Privacy guarantees in statistical estimation: How to formalize the problem?

Martin Wainwright

UC Berkeley
Departments of Statistics, and EECS

van Dantzig Seminar, University of Leiden

The modern landscape

Modern data sets are often very large

- biological data (genes, proteins, etc.)
- medical imaging (MRI, fMRI etc.)
- astronomy datasets
- social network data
- recommender systems (Amazon, Netflix etc.)

The modern landscape

Modern data sets are often very large

- biological data (genes, proteins, etc.)
- medical imaging (MRI, fMRI etc.)
- astronomy datasets
- social network data
- recommender systems (Amazon, Netflix etc.)

Statistical considerations interact with:

1. Computational constraints: (low-order) polynomial-time is essential!

The modern landscape

Modern data sets are often very large

- biological data (genes, proteins, etc.)
- medical imaging (MRI, fMRI etc.)
- astronomy datasets
- social network data
- recommender systems (Amazon, Netflix etc.)

Statistical considerations interact with:

1. Computational constraints: (low-order) polynomial-time is essential!
2. Communication/storage constraints: distributed implementations are often needed

The modern landscape

Modern data sets are often very large

- biological data (genes, proteins, etc.)
- medical imaging (MRI, fMRI etc.)
- astronomy datasets
- social network data
- recommender systems (Amazon, Netflix etc.)

Statistical considerations interact with:

1. Computational constraints: (low-order) polynomial-time is essential!
2. Communication/storage constraints: distributed implementations are often needed
3. Privacy constraints: tension between hiding/sharing data

From Classical Minimax Risk...

Choose estimator to minimize the worst-case risk

$$\text{Classical minimax risk} = \inf_{\widehat{\theta}_n} \sup_{\theta \in \Omega} \mathbb{E}\Big[\mathcal{L}(\widehat{\theta}_n, \theta).\Big]$$

Abraham Wald
1902–1950

From Classical Minimax Risk...

Choose estimator to minimize the worst-case risk

$$\text{Classical minimax risk} = \inf_{\widehat{\theta}_n} \sup_{\theta \in \Omega} \mathbb{E}\Big[\mathcal{L}(\widehat{\theta}_n, \theta).\Big]$$

Two party game:

- Nature chooses parameter $\theta \in \Omega$ in a potentially adversarial manner
- Statistician takes infimum over all estimators:

$$\underbrace{(X_1, \ldots, X_n) \mapsto \widehat{\theta}_n \in \Omega}_{\text{arbitrary measurable function}}$$

Abraham Wald
1902–1950

From Classical Minimax Risk...

Choose estimator to minimize the worst-case risk

$$\text{Classical minimax risk} = \inf_{\widehat{\theta}_n} \sup_{\theta \in \Omega} \mathbb{E}\Big[\mathcal{L}(\widehat{\theta}_n, \theta).\Big]$$

Abraham Wald
1902–1950

Two party game:

- Nature chooses parameter $\theta \in \Omega$ in a potentially adversarial manner
- Statistician takes infimum over all estimators:

$$\underbrace{(X_1, \ldots, X_n) \mapsto \widehat{\theta}_n \in \Omega}_{\text{arbitrary measurable function}}$$

Classical questions about minimax risk:

- how fast does it decay as a function of sample size n?
- dependence on dimensionality, smoothness etc.?
- characterization of optimal estimators?

....to Constrained Minimax Risk

Classical framework imposes no constraints on the choice of estimators $\widehat{\theta}_n$.

....to Constrained Minimax Risk

Classical framework imposes no constraints on the choice of estimators $\widehat{\theta}_n$.

- Unbounded memory and computational power.
- Provided centralized access to all n samples.
- Data is fully revealed: no privacy-preserving properties.

....to Constrained Minimax Risk

Classical framework imposes no constraints on the choice of estimators $\widehat{\theta}_n$.

- Unbounded memory and computational power.
- Provided centralized access to all n samples.
- Data is fully revealed: no privacy-preserving properties.

On-going research: statistical minimax with constraints

- Computationally-constrained estimators
  (e.g., Rigollet & Berthet, 2013; Ma & Wu, 2014; Zhang, W. & Jordan, 2014)
- Communication constraints
  (e.g., Zhang et al., 2013; Ma et al. 2014; Braverman et al., 2015)
- Privacy constraints (e.g., Dwork, 2006; Hardt & Rothblum, 2010; Hall et al., 2011; Duchi, W. & Jordan, 2013)

Why be concerned with privacy?

Many sources of data have both statistical utility and privacy concerns.

(a) Personal genome project

Why be concerned with privacy?

Many sources of data have both statistical utility and privacy concerns.

(a) Personal genome project

(b) Privacy breach
Scientific American, August 2013

Why be concerned with privacy?

Many sources of data have both statistical utility and privacy concerns.

