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
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

Classical minimax risk = \( \inf_{\hat{\theta}_n} \sup_{\theta \in \Omega} \mathbb{E}[\mathcal{L}(\hat{\theta}_n, \theta)] \)
Choose estimator to minimize the worst-case risk

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

Two party game:

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

\[
(X_1, \ldots, X_n) \mapsto \widehat{\theta}_n \in \Omega
\]

arbitrary measurable function
From Classical Minimax Risk...

Choose estimator to minimize the worst-case risk

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

Two party game:

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

\[
(X_1, \ldots, X_n) \mapsto \hat{\theta}_n \in \Omega
\]

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?

Abraham Wald
1902–1950
....to Constrained Minimax Risk

Classical framework imposes no constraints on the choice of estimators $\hat{\theta}_n$. 
....to Constrained Minimax Risk

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

- 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 $\hat{\theta}_n$.

- 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?
Each individual $i \in \{1, 2, \ldots, n\}$ has personal data $X_i \sim \mathbb{P}_{\theta^*}$.

Conditional distribution $Q$ between private data $X^n_1$ and public data $Z^n_1$.

Estimator $Z^n_1 \mapsto \hat{\theta}$ of unknown parameter $\theta^*$.
Local privacy at level $\alpha$

<table>
<thead>
<tr>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conditional distribution $Q$ is locally $\alpha$-differentially private if</td>
</tr>
</tbody>
</table>

$$e^{-\alpha} \leq \sup_{z} \frac{Q(z \mid x^n_1)}{Q(z \mid \bar{x}^n_1)} \leq e^{\alpha} \quad \text{for all } x^n_1 \text{ and } \bar{x}^n_1 \text{ such that } d_{\text{HAM}}(x^n_1, \bar{x}^n_1) = 1.$$  

(Dwork et al., 2006)
Add $\alpha$-Laplacian noise (Dwork et al., 2006)

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

$$Z = x + W,$$

where $W$ has density $\propto e^{-\alpha |w|}$

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

$$\sup_{z \in \mathbb{R}} \left| \log \frac{Q(z | x)}{Q(z | x')} \right| = \alpha \left| \sup_{z \in \mathbb{R}} |z - x| - |z - x'| \right| \leq \alpha.$$
Various mechanisms for $\alpha$-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 $\alpha$-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 $\alpha$-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 $\alpha$-privacy and statistical utility?
- Can we identify optimal “mechanisms” for privacy?
Minimax optimality with $\alpha$-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 $X_1^n \mapsto \hat{\theta}(X_1^n)$
- loss function (e.g., squared error, 0-1 error, $\ell_1$-error)

\[
(\hat{\theta}, \theta) \mapsto L(\hat{\theta}, \theta)
\]

quality of $\hat{\theta}$ as estimate of $\theta$
Minimax optimality with \( \alpha \)-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 \( X_1^n \mapsto \hat{\theta}(X_1^n) \)

- loss function (e.g., squared error, 0-1 error, \( \ell_1 \)-error)
  \[
  (\hat{\theta}, \theta) \quad \mapsto \quad \mathcal{L}(\hat{\theta}, \theta)
  \]

  quality of \( \hat{\theta} \) as estimate of \( \theta \)

**Ordinary minimax risk:**

\[
M_n(\mathcal{F}) := \inf_{\hat{\theta}} \sup_{\mathbb{P} \in \mathcal{F}} \mathbb{E} \left[ \mathcal{L}(\hat{\theta}(X_1^n), \theta(\mathbb{P})) \right]
\]

Best estimator \best\ Worst-case distribution
Minimax optimality with $\alpha$-privacy

- family of distributions $\{P \in \mathcal{F}\}$, and functional $P \mapsto \theta(P)$
- samples $X_1^n \equiv \{X_1, \ldots, X_n\} \sim P$ and estimator $X_1^n \mapsto \hat{\theta}(X_1^n)$
- loss function (e.g., squared error, 0-1 error, $\ell_1$-error)
  \[
  (\hat{\theta}, \theta) \quad \mapsto \quad \mathcal{L}(\hat{\theta}, \theta)
  \]
  quality of $\hat{\theta}$ as estimate of $\theta$

