Local Privacy and Statistical Minimax Rates

John C. Duchi, Michael I. Jordan, Martin J. Wainwright

University of California, Berkeley

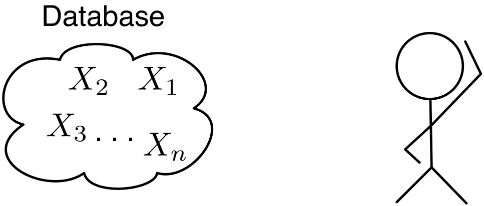
December 2013

Goals for this talk

Bring together some classical concepts of decision theory
and newer concepts of privacy

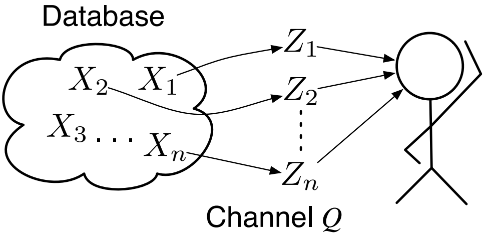
Illustration of problem

Illustration of problem



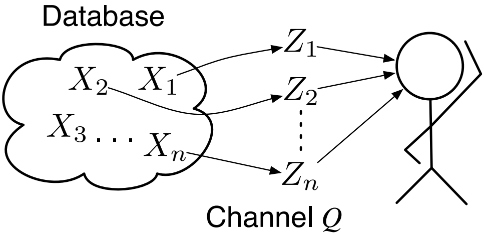
- Have data $X_1, \ldots, X_n$

Illustration of problem



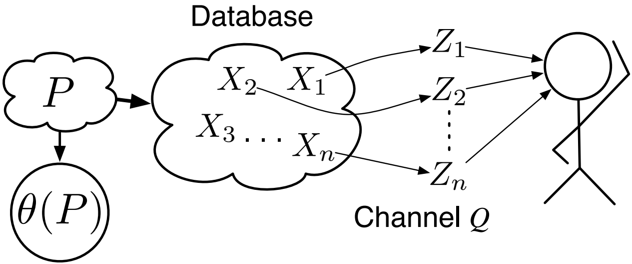
- Have data $X_1, \ldots, X_n$
- Private views $Z_1, \ldots, Z_n$ constructed from X_i

Illustration of problem



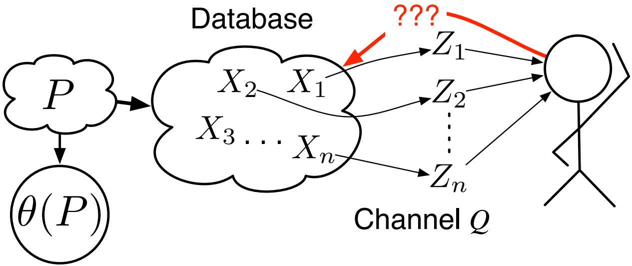
- Have data $X_1, \ldots, X_n$
- Private views $Z_1, \ldots, Z_n$ constructed from X_i
  - Often: goal to get statistics of $\{X_1, \ldots, X_n\}$ (e.g. average salary)

Illustration of problem



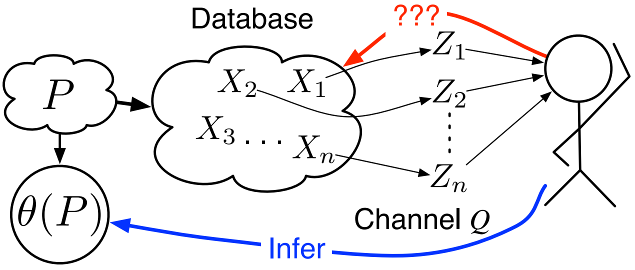
- Have data $X_1, \ldots, X_n$
- Private views $Z_1, \ldots, Z_n$ constructed from X_i
  - Often: goal to get statistics of $\{X_1, \ldots, X_n\}$ (e.g. average salary)
- Distribution P and parameter $\theta(P)$ *generate data*

Illustration of problem



- Have data $X_1, \ldots, X_n$
- Private views $Z_1, \ldots, Z_n$ constructed from X_i
  - Often: goal to get statistics of $\{X_1, \ldots, X_n\}$ (e.g. average salary)
- Distribution P and parameter $\theta(P)$ *generate data*
- Sample $X_1, \ldots, X_n$ from P *not* observed (only Z_i)

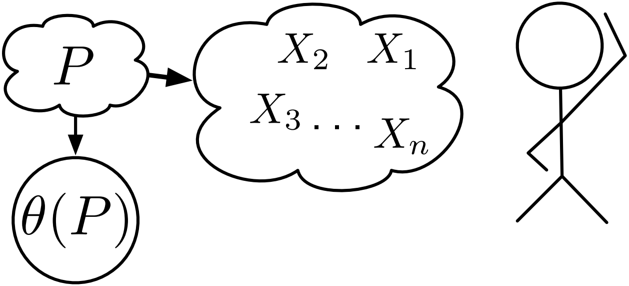
Illustration of problem



- Have data $X_1, \ldots, X_n$
- Private views $Z_1, \ldots, Z_n$ constructed from X_i
  - Often: goal to get statistics of $\{X_1, \ldots, X_n\}$ (e.g. average salary)
- Distribution P and parameter $\theta(P)$ *generate data*
- Sample $X_1, \ldots, X_n$ from P *not* observed (only Z_i)
- Goal: infer population parameter $\theta(P)$ based on $Z_1, \ldots, Z_n$

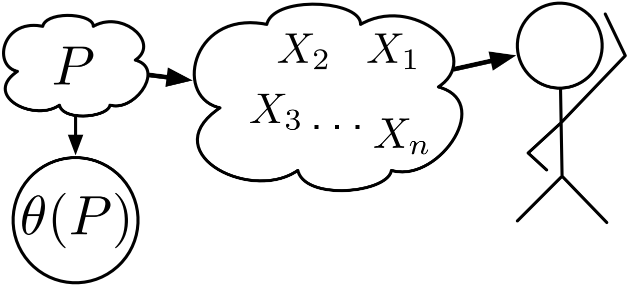
Primer on minimax rates of convergence and statistical inference

Illustration of classical problem



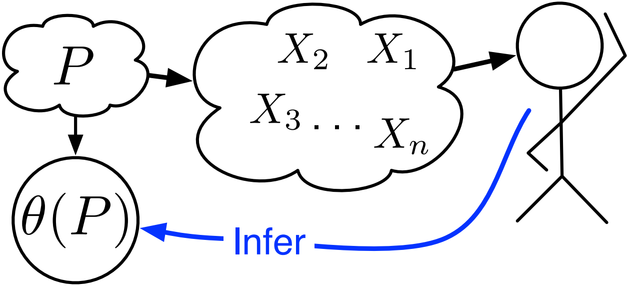
- Have distribution P and parameter $\theta(P)$ of P

Illustration of classical problem



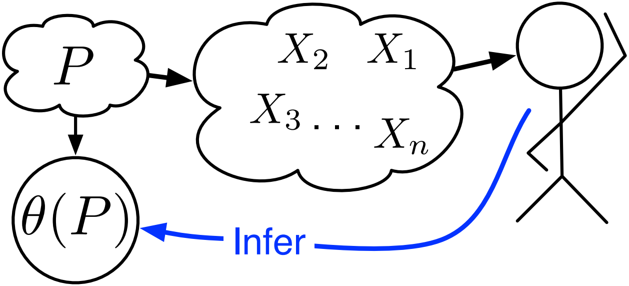
- Have distribution P and parameter $\theta(P)$ of P
- Sample $X_1, \ldots, X_n$ drawn from P and observed

Illustration of classical problem



- Have distribution P and parameter $\theta(P)$ of P
- Sample $X_1, \ldots, X_n$ drawn from P and observed
- Goal: infer population parameter $\theta(P)$

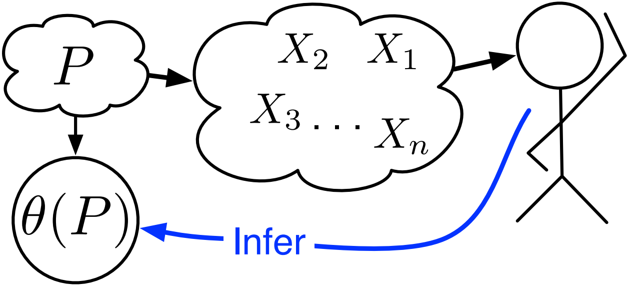
Illustration of classical problem



- Have distribution P and parameter $\theta(P)$ of P
- Sample $X_1, \ldots, X_n$ drawn from P and observed
- Goal: infer population parameter $\theta(P)$