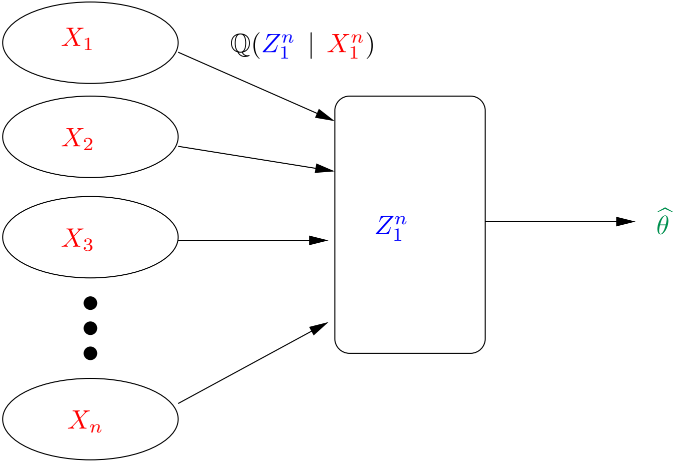
(a) Personal genome project

(b) Privacy breach
Scientific American, August 2013

Question

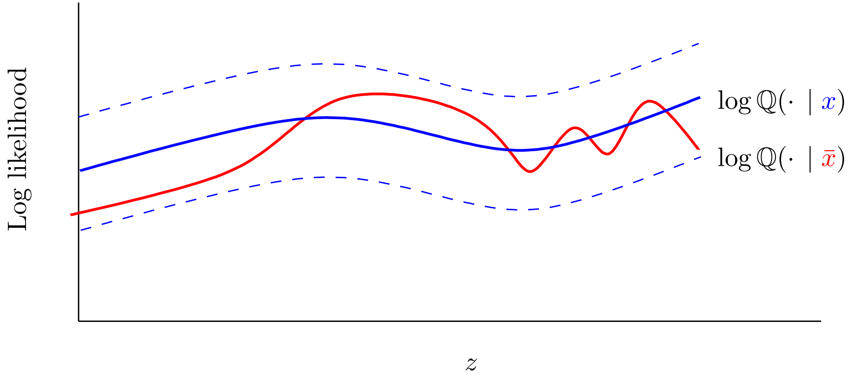
How to obtain principled tradeoffs between these competing criteria?

Basic model of local privacy



- each individual $i \in \{1, 2, \ldots, n\}$ has personal data $X_i \sim \mathbb{P}_{\theta^*}$
- conditional distribution $\mathbb{Q}$ between private data X_1^n and public data Z_1^n
- estimator $Z_1^n \mapsto \widehat{\theta}$ of unknown parameter θ^*.

Local privacy at level α



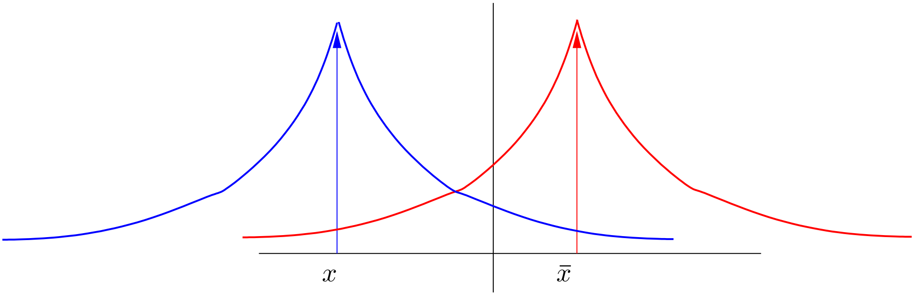
Definition

Conditional distribution $\mathbb{Q}$ is locally α-differentially private if

$$e^{-\alpha} \le \sup_z \frac{\mathbb{Q}(z \mid x_1^n)}{\mathbb{Q}(z \mid \bar{x}_1^n)} \le e^{\alpha} \qquad \text{for all } x_1^n \text{ and } \bar{x}_1^n \text{ such that } d_{\text{HAM}}(x_1^n, \bar{x}_1^n) = 1.$$

(Dwork et al., 2006)

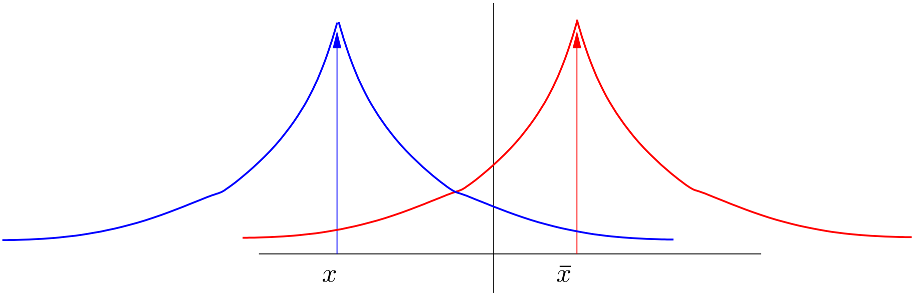
Illustration of Laplacian mechanism



Add α-Laplacian noise (Dwork et al., 2006)

$$Z = x + W, \qquad \text{where } W \text{ has density} \propto e^{-\alpha\,|w|}$$

Illustration of Laplacian mechanism



Add α-Laplacian noise (Dwork et al., 2006)

$$Z = x + W, \qquad \text{where } W \text{ has density} \propto e^{-\alpha\,|w|}$$

For all $x, x' \in [-1/2, 1/2]$:

$$\sup_{z \in \mathbb{R}} \left| \log \frac{\mathbb{Q}(z \mid x)}{\mathbb{Q}(z \mid \bar{x})} \right| \;=\; \alpha \left| \sup_{z \in \mathbb{R}} |z - x| - |z - \bar{x}| \right| \;\leq\; \alpha.$$

Various mechanisms for α-privacy

Choices from past work:

- randomized response in survey questions (Warner, 1965)
- Laplacian noise (Dwork et al., 2006)
- exponential mechanism (McSherry & Talwar, 2007)

Various mechanisms for α-privacy

Choices from past work:

- randomized response in survey questions (Warner, 1965)
- Laplacian noise (Dwork et al., 2006)
- exponential mechanism (McSherry & Talwar, 2007)

Some past work on privacy and estimation:

- local differential privacy and PAC learning (Kasiviswanathan et al., 2008)
- linear queries over discrete-valued data sets (Hardt & Rothblum, 2010)
- global differential privacy and histogram estimators (Hall et al., 2011)
- lower bounds for certain 1-D statistics (Chaudhuri & Hsu, 2012)

Various mechanisms for α-privacy

Choices from past work:

- randomized response in survey questions (Warner, 1965)
- Laplacian noise (Dwork et al., 2006)
- exponential mechanism (McSherry & Talwar, 2007)

Some past work on privacy and estimation:

- local differential privacy and PAC learning (Kasiviswanathan et al., 2008)
- linear queries over discrete-valued data sets (Hardt & Rothblum, 2010)
- global differential privacy and histogram estimators (Hall et al., 2011)
- lower bounds for certain 1-D statistics (Chaudhuri & Hsu, 2012)

Questions:

- Can we provide a general characterization of trade-offs between α-privacy and statistical utility?
- Can we identify optimal "mechanisms" for privacy?

Minimax optimality with α-privacy

- family of distributions $\{\mathbb{P} \in \mathcal{F}\}$, and functional $\mathbb{P} \mapsto \theta(\mathbb{P})$
- samples $X_1^n \equiv \{X_1, \ldots, X_n\} \sim \mathbb{P}$ and estimator $X_1^n \mapsto \widehat{\theta}(X_1^n)$
- loss function (e.g., squared error, 0-1 error, ℓ_1-error)

$$(\widehat{\theta}, \theta) \qquad \mapsto \qquad \underbrace{\mathcal{L}(\widehat{\theta}, \theta)}_{\text{quality of } \widehat{\theta} \text{ as estimate of } \theta}$$

Minimax optimality with α-privacy

- family of distributions $\{\mathbb{P} \in \mathcal{F}\}$, and functional $\mathbb{P} \mapsto \theta(\mathbb{P})$
- samples $X_1^n \equiv \{X_1, \ldots, X_n\} \sim \mathbb{P}$ and estimator $X_1^n \mapsto \widehat{\theta}(X_1^n)$
- loss function (e.g., squared error, 0-1 error, ℓ_1-error)

$$(\widehat{\theta}, \theta) \qquad \mapsto \qquad \underbrace{\mathcal{L}(\widehat{\theta}, \theta)}_{\text{quality of } \widehat{\theta} \text{ as estimate of } \theta}$$

Ordinary minimax risk:

$$\mathfrak{M}_n(\mathcal{F}) := \underbrace{\inf_{\widehat{\theta}}}_{\text{Best estimator}} \underbrace{\sup_{\mathbb{P} \in \mathcal{F}}}_{\text{Worst-case distribution}} \mathbb{E}\Big[\mathcal{L}\big(\widehat{\theta}(X_1^n), \theta(\mathbb{P})\big)\Big]$$

Minimax optimality with α-privacy

- family of distributions $\{\mathbb{P} \in \mathcal{F}\}$, and functional $\mathbb{P} \mapsto \theta(\mathbb{P})$
- samples $X_1^n \equiv \{X_1, \ldots, X_n\} \sim \mathbb{P}$ and estimator $X_1^n \mapsto \widehat{\theta}(X_1^n)$
- loss function (e.g., squared error, 0-1 error, ℓ_1-error)

$$(\widehat{\theta}, \theta) \quad \mapsto \quad \underbrace{\mathcal{L}(\widehat{\theta}, \theta)}_{\text{quality of } \widehat{\theta} \text{ as estimate of } \theta}$$

Ordinary minimax risk:

$$\mathfrak{M}_n(\mathcal{F}) := \underbrace{\inf_{\widehat{\theta}}}_{\text{Best estimator}} \quad \underbrace{\sup_{\mathbb{P} \in \mathcal{F}}}_{\text{Worst-case distribution}} \mathbb{E}\Big[\mathcal{L}\big(\widehat{\theta}(X_1^n), \theta(\mathbb{P})\big)\Big]$$

Minimax risk with α-privacy

Estimators now depend on privatized samples Z_1^n

$$\mathfrak{M}_n(\alpha; \mathcal{F}) := \underbrace{\inf_{\mathbb{Q} \in \mathcal{Q}_\alpha}}_{\text{Best } \alpha\text{-private channel}} \quad \inf_{\widehat{\theta}} \sup_{\mathbb{P} \in \mathcal{F}} \mathbb{E}\Big[\mathcal{L}\big(\widehat{\theta}(Z_1^n), \; \theta(\mathbb{P})\big)\Big]$$

Vignette A: α-private location estimation

Consider estimation of mean functional $\theta(\mathbb{P}) = \mathbb{E}[X]$ over family

$$\mathcal{F}_k := \left\{ \text{distributions } \mathbb{P} \text{ such that } \mathbb{E}[X] \in [-1, 1] \text{ and } \mathbb{E}[|X|^k|] \leq 1 \right\}$$

Vignette A: α-private location estimation

Consider estimation of mean functional $\theta(\mathbb{P}) = \mathbb{E}[X]$ over family

$$\mathcal{F}_k := \left\{ \text{distributions } \mathbb{P} \text{ such that } \mathbb{E}[X] \in [-1, 1] \text{ and } \mathbb{E}[|X|^k] \le 1 \right\}$$

For $k \geq 2$ and non-private setting, sample mean $\widehat{\theta} = \frac{1}{n} \sum_{i=1}^n X_i$ achieves rate $1/n$.

Vignette A: α-private location estimation

Consider estimation of mean functional $\theta(\mathbb{P}) = \mathbb{E}[X]$ over family

$$\mathcal{F}_k := \left\{\text{distributions } \mathbb{P} \text{ such that } \mathbb{E}[X] \in [-1, 1] \text{ and } \mathbb{E}[|X|^k|] \leq 1\right\}$$

For $k \geq 2$ and non-private setting, sample mean $\widehat{\theta} = \frac{1}{n}\sum_{i=1}^{n} X_i$ achieves rate $1/n$.

Theorem

For all $k \geq 2$ and $\alpha \in (0, 1/4]$, the α-private minimax risk scales as

$$\mathfrak{M}_n(\alpha; \mathcal{F}_k) \asymp \min\left\{1,\ \left(\frac{1}{\alpha^2 n}\right)^{\frac{k-1}{k}}\right\}.$$

Vignette A: α-private location estimation

Consider estimation of mean functional $\theta(\mathbb{P}) = \mathbb{E}[X]$ over family

$$\mathcal{F}_k := \big\{\text{distributions } \mathbb{P} \text{ such that } \mathbb{E}[X] \in [-1,1] \text{ and } \mathbb{E}[|X|^k|] \leq 1\big\}$$

For $k \geq 2$ and non-private setting, sample mean $\widehat{\theta} = \frac{1}{n}\sum_{i=1}^n X_i$ achieves rate $1/n$.

Theorem

For all $k \geq 2$ and $\alpha \in (0, 1/4]$, the α-private minimax risk scales as

$$\mathfrak{M}_n(\alpha; \mathcal{F}_k) \asymp \min\Big\{1,\ \big(\frac{1}{\alpha^2 n}\big)^{\frac{k-1}{k}}\Big\}.$$

Examples:

- For two moments $k = 2$, rate is reduced from parametric $1/n$ to $1/(\alpha\sqrt{n})$.
- As $k \to \infty$ (roughly bounded random variables), private rate converges to the parametric one (with a pre-factor of $1/\alpha^2$).

Sample size reduction: $n \mapsto \alpha^2 n$

Given an α-private channel, any pair $\{\mathbb{P}_j, j = 1, 2\}$ induces marginals

$$\mathbb{M}_j^n(A) := \int \mathbb{Q}(A \mid x_1, \ldots, x_n) d\mathbb{P}_j^n(x_1, \ldots, x_n) \quad \text{for } j = 1, 2.$$

Sample size reduction: $n \mapsto \alpha^2 n$

Given an α-private channel, any pair $\{\mathbb{P}_j, j = 1, 2\}$ induces marginals

$$\mathbb{M}_j^n(A) := \int \mathbb{Q}(A \mid x_1, \ldots, x_n) d\mathbb{P}_j^n(x_1, \ldots, x_n) \quad \text{for } j = 1, 2.$$

Question:

How much "contraction" induced by local α-privacy?

Sample size reduction: $n \mapsto \alpha^2 n$

Given an α-private channel, any pair $\{\mathbb{P}_j, j = 1, 2\}$ induces marginals

$$\mathbb{M}_j^n(A) := \int \mathbb{Q}(A \mid x_1, \ldots, x_n) d\mathbb{P}_j^n(x_1, \ldots, x_n) \quad \text{for } j = 1, 2.$$

Question:

How much "contraction" induced by local α-privacy?

Theorem (Duchi, W., & Jordan, 2013)

Given n i.i.d. samples from any α-private channel with $\alpha \in (0, 1/2]$, we have

$$\frac{1}{n}\Big\{ \underbrace{D(\mathbb{M}_1^n \parallel \mathbb{M}_0^n) + D(\mathbb{M}_0^n \parallel \mathbb{M}_1^n)}_{\textit{Symmetrized KL divergence}} \Big\} \precsim (e^{\alpha} - 1)^2 \quad \underbrace{\|\mathbb{P}_1 - \mathbb{P}_0\|_{TV}^2}_{\textit{Total variation}}$$

Sample size reduction: $n \mapsto \alpha^2 n$

Given an α-private channel, any pair $\{\mathbb{P}_j, j = 1, 2\}$ induces marginals

$$\mathbb{M}_j^n(A) := \int \mathbb{Q}(A \mid x_1, \ldots, x_n) d\mathbb{P}_j^n(x_1, \ldots, x_n) \quad \text{for } j = 1, 2.$$

Question:

How much "contraction" induced by local α-privacy?