Ordinary minimax risk:

\[
\mathcal{M}_n(\mathcal{F}) := \inf_{\hat{\theta}} \sup_{P \in \mathcal{F}} \mathbb{E} \left[ \mathcal{L}(\hat{\theta}(X_1^n), \theta(P)) \right]
\]

Best estimator Worst-case distribution

Minimax risk with $\alpha$-privacy

Estimators now depend on privatized samples $Z_1^n$

\[
\mathcal{M}_n(\alpha; \mathcal{F}) := \inf_{Q \in \mathcal{Q}_\alpha} \inf_{\theta} \sup_{P \in \mathcal{F}} \mathbb{E} \left[ \mathcal{L}(\hat{\theta}(Z_1^n), \theta(P)) \right]
\]

Best $\alpha$-private channel
Vignette A: $\alpha$-private location estimation

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

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

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

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

For $k \geq 2$ and non-private setting, sample mean $\hat{\theta} = \frac{1}{n} \sum_{i=1}^{n} X_i$ achieves rate $1/n$. 
Vignette A: \( \alpha \)-private location estimation

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

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

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

**Theorem**

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

\[
\mathcal{M}_n(\alpha; \mathcal{F}_k) \asymp \min \left\{ 1, \left( \frac{1}{\alpha^2 n} \right)^{k-1/k} \right\}.
\]
Vignette A: $\alpha$-private location estimation

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

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

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

**Theorem**

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

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

**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 $\alpha$-private channel, any pair $\{P_j, j = 1, 2\}$ induces marginals

$$M_j^n(A) := \int Q(A \mid x_1, \ldots, x_n) dP_j^n(x_1, \ldots, x_n) \text{ for } j = 1, 2.$$
Sample size reduction: \( n \mapsto \alpha^2 n \)

Given an \( \alpha \)-private channel, any pair \( \{P_j, j = 1, 2\} \) induces marginals

\[
M^n_j(A) := \int Q(A \mid x_1, \ldots, x_n) dP^n_j(x_1, \ldots, x_n) \quad \text{for} \ j = 1, 2.
\]

Question:
How much “contraction” induced by local \( \alpha \)-privacy?
Sample size reduction: \( n \mapsto \alpha^2 n \)

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

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

**Question:**

How much “contraction” induced by local \( \alpha \)-privacy?

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

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

\[
\frac{1}{n} \left\{ D(\mathbb{M}_1^n \parallel \mathbb{M}_0^n) + D(\mathbb{M}_0^n \parallel \mathbb{M}_1^n) \right\} \sim (e^\alpha - 1)^2 \left\| \mathbb{P}_1 - \mathbb{P}_0 \right\|_{TV}^2
\]

Symmetrized KL divergence

Total variation
Sample size reduction: \( n \mapsto \alpha^2 n \)

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

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

**Question:**

How much “contraction” induced by local \( \alpha \)-privacy?

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

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

\[
\frac{1}{n} \left\{ \frac{D(M^n_1 \parallel M^n_0) + D(M^n_0 \parallel M^n_1)}{\text{Symmetrized KL divergence}} \right\} \preceq (e^\alpha - 1)^2 \quad \text{\|P_1 - P_0\|}^2_{TV} \quad \text{Total variation}
\]

Note that \((e^\alpha - 1)^2 \preceq \alpha^2 \) for \( \alpha \in (0, 1/4] \).
Vignette B: Non-parametric density estimation

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

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

$$M_n(F(\beta)) \asymp \left(\frac{1}{n}\right)^{\frac{2\beta}{2\beta+1}}.$$

(Ibragimov & Hasminskii, 1978; Stone, 1980)
Optimal rates for $\alpha$-private density estimation

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

Consider density estimation based on $\alpha$-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 $\beta > 1/2$:

$$M_n(\alpha; F(\beta)) \asymp \left(\frac{1}{\alpha^2 n}\right)^{\frac{2\beta}{2\beta + 2}}$$
Optimal rates for $\alpha$-private density estimation

Consider density estimation based on $\alpha$-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 $\beta > 1/2$:

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

- can give a simple/explicit scheme that achieves this optimal rate.
Optimal rates for $\alpha$-private density estimation

Consider density estimation based on $\alpha$-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 $\beta > 1/2$:

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

- 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!
Optimal rates for $\alpha$-private density estimation

Consider density estimation based on $\alpha$-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 $\beta > 1/2$:

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

- 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

$Z_i = X_i + W_i, \quad \text{with } W_i \sim \frac{\alpha}{2} e^{-\alpha |w|}$
How to achieve a matching upper bound?