Why? Care about making future predictions

Illustration of classical problem

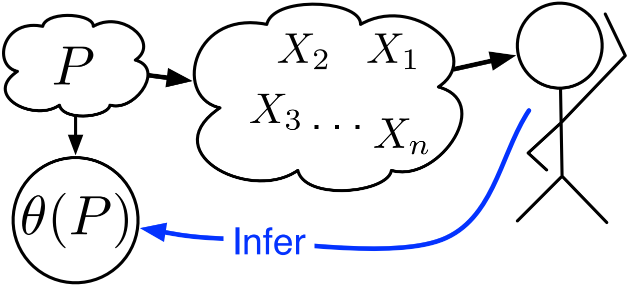


- Have distribution P and parameter $\theta(P)$ of P
- Sample $X_1, \ldots, X_n$ drawn from P and observed
- Goal: infer population parameter $\theta(P)$

Why? Care about making future predictions

- What is likelihood new resident of San Francisco needs food stamps

Illustration of classical problem



- Have distribution P and parameter $\theta(P)$ of P
- Sample $X_1, \ldots, X_n$ drawn from P and observed
- Goal: infer population parameter $\theta(P)$

Why? Care about making future predictions

- What is likelihood new resident of San Francisco needs food stamps
- Biological prediction, web advertising, search, ...

Minimax risk

Central object of study: Minimax risk

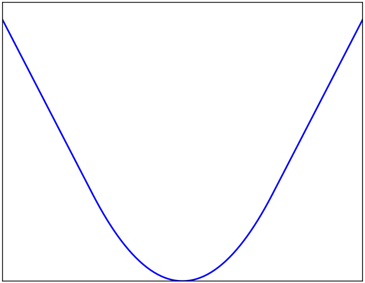
- Parameter $\theta(P)$ of distribution P
  - E.g. mean: $\theta(P) = \mathbb{E}_P[X]$

Minimax risk

Central object of study: Minimax risk

- Parameter $\theta(P)$ of distribution P
  - E.g. mean: $\theta(P) = \mathbb{E}_P[X]$
- Loss ρ that measures error in estimate of $\widehat{\theta}$ for θ: $\rho(\widehat{\theta}, \theta)$
  - E.g. $\rho(\widehat{\theta}, \theta) = \|\widehat{\theta} - \theta\|_2^2$ or more esoteric/robust

$$\rho_u(\widehat{\theta}, \theta) = \begin{cases} \frac{1}{2}\|\widehat{\theta} - \theta\|_2^2 & \text{if } \|\widehat{\theta} - \theta\|_2 \le u \\ u\|\widehat{\theta} - \theta\|_2 - u^2/2 & \text{otherwise} \end{cases}$$

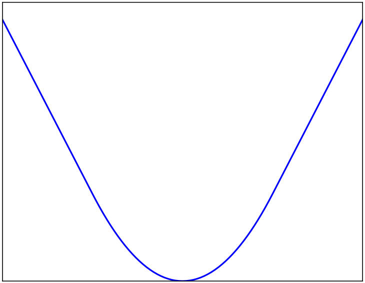


Minimax risk

Central object of study: Minimax risk

- Parameter $\theta(P)$ of distribution P
  - E.g. mean: $\theta(P) = \mathbb{E}_P[X]$
- Loss ρ that measures error in estimate of $\widehat{\theta}$ for θ: $\rho(\widehat{\theta}, \theta)$
  - E.g. $\rho(\widehat{\theta}, \theta) = \|\widehat{\theta} - \theta\|_2^2$ or more esoteric/robust

$$\rho_u(\widehat{\theta}, \theta) = \begin{cases} \frac{1}{2}\|\widehat{\theta} - \theta\|_2^2 & \text{if } \|\widehat{\theta} - \theta\|_2 \le u \\ u\|\widehat{\theta} - \theta\|_2 - u^2/2 & \text{otherwise} \end{cases}$$



- Family of distributions $\mathcal{P}$ that we study
  - E.g. P such that $\mathbb{E}_P[X^2] \le 1$

Minimax risk

Central object of study: Minimax risk

- Parameter $\theta(P)$ of distribution P
- Loss ρ that measures error in estimate of $\widehat{\theta}$ for θ: $\rho(\widehat{\theta}, \theta)$
  - E.g. $\rho(\widehat{\theta}, \theta) = \|\widehat{\theta} - \theta\|_2^2$
- Family of distributions $\mathcal{P}$ that we study

Minimax risk

Central object of study: Minimax risk

- Parameter $\theta(P)$ of distribution P
- Loss ρ that measures error in estimate of $\widehat{\theta}$ for θ: $\rho(\widehat{\theta}, \theta)$
  - E.g. $\rho(\widehat{\theta}, \theta) = \|\widehat{\theta} - \theta\|_2^2$
- Family of distributions $\mathcal{P}$ that we study

Look at expected loss

$$\mathbb{E}_P \left[\rho(\widehat{\theta}(X_1, \ldots, X_n), \theta(P))\right]$$

Minimax risk

Central object of study: Minimax risk

- Parameter $\theta(P)$ of distribution P
- Loss ρ that measures error in estimate of $\widehat{\theta}$ for θ: $\rho(\widehat{\theta}, \theta)$
  - E.g. $\rho(\widehat{\theta}, \theta) = \|\widehat{\theta} - \theta\|_2^2$
- Family of distributions $\mathcal{P}$ that we study

Look at expected loss

$$\sup_{P \in \mathcal{P}} \mathbb{E}_P \left[\rho(\widehat{\theta}(X_1, \ldots, X_n), \theta(P)) \right]$$

- Worst case over distributions $\mathcal{P}$

Minimax risk

Central object of study: Minimax risk

- Parameter $\theta(P)$ of distribution P
- Loss ρ that measures error in estimate of $\widehat{\theta}$ for θ: $\rho(\widehat{\theta}, \theta)$
  - E.g. $\rho(\widehat{\theta}, \theta) = \|\widehat{\theta} - \theta\|_2^2$
- Family of distributions $\mathcal{P}$ that we study

Look at expected loss

$$\inf_{\widehat{\theta}} \sup_{P \in \mathcal{P}} \mathbb{E}_P \left[\rho(\widehat{\theta}(X_1, \ldots, X_n), \theta(P)) \right]$$

- Worst case over distributions $\mathcal{P}$
- Best case over all estimators $\widehat{\theta} : \mathcal{Z}^n \to \Theta$

Minimax risk

Central object of study: Minimax risk

- Parameter $\theta(P)$ of distribution P
- Loss ρ that measures error in estimate of $\widehat{\theta}$ for θ: $\rho(\widehat{\theta}, \theta)$
  - E.g. $\rho(\widehat{\theta}, \theta) = \|\widehat{\theta} - \theta\|_2^2$
- Family of distributions $\mathcal{P}$ that we study

Look at expected loss

$$\mathfrak{M}_n(\theta(\mathcal{P}), \rho) := \underbrace{\inf_{\widehat{\theta}} \sup_{P \in \mathcal{P}} \mathbb{E}_P \left[\rho(\widehat{\theta}(X_1, \ldots, X_n), \theta(P)) \right]}_{\text{Classical minimax risk}}$$

- Worst case over distributions $\mathcal{P}$
- Best case over all estimators $\widehat{\theta} : \mathcal{Z}^n \to \Theta$

Minimax risk

Central object of study: Minimax risk

- Parameter $\theta(P)$ of distribution P
- Loss ρ that measures error in estimate of $\widehat{\theta}$ for θ: $\rho(\widehat{\theta}, \theta)$
  - E.g. $\rho(\widehat{\theta}, \theta) = \|\widehat{\theta} - \theta\|_2^2$
- Family of distributions $\mathcal{P}$ that we study

Look at expected loss

$$\mathfrak{M}_n(\theta(\mathcal{P}), \rho) := \underbrace{\inf_{\widehat{\theta}} \sup_{P \in \mathcal{P}} \mathbb{E}_P \left[\rho(\widehat{\theta}(X_1, \ldots, X_n), \theta(P)) \right]}_{\text{Classical minimax risk}}$$

- Worst case over distributions $\mathcal{P}$
- Best case over all estimators $\widehat{\theta} : \mathcal{Z}^n \to \Theta$

To study: rate of $\mathfrak{M}_n(\theta(\mathcal{P}), \rho) \to 0$ as n grows

Proving minimax bounds

This talk:

Proving minimax bounds

This talk:

- Upper bounds will be ad-hoc

Proving minimax bounds

This talk:

- Upper bounds will be ad-hoc
- Lower bounds will be information theoretic [Hasminskii 78, Birge 83, Ibragimov and Hasminskii 81, Yang and Barron 99, Yu97]

Proving minimax bounds

This talk:

- Upper bounds will be ad-hoc
- Lower bounds will be information theoretic [Hasminskii 78, Birge 83, Ibragimov and Hasminskii 81, Yang and Barron 99, Yu97]
- **NB:** Many known information-theoretic upper bounds [Barron, Birge, Kivinen, ...]

Proving minimax lower bounds

Step 1: Reduce from estimation to testing

Proving minimax lower bounds

Step 1: Reduce from estimation to testing

Step 2: Canonical testing problem

Proving minimax lower bounds

Step 1: Reduce from estimation to testing

Step 2: Canonical testing problem

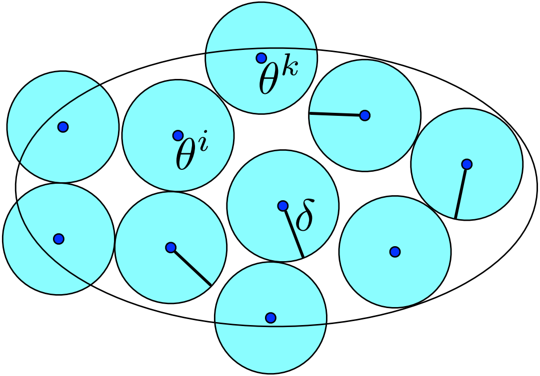
Step 3: Classical bounds on error probabilities

Proving minimax lower bounds

Step 1: Reduce from estimation to testing
Step 2: Canonical testing problem
Step 3: Classical bounds on error probabilities

Proving minimax lower bounds

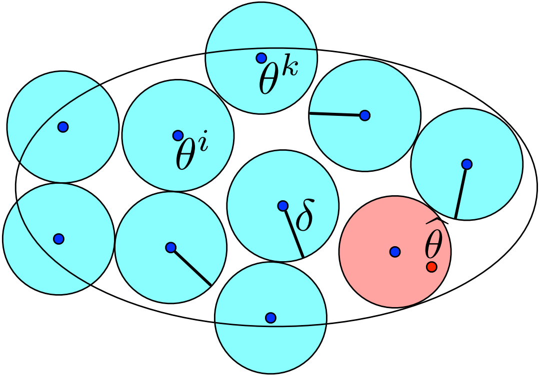
Step 1: Reduce from estimation to testing
Step 2: Canonical testing problem
Step 3: Classical bounds on error probabilities



- Have 2δ-packing of Θ, i.e. $\{\theta^1, \theta^2, \ldots, \theta^K\}$

Proving minimax lower bounds

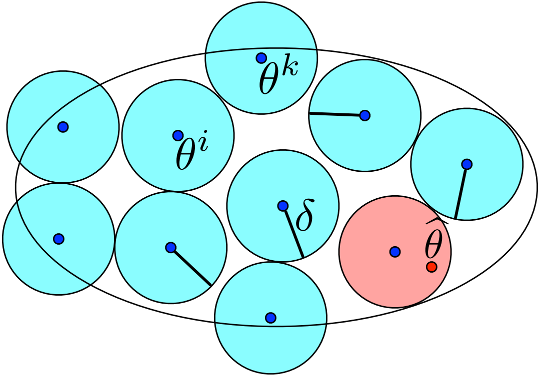
Step 1: Reduce from estimation to testing
Step 2: Canonical testing problem
Step 3: Classical bounds on error probabilities



- Have 2δ-packing of Θ, i.e. $\{\theta^1, \theta^2, \ldots, \theta^K\}$
- Estimator $\widehat{\theta}$

Proving minimax lower bounds

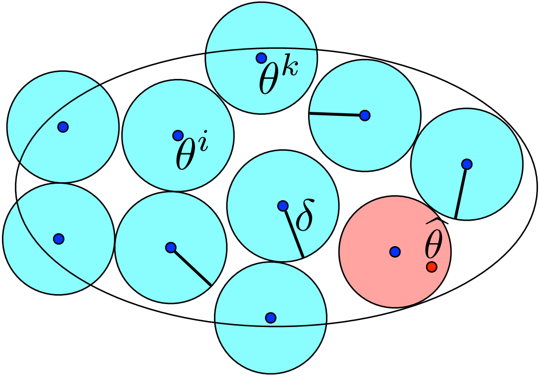
Step 1: Reduce from estimation to testing
Step 2: Canonical testing problem
Step 3: Classical bounds on error probabilities



- Have 2δ-packing of Θ, i.e. $\{\theta^1, \theta^2, \ldots, \theta^K\}$
- Estimator $\widehat{\theta}$
- At most *one* index close to $\widehat{\theta}$

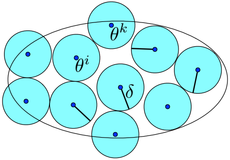
Proving minimax lower bounds

Step 1: Reduce from estimation to testing
Step 2: Canonical testing problem
Step 3: Classical bounds on error probabilities



- Have 2δ-packing of Θ, i.e. $\{\theta^1, \theta^2, \ldots, \theta^K\}$
- Estimator $\widehat{\theta}$
- At most *one* index close to $\widehat{\theta}$
- Can *test* index $i \in [K]$

Proving minimax lower bounds

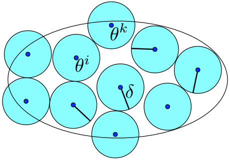


Step 1: Reduce from estimation to testing

Step 2: Canonical testing problem

Step 3: Classical bounds on error probabilities

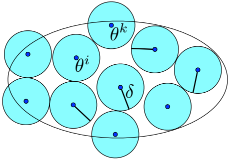
Proving minimax lower bounds



Step 1: Reduce from estimation to testing
Step 2: Canonical testing problem
Step 3: Classical bounds on error probabilities

- Nature chooses random index $V \in [K]$

Proving minimax lower bounds



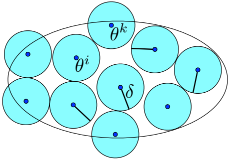
Step 1: Reduce from estimation to testing

Step 2: Canonical testing problem

Step 3: Classical bounds on error probabilities

- Nature chooses random index $V \in [K]$
- Conditional on $V = v$, sample $X_1, \ldots, X_n$ i.i.d. from P_v

Proving minimax lower bounds



Step 1: Reduce from estimation to testing
Step 2: Canonical testing problem
Step 3: Classical bounds on error probabilities

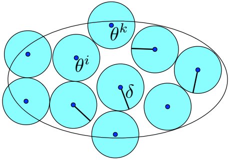
- Nature chooses random index $V \in [K]$
- Conditional on $V = v$, sample $X_1, \ldots, X_n$ i.i.d. from P_v
- Lower bound minimax error:

$$\sup_{P \in \mathcal{P}} \mathbb{E}_P \left[\rho(\widehat{\theta}, \theta(P)) \right] \geq \frac{1}{K} \sum_{v=1}^{K} \mathbb{E}_v \left[\rho(\widehat{\theta}, \theta_v) \right]$$

$$\geq \frac{1}{K} \sum_{v=1}^{K} \rho(\delta) P_v(\rho(\widehat{\theta}, \theta_v) \geq 2\delta)$$

Proving minimax lower bounds

Step 1: Reduce from estimation to testing
Step 2: Canonical testing problem
Step 3: Classical bounds on error probabilities



- Nature chooses random index $V \in [K]$
- Conditional on $V = v$, sample $X_1, \ldots, X_n$ i.i.d. from P_v
- Lower bound minimax error:

$$\sup_{P \in \mathcal{P}} \mathbb{E}_P \left[\rho(\widehat{\theta}, \theta(P)) \right] \geq \frac{1}{K} \sum_{v=1}^{K} \mathbb{E}_v \left[\rho(\widehat{\theta}, \theta_v) \right]$$

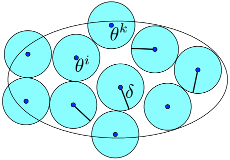
$$\geq \frac{1}{K} \sum_{v=1}^{K} \rho(\delta) P_v(\rho(\widehat{\theta}, \theta_v) \geq 2\delta)$$

- Final canonical testing problem:

$$\mathfrak{M}_n(\theta(\mathcal{P}), \rho) \geq \rho(\delta) \min_{\widehat{v}} \mathbb{P}(\widehat{v}(X_1, \ldots, X_n) \neq V).$$

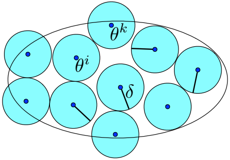
Proving minimax lower bounds

Step 1: Reduce from estimation to testing
Step 2: Canonical testing problem
Step 3: Classical bounds on error probabilities



Proving minimax lower bounds

Step 1: Reduce from estimation to testing
Step 2: Canonical testing problem
Step 3: Classical bounds on error probabilities

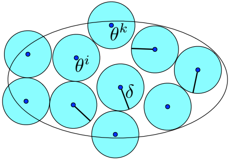


Le Cam's method

$$P_0(\widehat{v} \neq 0) + P_1(\widehat{v} \neq 1) \geq 1 - \|P_0 - P_1\|_{\mathrm{TV}}$$

Proving minimax lower bounds

Step 1: Reduce from estimation to testing
Step 2: Canonical testing problem
Step 3: Classical bounds on error probabilities



Le Cam's method

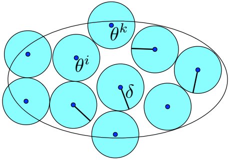
$$P_0(\widehat{v} \neq 0) + P_1(\widehat{v} \neq 1) \geq 1 - \|P_0 - P_1\|_{\mathrm{TV}}$$

Fano's inequality (for $K > 2$)

$$\mathbb{P}(\widehat{v} \neq V) \geq 1 - \frac{I(X_1, \ldots, X_n; V) + \log 2}{\log K}$$

Proving minimax lower bounds

Step 1: Reduce from estimation to testing
Step 2: Canonical testing problem
Step 3: Classical bounds on error probabilities



Le Cam's method

$$P_0(\widehat{v} \neq 0) + P_1(\widehat{v} \neq 1) \geq 1 - \|P_0 - P_1\|_{\mathrm{TV}}$$

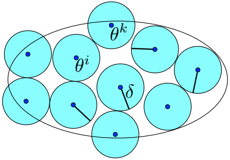
Fano's inequality (for $K > 2$)

$$\mathbb{P}(\widehat{v} \neq V) \geq 1 - \frac{I(X_1, \ldots, X_n; V) + \log 2}{\log K}$$

Summarizing

Proving minimax lower bounds

Step 1: Reduce from estimation to testing
Step 2: Canonical testing problem
Step 3: Classical bounds on error probabilities



Fano's inequality (for $K > 2$)

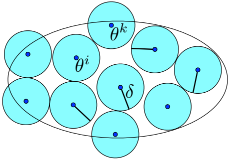
Le Cam's method

$$P_0(\widehat{v} \neq 0) + P_1(\widehat{v} \neq 1) \geq 1 - \|P_0 - P_1\|_{\text{TV}}$$

$$\mathbb{P}(\widehat{v} \neq V) \geq 1 - \frac{I(X_1, \ldots, X_n; V) + \log 2}{\log K}$$

Summarizing

Proving minimax lower bounds



Step 1: Reduce from estimation to testing
Step 2: Canonical testing problem
Step 3: Classical bounds on error probabilities

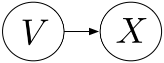
Le Cam's method

$$P_0(\widehat{v} \neq 0) + P_1(\widehat{v} \neq 1) \geq 1 - \|P_0 - P_1\|_{\mathrm{TV}}$$

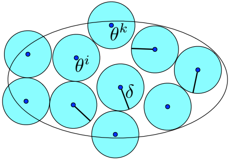
Fano's inequality (for $K > 2$)

$$\mathbb{P}(\widehat{v} \neq V) \geq 1 - \frac{I(X_1, \ldots, X_n; V) + \log 2}{\log K}$$

Summarizing



Proving minimax lower bounds



Step 1: Reduce from estimation to testing
Step 2: Canonical testing problem
Step 3: Classical bounds on error probabilities

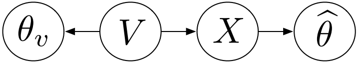
Le Cam's method

$$P_0(\widehat{v} \neq 0) + P_1(\widehat{v} \neq 1) \geq 1 - \|P_0 - P_1\|_{\mathrm{TV}}$$

Fano's inequality (for $K > 2$)

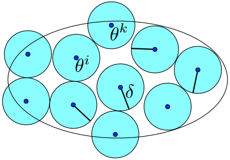
$$\mathbb{P}(\widehat{v} \neq V) \geq 1 - \frac{I(X_1, \ldots, X_n; V) + \log 2}{\log K}$$

Summarizing



Proving minimax lower bounds

Step 1: Reduce from estimation to testing
Step 2: Canonical testing problem
Step 3: Classical bounds on error probabilities



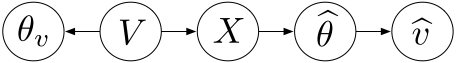
Le Cam's method

$$P_0(\widehat{v} \neq 0) + P_1(\widehat{v} \neq 1) \geq 1 - \|P_0 - P_1\|_{\mathrm{TV}}$$

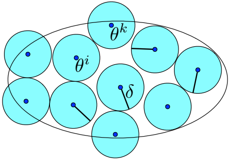
Fano's inequality (for $K > 2$)

$$\mathbb{P}(\widehat{v} \neq V) \geq 1 - \frac{I(X_1, \ldots, X_n; V) + \log 2}{\log K}$$

Summarizing



Proving minimax lower bounds



Step 1: Reduce from estimation to testing
Step 2: Canonical testing problem
Step 3: Classical bounds on error probabilities

Le Cam's method

$$P_0(\widehat{v} \neq 0) + P_1(\widehat{v} \neq 1) \geq 1 - \|P_0 - P_1\|_{\mathrm{TV}}$$

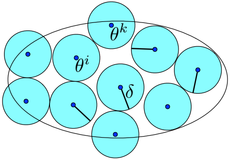
Fano's inequality (for $K > 2$)

$$\mathbb{P}(\widehat{v} \neq V) \geq 1 - \frac{I(X_1, \ldots, X_n; V) + \log 2}{\log K}$$

Summarizing

Proving minimax lower bounds

Step 1: Reduce from estimation to testing
Step 2: Canonical testing problem
Step 3: Classical bounds on error probabilities



Le Cam's method

$$P_0(\widehat{v} \neq 0) + P_1(\widehat{v} \neq 1) \geq 1 - \|P_0 - P_1\|_{\mathrm{TV}}$$

Fano's inequality (for $K > 2$)

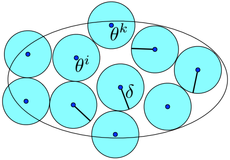
$$\mathbb{P}(\widehat{v} \neq V) \geq 1 - \frac{I(X_1, \ldots, X_n; V) + \log 2}{\log K}$$

Summarizing

$$\mathfrak{M}_n(\theta(\mathcal{P}), \rho) \geq \rho(\delta) \min_{\widehat{v}} \mathbb{P}\left(\widehat{v}(X_1, \ldots, X_n) \neq V\right)$$

Proving minimax lower bounds

Step 1: Reduce from estimation to testing
Step 2: Canonical testing problem
Step 3: Classical bounds on error probabilities

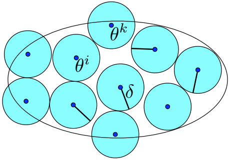


Summarizing

$$\mathfrak{M}_n(\theta(\mathcal{P}), \rho) \geq \rho(\delta) \min_{\widehat{v}} \mathbb{P}\left(\widehat{v}(X_1, \ldots, X_n) \neq V\right)$$

Proving minimax lower bounds

Step 1: Reduce from estimation to testing
Step 2: Canonical testing problem
Step 3: Classical bounds on error probabilities



Summarizing

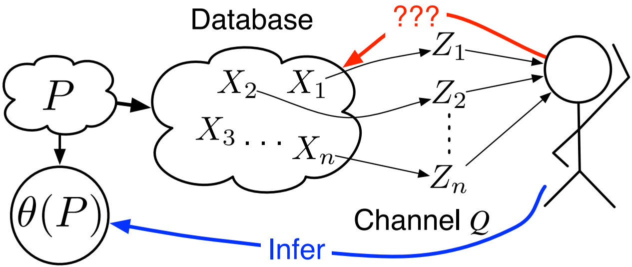
$$\mathfrak{M}_n(\theta(\mathcal{P}), \rho) \geq \rho(\delta) \min_{\widehat{v}} \mathbb{P}\left(\widehat{v}(X_1, \ldots, X_n) \neq V\right)$$

Key idea: Control information-theoretic divergences

$$\|P_0 - P_1\|_{\mathrm{TV}} \quad \text{or} \quad I(X_1, \ldots, X_n; V)$$

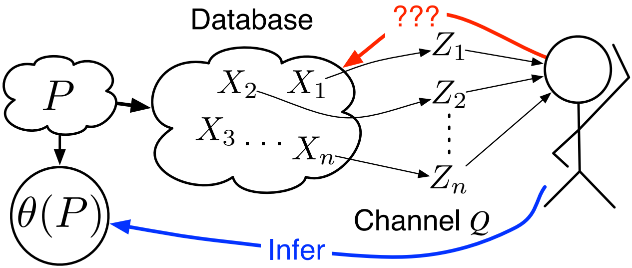
to attain minimax rate

Inference under privacy constraints



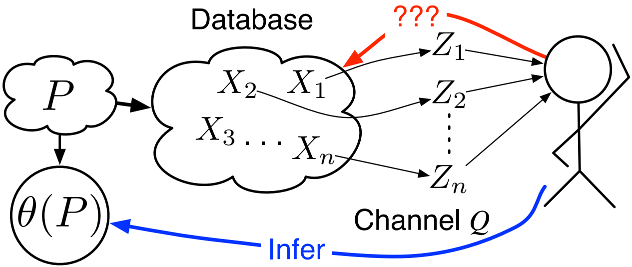
- Have distribution P and parameter $\theta(P)$
- Sample $X_1, \ldots, X_n$ drawn from P and *not* observed
- Private views $Z_1, \ldots, Z_n$ constructed from X_i

Inference under privacy constraints



- Have distribution P and parameter $\theta(P)$
- Sample $X_1, \ldots, X_n$ drawn from P and *not* observed
- Private views $Z_1, \ldots, Z_n$ constructed from X_i
- Goal: infer population parameter $\theta(P)$

Inference under privacy constraints



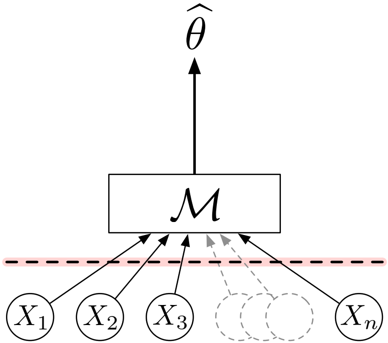
- Have distribution P and parameter $\theta(P)$
- Sample $X_1, \ldots, X_n$ drawn from P and *not* observed
- Private views $Z_1, \ldots, Z_n$ constructed from X_i
- Goal: infer population parameter $\theta(P)$ based on $X_1, \ldots, X_n$

Model of privacy

Local Privacy: Don't trust collector of data (Evfimievski et al. 2003, Warner 1965)

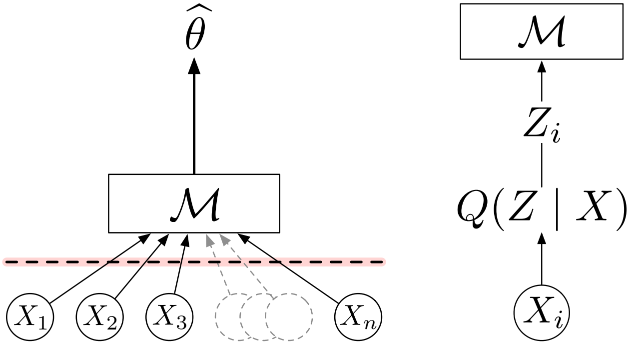
Model of privacy

Local Privacy: Don't trust collector of data (Evfimievski et al. 2003, Warner 1965)



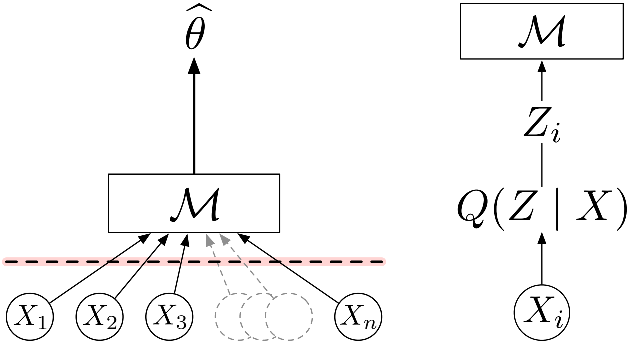
Model of privacy

Local Privacy: Don't trust collector of data (Evfimievski et al. 2003, Warner 1965)



Model of privacy

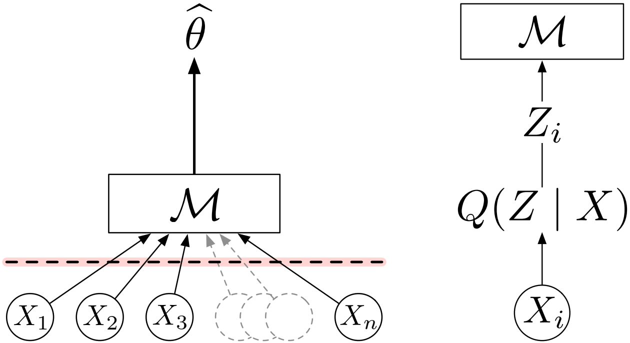
Local Privacy: Don't trust collector of data (Evfimievski et al. 2003, Warner 1965)



- Individuals $i \in \{1, \ldots, n\}$ have personal data $X_i \sim P_\theta$

Model of privacy

Local Privacy: Don't trust collector of data (Evfimievski et al. 2003, Warner 1965)



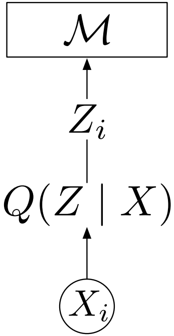
- Individuals $i \in \{1, \ldots, n\}$ have personal data $X_i \sim P_\theta$
- Estimator $Z_1^n \mapsto \widehat{\theta}(Z_{1:n})$

Differential privacy

Definition: The channel Q is α*-differentially private* if

$$\max_{z,x,x'} \frac{Q(Z = z \mid x)}{Q(Z = z \mid x')} \le e^{\alpha}.$$

[Dwork, McSherry, Nissim, Smith 2006]

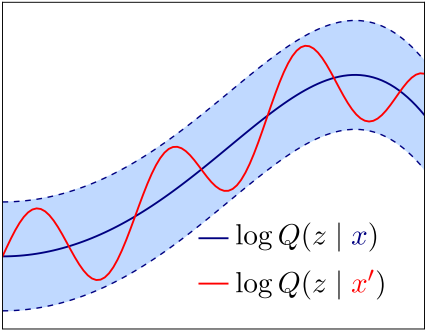


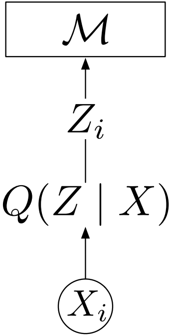
Differential privacy

Definition: The channel Q is α-*differentially private* if

$$\max_{z,x,x'} \frac{Q(Z = z \mid x)}{Q(Z = z \mid x')} \leq e^{\alpha}.$$

[Dwork, McSherry, Nissim, Smith 2006]





Differential privacy

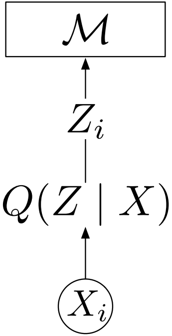
Definition: The channel Q is *α-differentially private* if

$$\max_{z,x,x'} \frac{Q(Z = z \mid x)}{Q(Z = z \mid x')} \leq e^{\alpha}.$$

[Dwork, McSherry, Nissim, Smith 2006]

What does this mean?

- Given Z, *cannot tell* what x gave Z



Differential privacy

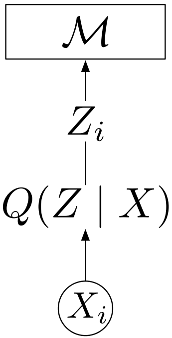
Definition: The channel Q is *α-differentially private* if

$$\max_{z,x,x'} \frac{Q(Z = z \mid x)}{Q(Z = z \mid x')} \leq e^{\alpha}.$$

[Dwork, McSherry, Nissim, Smith 2006]

What does this mean?