Theorem (Duchi, W., & Jordan, 2013)

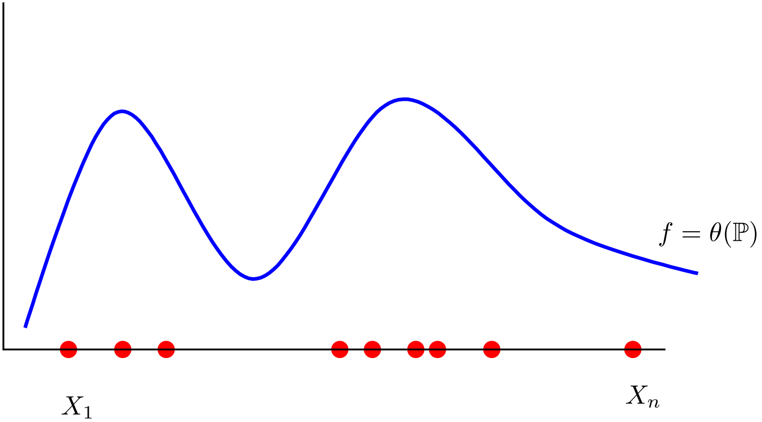
Given n i.i.d. samples from any α-private channel with $\alpha \in (0, 1/2]$, we have

$$\frac{1}{n}\Big\{ \underbrace{D(\mathbb{M}_1^n \parallel \mathbb{M}_0^n) + D(\mathbb{M}_0^n \parallel \mathbb{M}_1^n)}_{\textit{Symmetrized KL divergence}} \Big\} \precsim (e^\alpha - 1)^2 \quad \underbrace{\|\mathbb{P}_1 - \mathbb{P}_0\|_{TV}^2}_{\textit{Total variation}}$$

Note that $(e^\alpha - 1)^2 \precsim \alpha^2$ for $\alpha \in (0, 1/4]$.

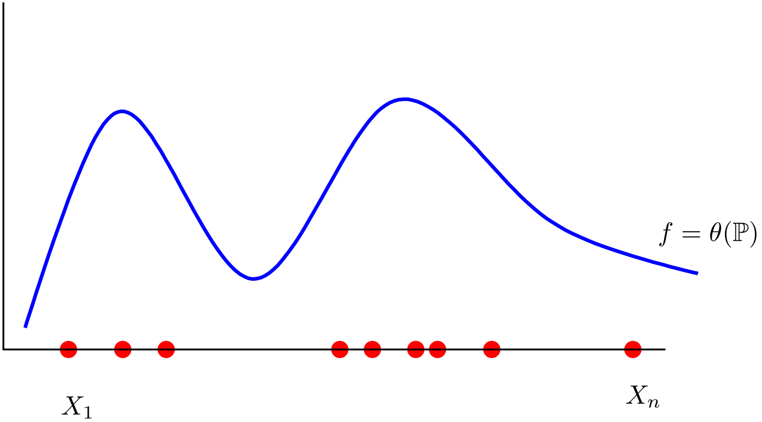
Vignette B: Non-parametric density estimation

Suppose that we want to estimate the quantity $\mathbb{P} \mapsto \theta(\mathbb{P}) \equiv$ density f



Vignette B: Non-parametric density estimation

Suppose that we want to estimate the quantity $\mathbb{P} \mapsto \theta(\mathbb{P}) \equiv$ density f



Ordinary minimax rates depend on number of derivatives $\beta > 1/2$ of density f:

$$\mathfrak{M}_n\big(\mathcal{F}(\beta)\big) \asymp \Big(\frac{1}{n}\Big)^{\frac{2\beta}{2\beta+1}}.$$

(Ibragimov & Hasminskii, 1978; Stone, 1980)

Optimal rates for α-private density estimation

Consider density estimation based on α-private views $(Z_1, \ldots, Z_n)$ of original samples $(X_1, \ldots, X_n)$.

Optimal rates for α-private density estimation

Consider density estimation based on α-private views $(Z_1, \ldots, Z_n)$ of original samples $(X_1, \ldots, X_n)$.

Theorem (Duchi, W. & Jordan, 2013)

For all privacy levels $\alpha \in (0, 1/4]$ and smoothness levels $\beta > 1/2$:

$$\mathfrak{M}_n\big(\alpha; \mathcal{F}(\beta)\big) \asymp \Big(\frac{1}{\alpha^2 n}\Big)^{\frac{2\beta}{2\beta+2}}$$

Optimal rates for α-private density estimation

Consider density estimation based on α-private views $(Z_1, \ldots, Z_n)$ of original samples $(X_1, \ldots, X_n)$.

Theorem (Duchi, W. & Jordan, 2013)

For all privacy levels $\alpha \in (0, 1/4]$ and smoothness levels $\beta > 1/2$:

$$\mathfrak{M}_n\big(\alpha; \mathcal{F}(\beta)\big) \asymp \Big(\frac{1}{\alpha^2 n}\Big)^{\frac{2\beta}{2\beta+2}}$$

- can give a simple/explicit scheme that achieves this optimal rate.

Optimal rates for α-private density estimation

Consider density estimation based on α-private views $(Z_1, \ldots, Z_n)$ of original samples $(X_1, \ldots, X_n)$.

Theorem (Duchi, W. & Jordan, 2013)

For all privacy levels $\alpha \in (0, 1/4]$ and smoothness levels $\beta > 1/2$:

$$\mathfrak{M}_n\big(\alpha; \mathcal{F}(\beta)\big) \asymp \left(\frac{1}{\alpha^2 n}\right)^{\frac{2\beta}{2\beta+2}}$$

- can give a simple/explicit scheme that achieves this optimal rate.
- contrast with classical rate $(1/n)^{\frac{2\beta}{2\beta+1}}$: Penalty for privacy can be significant!

Optimal rates for α-private density estimation

Consider density estimation based on α-private views $(Z_1, \ldots, Z_n)$ of original samples $(X_1, \ldots, X_n)$.

Theorem (Duchi, W. & Jordan, 2013)

For all privacy levels $\alpha \in (0, 1/4]$ and smoothness levels $\beta > 1/2$:

$$\mathfrak{M}_n(\alpha; \mathcal{F}(\beta)) \asymp \left(\frac{1}{\alpha^2 n}\right)^{\frac{2\beta}{2\beta+2}}$$

- can give a simple/explicit scheme that achieves this optimal rate.
- contrast with classical rate $(1/n)^{\frac{2\beta}{2\beta+1}}$: Penalty for privacy can be significant!

Example: How many samples $N(\epsilon)$ to achieve error $\epsilon = 0.01$ for Lipschitz densities $(\beta = 1)$?

Classical case $N \approx 1,000$ versus Private case $N \approx 10,000$.

How to achieve a matching upper bound?

Naive approach: Add Laplacian noise directly to samples

$$Z_i = X_i + W_i, \qquad \text{with } W_i \sim \frac{\alpha}{2} e^{-\alpha|w|}$$

How to achieve a matching upper bound?

Naive approach: Add Laplacian noise directly to samples

$$Z_i = X_i + W_i, \qquad \text{with } W_i \sim \tfrac{\alpha}{2} e^{-\alpha|w|}$$

Transforms problem into non-parametric deconvolution.

How to achieve a matching upper bound?

Naive approach: Add Laplacian noise directly to samples

$$Z_i = X_i + W_i, \qquad \text{with } W_i \sim \tfrac{\alpha}{2} e^{-\alpha |w|}$$

Transforms problem into non-parametric deconvolution.

Lower bound for this mechanism

For any estimator $\widehat{f}$ based on $(Z_1, \ldots, Z_n)$:

$$\sup_{f^* \in \mathcal{F}(\beta)} \mathbb{E}[\|\widehat{f} - f^*\|_2^2] \gtrsim \left(\frac{1}{n}\right)^{\frac{2\beta}{2\beta+5}}$$

Follows from known lower bounds for deconvolution (Carroll & Hall, 1988)

An optimal mechanism

1. For a given orthonormal basis $\{\phi_j\}_{j=1}^{\infty}$ of $L^2[0,1]$, individual i computes

   $$\Phi_1^D(X_i) := \{\phi_1(X_i), \phi_2(X_i), \dots, \phi_D(X_i)\} \qquad \text{for dimension } D \text{ to be chosen}$$

An optimal mechanism

1. For a given orthonormal basis $\{\phi_j\}_{j=1}^{\infty}$ of $L^2[0,1]$, individual i computes

   $\Phi_1^D(X_i) := \{\phi_1(X_i), \phi_2(X_i), \ldots, \phi_D(X_i)\}$ for dimension D to be chosen

2. Privatized D-dimensional vector:

$$\text{Hypercube sampling scheme with} \quad \mathbb{E}[Z_i \mid X_i] = \Phi_1^D(X_i)$$

An optimal mechanism

1. For a given orthonormal basis $\{\phi_j\}_{j=1}^{\infty}$ of $L^2[0,1]$, individual i computes

   $\Phi_1^D(X_i) := \{\phi_1(X_i), \phi_2(X_i), \ldots, \phi_D(X_i)\}$ for dimension D to be chosen

2. Privatized D-dimensional vector:

   Hypercube sampling scheme with $\mathbb{E}[Z_i \mid X_i] = \Phi_1^D(X_i)$

3. Statistician can compute noisy versions of D basis expansion coefficients

$$\widehat{B}_j = \frac{1}{n}\sum_{i=1}^{n} Z_{ij}, \quad \text{and} \quad \widehat{f} = \sum_{j=1}^{D} \widehat{B}_j \phi_j$$

An optimal mechanism

1. For a given orthonormal basis $\{\phi_j\}_{j=1}^{\infty}$ of $L^2[0,1]$, individual i computes

   $\Phi_1^D(X_i) := \{\phi_1(X_i), \phi_2(X_i), \dots, \phi_D(X_i)\}$ for dimension D to be chosen

2. Privatized D-dimensional vector:

   Hypercube sampling scheme with $\mathbb{E}[Z_i \mid X_i] = \Phi_1^D(X_i)$

3. Statistician can compute noisy versions of D basis expansion coefficients

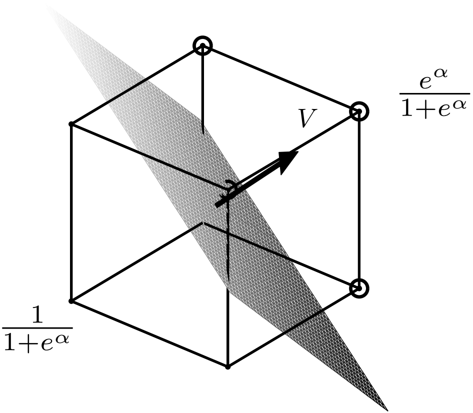
$$\widehat{B}_j = \frac{1}{n}\sum_{i=1}^{n} Z_{ij}, \quad \text{and} \quad \widehat{f} = \sum_{j=1}^{D} \widehat{B}_j \phi_j$$

Upper bound

For any $D \geq 1$, the privatized density estimate satisfies

$$\mathbb{E}\big[\|\widehat{f} - f^*\|_2^2\big] \precsim \frac{D^2}{n\alpha^2} + \frac{1}{D^{2\beta}}$$

Hypercube sampling: Optimal privacy mechanism



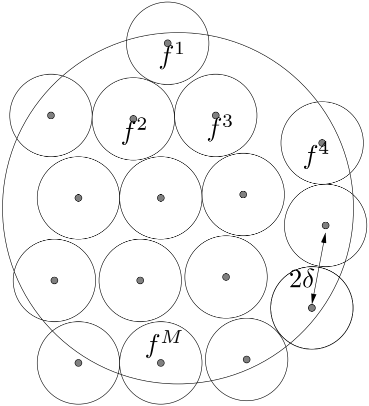
- Given $V = \Phi_1^D(X)$ with $\|V\|_\infty \le C$, form D-dimensional random vector

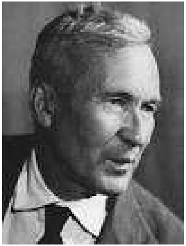
$$\widetilde{V}_j = \begin{cases} +C & \text{with prob. } \frac{1}{2} + \frac{V_j}{2C} \\ -C & \text{with prob. } \frac{1}{2} - \frac{V_j}{2C}. \end{cases}$$

- Draw $T \sim \mathrm{Ber}\left(\frac{e^\alpha}{1+e^\alpha}\right)$ and set

$$Z \sim \begin{cases} \mathrm{Uni}\left(\{-C, +C\}^D \mid \langle Z, \widetilde{V} \rangle > 0\right) & \text{if } T = 1 \\ \mathrm{Uni}\left(\{-C, +C\}^D \mid \langle Z, \widetilde{V} \rangle \le 0\right) & \text{if } T = 0 \end{cases}$$

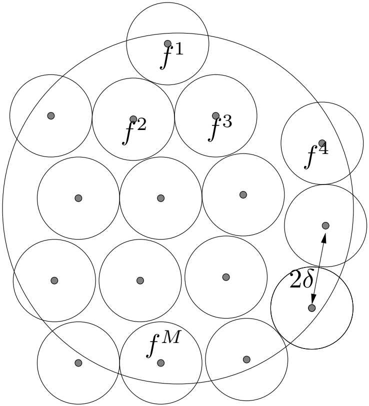
Lower bounds via metric entropy

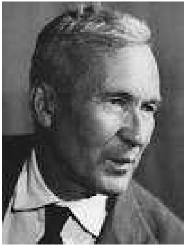




Andrey Kolmogorov
1903–1987

Lower bounds via metric entropy





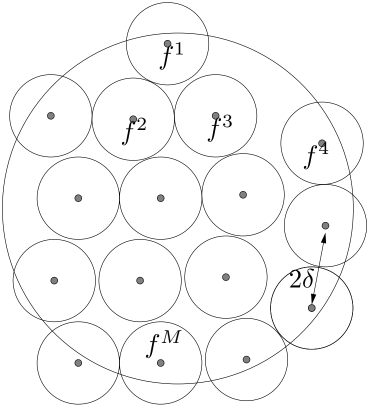
Andrey Kolmogorov
1903–1987

Packing number

Given a metric ρ and function class $\mathcal{F}$, a δ-packing is a collection $\{f^1, \ldots, f^M\}$ contained in $\mathcal{F}$ such that

$$\rho(f^j, f^k) > 2\delta \qquad \text{for all } j \neq k.$$

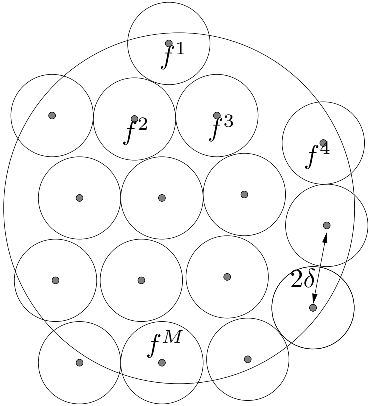
From metric entropy to hypothesis testing



Two-person game:

- Nature chooses a random index $J \in \{1, \ldots, M\}$
- Statistician estimates density based on n i.i.d. samples from f^J

From metric entropy to hypothesis testing



Two-person game:

- Nature chooses a random index $J \in \{1, \ldots, M\}$
- Statistician estimates density based on n i.i.d. samples from f^J

Reduction to hypothesis testing

Any estimator $\widehat{f}$ for which $\rho(\widehat{f}, f^J) < \delta$ with high probability can be used to decode the index J.

A quantitative data processing inequality

$$J \longrightarrow X \xrightarrow{\mathbb{Q}} Z$$

- packing index $J \in \{1, 2, \ldots, M\}$
- non-private variables $(X \mid J = j) \sim \mathbb{P}_j$
- mixture distribution $\bar{\mathbb{P}} = \frac{1}{M} \sum_{j=1}^{M} \mathbb{P}_j$.

A quantitative data processing inequality

$$J \longrightarrow X \xrightarrow{\mathbb{Q}} Z$$

- packing index $J \in \{1, 2, \ldots, M\}$
- non-private variables $(X \mid J = j) \sim \mathbb{P}_j$
- mixture distribution $\bar{\mathbb{P}} = \frac{1}{M} \sum_{j=1}^{M} \mathbb{P}_j$.

Theorem (Duchi, W. & Jordan, 2013)

For any non-interactive α-private channel $\mathbb{Q}$, we have

$$\frac{I(Z_1, \ldots, Z_n; J)}{n} \leq (e^{\alpha} - 1)^2 \underbrace{\sup_{\|\gamma\|_\infty \leq 1} \left\{ \frac{1}{M} \sum_{j=1}^{M} \left[\int_{\mathcal{X}} \gamma(x) \big(d\mathbb{P}_j(x) - d\bar{\mathbb{P}}(x) \big) \right]^2 \right\}}_{\textit{dimension-dependent contraction}}$$

High-level and extensions

High-level

Two main theorems are forms of "information contraction":

1. Pairwise contraction: consequences for Le Cam's method
2. Mutual information contraction: consequences for Fano's method

High-level and extensions

High-level

Two main theorems are forms of "information contraction":

1. Pairwise contraction: consequences for Le Cam's method
2. Mutual information contraction: consequences for Fano's method

Some extensions:

1. Matching rates for linear regression ($n \mapsto n\alpha^2$)
2. Matching rates for multinomial estimation ($n \mapsto \frac{n\alpha^2}{d}$)

High-level and extensions

High-level

Two main theorems are forms of "information contraction":

1. Pairwise contraction: consequences for Le Cam's method
2. Mutual information contraction: consequences for Fano's method

Some extensions:

1. Matching rates for linear regression ($n \mapsto n\alpha^2$)
2. Matching rates for multinomial estimation ($n \mapsto \frac{n\alpha^2}{d}$)
3. Convex risk minimization: dimension-dependent effects.
   Sparse optimization no longer depends logarithmically on dimension.

High-level and extensions

High-level

Two main theorems are forms of "information contraction":

1. Pairwise contraction: consequences for Le Cam's method
2. Mutual information contraction: consequences for Fano's method

Some extensions:

1. Matching rates for linear regression ($n \mapsto n\alpha^2$)
2. Matching rates for multinomial estimation ($n \mapsto \frac{n\alpha^2}{d}$)
3. Convex risk minimization: dimension-dependent effects.
   Sparse optimization no longer depends logarithmically on dimension.
4. Laplacian mechanism can be sub-optimal. Need to consider geometry of set.

Summary

- interesting trade-offs between privacy and statistical utility
- new notion of locally α-private minimax risk
- provided some general bounds and techniques:

bounds on total variation	useful for Le Cam's method
bounds on mutual information	useful for Fano's method

- sharp bounds for several parametric/non-parametric problems

Summary

- interesting trade-offs between privacy and statistical utility
- new notion of locally α-private minimax risk
- provided some general bounds and techniques:

bounds on total variation	useful for Le Cam's method
bounds on mutual information	useful for Fano's method

- sharp bounds for several parametric/non-parametric problems
- many open problems and issues:
  - benefits of partially local privacy?
  - other models for privacy?
  - privacy for multiple statistical objectives?

Summary

- interesting trade-offs between privacy and statistical utility
- new notion of locally α-private minimax risk
- provided some general bounds and techniques:

bounds on total variation	useful for Le Cam's method
bounds on mutual information	useful for Fano's method

- sharp bounds for several parametric/non-parametric problems
- many open problems and issues:
  - benefits of partially local privacy?
  - other models for privacy?
  - privacy for multiple statistical objectives?

Some papers:

- Duchi, W. & Jordan (2013). Local privacy and statistical minimax rates. http://arxiv.org/abs/1302.3203, February 2013.
- W. (2015). Constrained forms of statistical minimax: Computation, communication, and privacy. *Proceedings of the International Congress of Mathematicians*.