Naive approach: Add Laplacian noise directly to samples

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

Transforms problem into non-parametric deconvolution.
How to achieve a matching upper bound?

Naive approach: Add Laplacian noise directly to samples

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

Transforms problem into non-parametric deconvolution.

Lower bound for this mechanism

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

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

Follows from known lower bounds for deconvolution \(\text{ (Carroll & Hall, 1988)}\)
An optimal mechanism

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

\[
\Phi_i^D(X_i) := \{ \phi_1(X_i), \phi_2(X_i), \ldots, \phi_D(X_i) \}
\]

for dimension \( D \) 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

\[
\Phi^D_i(X_i) := \{\phi_1(X_i), \phi_2(X_i), \ldots, \phi_D(X_i)\}
\]

for dimension \( D \) to be chosen.

2. Privatized \( D \)-dimensional vector:

Hypercube sampling scheme with 

\[
\mathbb{E}[Z_i \mid X_i] = \Phi^D_i(X_i)
\]
An optimal mechanism

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

\[
\Phi^D_1(X_i) := \{\phi_1(X_i), \phi_2(X_i), \ldots, \phi_D(X_i)\}
\]

for dimension \( D \) to be chosen.

2. Privatized \( D \)-dimensional vector:

Hypercube sampling scheme with

\[
\mathbb{E}[Z_i \mid X_i] = \Phi^D_1(X_i)
\]

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

\[
\hat{B}_j = \frac{1}{n} \sum_{i=1}^{n} Z_{ij}, \quad \text{and} \quad \hat{f} = \sum_{j=1}^{D} \hat{B}_j \phi_j
\]
An optimal mechanism

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

\[
\Phi^D_i(X_i) := \{\phi_1(X_i), \phi_2(X_i), \ldots, \phi_D(X_i)\}
\]

for dimension \( D \) to be chosen

2. Privatized \( D \)-dimensional vector:

Hypercube sampling scheme with \( \mathbb{E}[Z_i | X_i] = \Phi^D_1(X_i) \)

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

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

Upper bound

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

\[
\mathbb{E}[\|\hat{f} - f^*\|_2^2] \lesssim \frac{D^2}{n\alpha^2} + \frac{1}{D^{2\beta}}
\]
Hypercube sampling: Optimal privacy mechanism

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

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

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

\[
Z \sim \begin{cases} 
\text{Uni}(\{-C, +C\}^D | \langle Z, \tilde{V} \rangle > 0) & \text{if } T = 1 \\
\text{Uni}(\{-C, +C\}^D | \langle Z, \tilde{V} \rangle \leq 0) & \text{if } T = 0
\end{cases}
\]
Lower bounds via metric entropy

Andrey Kolmogorov
1903–1987
Lower bounds via metric entropy

Packing number

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

$$\rho(f^j, f^k) > 2\delta \quad \text{for all } j \neq k.$$
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 $\hat{f}$ for which $\rho(\hat{f}, f^J) < \delta$ with high probability can be used to decode the index $J$. 
A quantitative data processing inequality

packing index $J \in \{1, 2, \ldots, M\}$
non-private variables $(X \mid J = j) \sim P_j$
mixture distribution $\overline{P} = \frac{1}{M} \sum_{j=1}^{M} P_j$. 
A quantitative data processing inequality

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

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

For any non-interactive $\alpha$-private channel $Q$, we have

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

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 \((n \mapsto \frac{n\alpha^2}{d})\)
   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 \((n \mapsto \frac{n\alpha^2}{d})\)
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 $\alpha$-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 $\alpha$-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 $\alpha$-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: