according to

$$\theta_{i+1} = \max\{-r, \min(\theta_i - \eta_i Z_i, r)\}$$

where

$$Z_i = \frac{e^\alpha + 1}{e^\alpha - 1} W_i \,\mathrm{sign}(\theta_i - X_i).$$

The final estimator is

$$\hat{\theta} = \frac{\sum_{i=1}^n \eta_i \theta_i}{\sum_{i=1}^n \eta_i}.$$

Conclusion

We identify procedures that are simultaneously privacy-preserving and robust for a range of statistical problems.

The difficulty of private hypothesis selection makes a general theory more difficult...

But many existing private procedures are automatically robust, and ideas from robust statistics are useful for constructing α-LDP methods.

Thank you!

Li, M., B. and Yu, Y. (2023) On robustness and local differential privacy, *Ann. Statist.*, **51**(2), 717–737.

References

Acharya, J., Sun, Z. and Zhang, H. (2021) Robust testing and estimation under manipulation attacks. *ICML*.

Asi, H., Ullman, J. and Zakynthinou, L. (2023) From robustness to privacy and back. `arXiv:2302.01855`

Avella-Medina, M. (2021) Privacy-preserving parametric inference: a case for robust statistics. *J. Amer. Statist. Assoc.*, **116**, 969–983.

Brunel, V.-E. and Avella-Medina, M. (2020) Propose, test, release: Differentially private estimation with high probability. `arXiv:2002.08774`.

Butucea, C., Dubois, A., Kroll, M. and Saumard, A. (2020) Local differential privacy: Elbow effect in optimal density estimation and adaptation of Besov ellipsoids. *Bernoulli*, **26**, 1727–1764.

Chen, M., Gao, C. and Ren, Z. (2016) A general decision theory for Huber's ε-contamination model. *Electronic Journal of Statistics*, **10**, 3752–3774.

Cheu, A., Smith, A. and Ullman, J. (2021) Manipulation attacks in local differential privacy. *2021 IEEE Symposium of Security and Privacy*, 883–900.

Chhor, J. and Sentenac, F. (2023) Robust estimation of discrete distributions under local differential privacy. *The 34th International Conference on Algorithmic Learning Theory*.

Cohen, A. and Nassim, K. (2020) Towards formalizing the GDPR's notion of singling out. *PNAS*, **117**, 8344–8352.

Devroye, L. and Lugosi, G. (2001) *Combinatorial Methods in Density Estimation*, Springer.

Ding, B., Kulkarni, J., and Yekhanin S. (2017) Collecting telemetry data privately. *NeurIPS*, 3571–3580.

References

Donoho, D. L. and Liu, R. C. (1991) Geometrizing rates of convergence, III. *Ann. Statist.*, **19**, 668–701.

Duchi, J. C., Jordan, M. I. and Wainwright, M. J. (2013) Local privacy and minimax bounds: Sharp rates for probability estimation. *NeurIPS*, 1529–1537.

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

Dwork, C. and Lei, J. (2009) Differential privacy and robust statistics. *Annual ACM Symposium on Theory of Computing*, 371–380.

Dwork, C., McSherry, F., Nissim, K. and Smith, A. (2006) Calibrating noise to sensitivity in private data analysis. *Theory of Cryptography*, 265–284.

Dwork, C. (2019) Differential privacy and the US census. *PODS*.

Erlingsson, U., Pihur, V. and Korolova, A. (2014) Rappor: Randomized aggregatable privacy-preserving ordinal response. *Proc. 2014 ACM SIGSAC conference on computer and communications security*, 1054–1067.

Gopi, S., Kamath, G., Kulkarni, J., Nikolov, A., Wu, Z. S. and Zhang, H. (2020) Locally private hypothesis selection. *Conference on Learning Theory*, 1785–1816.

Hopkins, S. B., Kamath, G., Majid, M. and Narayanan, S. (2023) Robustness implies privacy in statistical estimation. *Proceedings of the 55th Annual ACM Symposium on Theory of Computing*, 497–506.

Kamath, G., Mouzakis, A., Singhal, V., Steinke, T. and Ullman, J. (2021) A private and computationally-efficient estimator for unbounded Gaussians. `arXiv:2111.04609`.

References

Near, J. (2018) Differential privacy at scale: Uber and Berkeley collaboration. *Enigma 2018*.

Prasad, A., Balakrishnan, S. and Ravikumar, P. (2019) A unified approach to robust mean estimation. *Available at*, `arXiv:1907.00927`.

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

Sweeney, L. (2002) k-anonymity: A model for protecting privacy. *Fuzziness and Knowledge-Based Systems*, **10**, 557–570.

Tang J., Korolova, A., Bai, X., Wang, X. and Wang X. (2017) Privacy loss in Apple's implementation of differential privacy on macOS 10.12. *Available at* `arXiv:1709.02753`.

Uppal, A., Singh, S. and Poczós, B. (2020) Robust density estimation under Besov IPM losses. *NeurIPS 33*.

Warner, S. L. (1965) Randomized sesponse: A survey technique for eliminating evasive answer bias. *J. Amer. Statist. Assoc.*, **60**, 63–69.