- Given Z, *cannot tell* what x gave Z
- Testing argument: based on Z, adversary must distinguish between x and x':



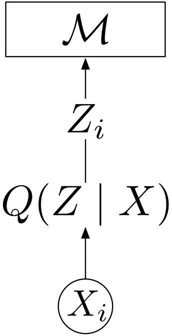
[Wasserman and Zhou 2011]

Differential privacy

Definition: The channel Q is α-*differentially private* if

$$\max_{z,x,x'} \frac{Q(Z = z \mid x)}{Q(Z = z \mid x')} \le e^{\alpha}.$$

[Dwork, McSherry, Nissim, Smith 2006]



What does this mean?

- Given Z, *cannot tell* what x gave Z
- Testing argument: based on Z, adversary must distinguish between x and x':

$$\text{FNR} + \text{FPR} \ge \frac{2}{1 + e^{\alpha}}$$

[Wasserman and Zhou 2011]

Minimax risk

Central object of study: Minimax risk

- Parameter $\theta(P)$ of distribution P
- Loss ρ that measures error in estimate of $\widehat{\theta}$ for θ: $\rho(\widehat{\theta}, \theta)$
  - E.g. $\rho(\widehat{\theta}, \theta) = \|\widehat{\theta} - \theta\|_2^2$
- Family of distributions $\mathcal{P}$ that we study

Minimax risk

Central object of study: Minimax risk

- Parameter $\theta(P)$ of distribution P
- Loss ρ that measures error in estimate of $\widehat{\theta}$ for θ: $\rho(\widehat{\theta}, \theta)$
  - E.g. $\rho(\widehat{\theta}, \theta) = \|\widehat{\theta} - \theta\|_2^2$
- Family of distributions $\mathcal{P}$ that we study

Look at expected loss

$$\mathbb{E}_P \left[\rho(\widehat{\theta}(Z_1, \ldots, Z_n), \theta(P)) \right]$$

Minimax risk

Central object of study: Minimax risk

- Parameter $\theta(P)$ of distribution P
- Loss ρ that measures error in estimate of $\widehat{\theta}$ for θ: $\rho(\widehat{\theta}, \theta)$
  - E.g. $\rho(\widehat{\theta}, \theta) = \|\widehat{\theta} - \theta\|_2^2$
- Family of distributions $\mathcal{P}$ that we study

Look at expected loss

$$\sup_{P \in \mathcal{P}} \mathbb{E}_P \left[\rho(\widehat{\theta}(Z_1, \ldots, Z_n), \theta(P)) \right]$$

- Worst case over distributions $\mathcal{P}$

Minimax risk

Central object of study: Minimax risk

- Parameter $\theta(P)$ of distribution P
- Loss ρ that measures error in estimate of $\widehat{\theta}$ for θ: $\rho(\widehat{\theta}, \theta)$
  - E.g. $\rho(\widehat{\theta}, \theta) = \|\widehat{\theta} - \theta\|_2^2$
- Family of distributions $\mathcal{P}$ that we study

Look at expected loss

$$\inf_{\widehat{\theta}} \sup_{P \in \mathcal{P}} \mathbb{E}_P \left[\rho(\widehat{\theta}(Z_1, \ldots, Z_n), \theta(P)) \right]$$

- Worst case over distributions $\mathcal{P}$
- Best case over all estimators $\widehat{\theta} : \mathcal{Z}^n \to \Theta$

Minimax risk

Central object of study: Minimax risk

- Parameter $\theta(P)$ of distribution P
- Loss ρ that measures error in estimate of $\widehat{\theta}$ for θ: $\rho(\widehat{\theta}, \theta)$
  - E.g. $\rho(\widehat{\theta}, \theta) = \|\widehat{\theta} - \theta\|_2^2$
- Family of distributions $\mathcal{P}$ that we study

Look at expected loss

$$\underbrace{\inf_{\widehat{\theta}} \sup_{P \in \mathcal{P}} \mathbb{E}_P \left[\rho(\widehat{\theta}(Z_1, \ldots, Z_n), \theta(P)) \right]}_{\text{Classical minimax risk}}$$

- Worst case over distributions $\mathcal{P}$
- Best case over all estimators $\widehat{\theta} : \mathcal{Z}^n \to \Theta$

Minimax risk

Central object of study: Minimax risk

- Parameter $\theta(P)$ of distribution P
- Loss ρ that measures error in estimate of $\widehat{\theta}$ for θ: $\rho(\widehat{\theta}, \theta)$
  - E.g. $\rho(\widehat{\theta}, \theta) = \|\widehat{\theta} - \theta\|_2^2$
- Family of distributions $\mathcal{P}$ that we study

Look at expected loss

$$\inf_{Q \in \mathcal{Q}_\alpha} \inf_{\widehat{\theta}} \sup_{P \in \mathcal{P}} \mathbb{E}_{P,Q} \left[\rho(\widehat{\theta}(Z_1, \ldots, Z_n), \theta(P)) \right]$$

- Worst case over distributions $\mathcal{P}$
- Best case over all estimators $\widehat{\theta} : \mathcal{Z}^n \to \Theta$
- Best case over all α-private channels $Q \in \mathcal{Q}_\alpha$ from X to Z

Minimax risk

Central object of study: Minimax risk

- Parameter $\theta(P)$ of distribution P
- Loss ρ that measures error in estimate of $\widehat{\theta}$ for θ: $\rho(\widehat{\theta}, \theta)$
  - E.g. $\rho(\widehat{\theta}, \theta) = \|\widehat{\theta} - \theta\|_2^2$
- Family of distributions $\mathcal{P}$ that we study

Look at expected loss

$$\underbrace{\mathfrak{M}_n(\theta(\mathcal{P}), \rho, \alpha) := \inf_{Q \in \mathcal{Q}_\alpha} \inf_{\widehat{\theta}} \sup_{P \in \mathcal{P}} \mathbb{E}_{P,Q} \left[\rho(\widehat{\theta}(Z_1, \ldots, Z_n), \theta(P)) \right]}_{\text{Private minimax risk}}$$

- Worst case over distributions $\mathcal{P}$
- Best case over all estimators $\widehat{\theta} : \mathcal{Z}^n \to \Theta$
- Best case over all α-private channels $Q \in \mathcal{Q}_\alpha$ from X to Z

Goal for the rest of the talk

How does the minimax risk

$$\mathfrak{M}_n(\theta(\mathcal{P}), \rho, \alpha) := \inf_{Q \in \mathcal{Q}_\alpha} \inf_{\widehat{\theta}} \sup_{P \in \mathcal{P}} \mathbb{E}_{P,Q} \left[\rho(\widehat{\theta}(Z_1, \ldots, Z_n), \theta(P)) \right]$$

change with privacy parameter α and number of samples n?

Goal for the rest of the talk

How does the minimax risk

$$\mathfrak{M}_n(\theta(\mathcal{P}), \rho, \alpha) := \inf_{Q \in \mathcal{Q}_\alpha} \inf_{\widehat{\theta}} \sup_{P \in \mathcal{P}} \mathbb{E}_{P,Q} \left[\rho(\widehat{\theta}(Z_1, \ldots, Z_n), \theta(P)) \right]$$

change with privacy parameter α and number of samples n?

Many related results

Goal for the rest of the talk

How does the minimax risk

$$\mathfrak{M}_n(\theta(\mathcal{P}), \rho, \alpha) := \inf_{Q \in \mathcal{Q}_\alpha} \inf_{\widehat{\theta}} \sup_{P \in \mathcal{P}} \mathbb{E}_{P,Q} \left[\rho(\widehat{\theta}(Z_1, \ldots, Z_n), \theta(P)) \right]$$

change with privacy parameter α and number of samples n?

Many related results

- Non-population lower bounds [Hardt and Talwar 10, Nikkolov, Talwar, Zhang 13; Hall, Rinaldo, Wasserman 11, Chaudhuri, Monteleoni, Sarwate 12]

Goal for the rest of the talk

How does the minimax risk

$$\mathfrak{M}_n(\theta(\mathcal{P}), \rho, \alpha) := \inf_{Q \in \mathcal{Q}_\alpha} \inf_{\widehat{\theta}} \sup_{P \in \mathcal{P}} \mathbb{E}_{P,Q}\left[\rho(\widehat{\theta}(Z_1, \ldots, Z_n), \theta(P))\right]$$

change with privacy parameter α and number of samples n?

Many related results

- Non-population lower bounds [Hardt and Talwar 10, Nikkolov, Talwar, Zhang 13; Hall, Rinaldo, Wasserman 11, Chaudhuri, Monteleoni, Sarwate 12]
- Related population bounds:

Goal for the rest of the talk

How does the minimax risk

$$\mathfrak{M}_n(\theta(\mathcal{P}), \rho, \alpha) := \inf_{Q \in \mathcal{Q}_\alpha} \inf_{\widehat{\theta}} \sup_{P \in \mathcal{P}} \mathbb{E}_{P,Q} \left[\rho(\widehat{\theta}(Z_1, \ldots, Z_n), \theta(P)) \right]$$

change with privacy parameter α and number of samples n?

Many related results

- Non-population lower bounds [Hardt and Talwar 10, Nikkolov, Talwar, Zhang 13; Hall, Rinaldo, Wasserman 11, Chaudhuri, Monteleoni, Sarwate 12]
- Related population bounds:
  - Two point hypotheses [Chaudhuri & Hsu 12, Beimel, Nissim, Omri 08] (e.g. for 1-dimensional bias estimation, get $1/(n\alpha)^2$ error)

Goal for the rest of the talk

How does the minimax risk

$$\mathfrak{M}_n(\theta(\mathcal{P}), \rho, \alpha) := \inf_{Q \in \mathcal{Q}_\alpha} \inf_{\widehat{\theta}} \sup_{P \in \mathcal{P}} \mathbb{E}_{P,Q}\left[\rho(\widehat{\theta}(Z_1, \ldots, Z_n), \theta(P))\right]$$

change with privacy parameter α and number of samples n?

Many related results

- Non-population lower bounds [Hardt and Talwar 10, Nikkolov, Talwar, Zhang 13; Hall, Rinaldo, Wasserman 11, Chaudhuri, Monteleoni, Sarwate 12]
- Related population bounds:
  - Two point hypotheses [Chaudhuri & Hsu 12, Beimel, Nissim, Omri 08] (e.g. for 1-dimensional bias estimation, get $1/(n\alpha)^2$ error)
  - PAC learning results [Beimel, Brenner, Kasiviswanathan, Nissim 13]

Examples

- Mean estimation
- Fixed-design regression
- Convex risk minimization (i.e. online learning)
- Multinomial (probability) estimation
- Nonparametric density estimation

Examples

- Mean estimation
- Fixed-design regression
- Convex risk minimization (i.e. online learning)
- Multinomial (probability) estimation
- Nonparametric density estimation

Example 1: Mean estimation

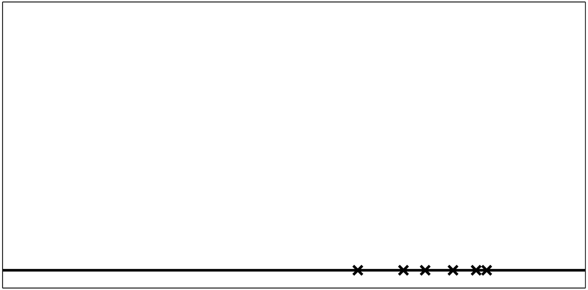
Problem: Estimate mean of distributions P with $k \geq 2$nd moment:

$$\theta(P) := \mathbb{E}_P[X], \quad \mathbb{E}_P[|X|^k] \leq 1.$$

Example 1: Mean estimation

Problem: Estimate mean of distributions P with $k \geq 2$nd moment:

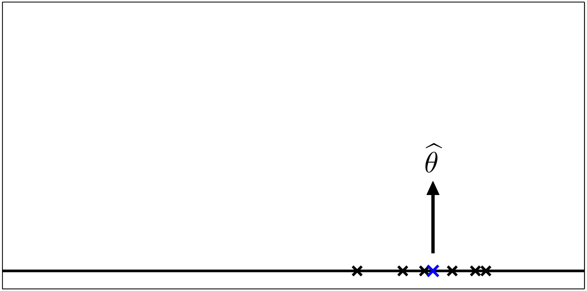
$$\theta(P) := \mathbb{E}_P[X], \quad \mathbb{E}_P[|X|^k] \leq 1.$$



Example 1: Mean estimation

Problem: Estimate mean of distributions P with $k \geq 2$nd moment:

$$\theta(P) := \mathbb{E}_P[X], \quad \mathbb{E}_P[|X|^k] \leq 1.$$



Example 1: Mean estimation

Problem: Estimate mean of distributions P with $k \geq 2$nd moment:

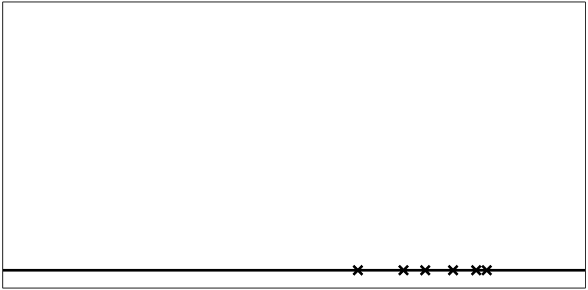
$$\theta(P) := \mathbb{E}_P[X], \quad \mathbb{E}_P[|X|^k] \leq 1.$$

$$\underbrace{\widehat{\theta} = \frac{1}{n}\sum_{i=1}^{n} X_i}_{\text{Standard estimator}} \qquad \mathbb{E}\left[\left(\widehat{\theta} - \mathbb{E}[X]\right)^2\right] \leq \frac{1}{n}.$$

Example 1: Mean estimation

Problem: Estimate mean of distributions P with $k \geq 2$nd moment:

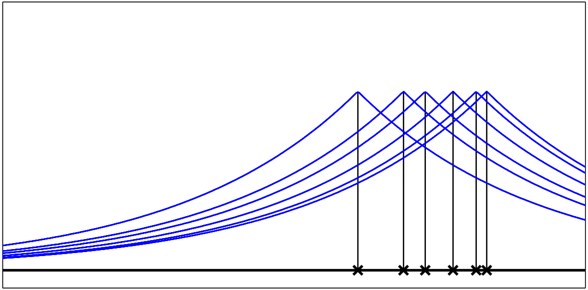
$$\theta(P) := \mathbb{E}_P[X], \quad \mathbb{E}_P[|X|^k] \leq 1.$$



Example 1: Mean estimation

Problem: Estimate mean of distributions P with $k \geq 2$nd moment:

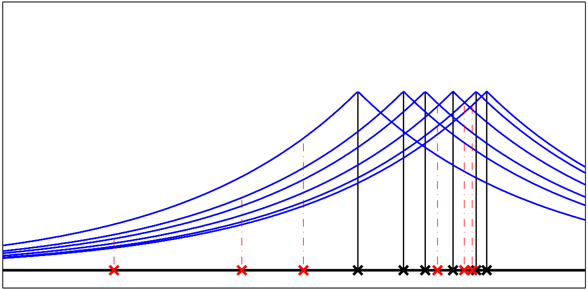
$$\theta(P) := \mathbb{E}_P[X], \quad \mathbb{E}_P[|X|^k] \leq 1.$$



Example 1: Mean estimation

Problem: Estimate mean of distributions P with $k \geq 2$nd moment:

$$\theta(P) := \mathbb{E}_P[X], \quad \mathbb{E}_P[|X|^k] \leq 1.$$



Example 1: Mean estimation

Problem: Estimate mean of distributions P with $k \geq 2$nd moment:

$$\theta(P) := \mathbb{E}_P[X], \quad \mathbb{E}_P[|X|^k] \leq 1.$$

Proposition (D., Jordan, Wainwright):
Non-private minimax rate

$$\frac{1}{n} \lesssim \mathbb{E}[(\widehat{\theta} - \mathbb{E}[X])^2] \lesssim \frac{1}{n}$$

Example 1: Mean estimation

Problem: Estimate mean of distributions P with $k \geq 2$nd moment:

$$\theta(P) := \mathbb{E}_P[X], \quad \mathbb{E}_P[|X|^k] \leq 1.$$

Proposition (D., Jordan, Wainwright):
Private minimax rate

$$\frac{1}{(n\alpha^2)^{\frac{k-1}{k}}} \lesssim \mathbb{E}[(\widehat{\theta} - \mathbb{E}[X])^2] \lesssim \frac{1}{(n\alpha^2)^{\frac{k-1}{k}}}$$

Example 1: Mean estimation

Problem: Estimate mean of distributions P with $k \geq 2$nd moment:

$$\theta(P) := \mathbb{E}_P[X], \quad \mathbb{E}_P[|X|^k] \leq 1.$$

Proposition (D., Jordan, Wainwright):
Private minimax rate

$$\frac{1}{(n\alpha^2)^{\frac{k-1}{k}}} \lesssim \mathfrak{M}_n(\mathbb{E}[X], (\cdot)^2, \alpha) \lesssim \frac{1}{(n\alpha^2)^{\frac{k-1}{k}}}$$

Example 1: Mean estimation

Problem: Estimate mean of distributions P with $k \ge 2$nd moment:

$$\theta(P) := \mathbb{E}_P[X], \quad \mathbb{E}_P[|X|^k] \le 1.$$

Proposition (D., Jordan, Wainwright):
Private minimax rate

$$\frac{1}{(n\alpha^2)^{\frac{k-1}{k}}} \lesssim \mathfrak{M}_n(\mathbb{E}[X], (\cdot)^2, \alpha) \lesssim \frac{1}{(n\alpha^2)^{\frac{k-1}{k}}}$$

Examples:

- For two moments $k = 2$, rate goes from parametric $1/n$ to $1/\sqrt{n\alpha^2}$

Example 1: Mean estimation

Problem: Estimate mean of distributions P with $k \geq 2$nd moment:

$$\theta(P) := \mathbb{E}_P[X], \quad \mathbb{E}_P[|X|^k] \leq 1.$$

Proposition (D., Jordan, Wainwright):
Private minimax rate

$$\frac{1}{(n\alpha^2)^{\frac{k-1}{k}}} \lesssim \mathfrak{M}_n(\mathbb{E}[X], (\cdot)^2, \alpha) \lesssim \frac{1}{(n\alpha^2)^{\frac{k-1}{k}}}$$

Examples:

- For two moments $k = 2$, rate goes from parametric $1/n$ to $1/\sqrt{n\alpha^2}$
- For $k \to \infty$ (bounded random variables) parametric decrease

$$n \mapsto n\alpha^2$$

Example 2: multinomial estimation

Problem: Get observations $X \in [d]$ and wish to estimate

$$\theta_j := P(X = j)$$

Example 2: multinomial estimation

Problem: Get observations $X \in [d]$ and wish to estimate

$$\theta_j := P(X = j)$$

Example:

Example 2: multinomial estimation

Problem: Get observations $X \in [d]$ and wish to estimate

$$\theta_j := P(X = j)$$

Example:

- ______ \$0–\$10,000
- ______ \$10,001–\$20,000
- ______ \$20,001–\$40,000
- ______ \$40,001–\$80,000
- ______ \$80,001–\$160,000
- ______ \$160,001–\$320,000
- ______ \$320,001+

Example 2: multinomial estimation

Problem: Get observations $X \in [d]$ and wish to estimate

$$\theta_j := P(X = j)$$

Example:

- _____ \$0–\$10,000
- _____ \$10,001–\$20,000
- _____ \$20,001–\$40,000
- _____ \$40,001–\$80,000
- _____ \$80,001–\$160,000
- _____ \$160,001–\$320,000
- _____ \$320,001+

$$\theta_1 = .05$$
$$\theta_2 = .1$$
$$\theta_3 = .2$$
$$\theta_4 = .4$$
$$\theta_5 = .2$$
$$\theta_6 = .04$$
$$\theta_7 = .01$$

Example 2: multinomial estimation

Problem: Get observations $X \in [d]$ and wish to estimate

$$\theta_j := P(X = j)$$

Example 2: multinomial estimation

Problem: Get observations $X \in [d]$ and wish to estimate

$$\theta_j := P(X = j)$$

$$\underbrace{\widehat{\theta}_j = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\left\{X_i = j\right\} = \widehat{P}(X = j)}_{\text{Standard estimator}}$$

Example 2: multinomial estimation

Problem: Get observations $X \in [d]$ and wish to estimate

$$\theta_j := P(X = j)$$

$$\underbrace{\widehat{\theta}_j = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{X_i = j\} = \widehat{P}(X = j)}_{\text{Standard estimator (counts)}}$$

Example 2: multinomial estimation

Problem: Get observations $X \in [d]$ and wish to estimate

$$\theta_j := P(X = j)$$

$$\underbrace{\widehat{\theta}_j = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{X_i = j\} = \widehat{P}(X = j)}_{\text{Standard estimator (counts)}}$$

Usual rate:

$$\mathbb{E}\left[\|\widehat{\theta} - \theta\|_2^2\right] \leq \frac{1}{n}.$$

Example 2: multinomial estimation

Problem: Get observations $X \in [d]$ and wish to estimate

$$\theta_j := P(X = j)$$

Example 2: multinomial estimation

Problem: Get observations $X \in [d]$ and wish to estimate

$$\theta_j := P(X = j)$$

Proposition: Non-private minimax rate

$$\frac{1}{n} \lesssim \mathbb{E}\left[\|\widehat{\theta} - \theta\|_2^2\right] \lesssim \frac{1}{n}$$

Example 2: multinomial estimation

Problem: Get observations $X \in [d]$ and wish to estimate

$$\theta_j := P(X = j)$$

Proposition: Private minimax rate

$$\frac{d}{(n\alpha^2)} \lesssim \mathbb{E}\left[\|\widehat{\theta} - \theta\|_2^2\right] \lesssim \frac{d}{(n\alpha^2)}$$

Example 2: multinomial estimation

Problem: Get observations $X \in [d]$ and wish to estimate

$$\theta_j := P(X = j)$$

Proposition: Private minimax rate

$$\frac{d}{(n\alpha^2)} \lesssim \mathfrak{M}_n([P(X = j)]_{j=1}^d, \|\cdot\|_2^2, \alpha) \lesssim \frac{d}{(n\alpha^2)}$$

Example 2: multinomial estimation

Problem: Get observations $X \in [d]$ and wish to estimate

$$\theta_j := P(X = j)$$

Proposition: Private minimax rate

$$\frac{d}{(n\alpha^2)} \lesssim \mathfrak{M}_n([P(X = j)]_{j=1}^d, \|\cdot\|_2^2, \alpha) \lesssim \frac{d}{(n\alpha^2)}$$

Take away: Sample size reduction

$$\boldsymbol{n \mapsto \frac{n\alpha^2}{d}}$$

Example 2: multinomial estimation

Problem: Get observations $X \in [d]$ and wish to estimate

$$\theta_j := P(X = j)$$

Example 2: multinomial estimation

Problem: Get observations $X \in [d]$ and wish to estimate

$$\theta_j := P(X = j)$$

- Optimal mechanism: randomized response. Resample each coordinate by Bernoulli coin flips

Example 2: multinomial estimation

Problem: Get observations $X \in [d]$ and wish to estimate

$$\theta_j := P(X = j)$$

- Optimal mechanism: randomized response. Resample each coordinate by Bernoulli coin flips
  - ____ \$0–\$10,000
  - ____ \$10,001–\$20,000
  - _X_ \$20,001–\$40,000
  - ____ \$40,001–\$80,000
  - ____ \$80,001–\$160,000
  - ____ \$160,001–\$320,000
  - ____ \$320,001+

Example 2: multinomial estimation

Problem: Get observations $X \in [d]$ and wish to estimate

$$\theta_j := P(X = j)$$

- Optimal mechanism: randomized response. Resample each coordinate by Bernoulli coin flips

  - ____ \$0–\$10,000
  - ____ \$10,001–\$20,000
  - _X_ \$20,001–\$40,000
  - ____ \$40,001–\$80,000
  - ____ \$80,001–\$160,000
  - ____ \$160,001–\$320,000
  - ____ \$320,001+

  - ____ \$0–\$10,000
  - _X_ \$10,001–\$20,000
  - _X_ \$20,001–\$40,000
  - ____ \$40,001–\$80,000
  - ____ \$80,001–\$160,000
  - ____ \$160,001–\$320,000
  - _X_ \$320,001+

Main consequences

Goal: Understand tradeoff between differential privacy bound α and sample size n

"Theorem 1" Effective sample size for *essentially any*[1] problem is made worse by at least

$$n \mapsto n\alpha^2$$

Main consequences

Goal: Understand tradeoff between differential privacy bound α and sample size n

"Theorem 1" Effective sample size for *essentially any*[1] problem is made worse by at least

$$n \mapsto n\alpha^2$$

[1] *essentially any*: any problem whose minimax rate can be controlled by information-theoretic techniques

Main consequences

Goal: Understand tradeoff between differential privacy bound α and sample size n

"Theorem 1" Effective sample size for *essentially any*[1] problem is made worse by at least

$$n \mapsto n\alpha^2$$

[1] *essentially any*: any problem whose minimax rate can be controlled by information-theoretic techniques

"Theorem 2" Effective sample size for d-dimensional problems scales as

$$n \mapsto \frac{n\alpha^2}{d}$$

General theory

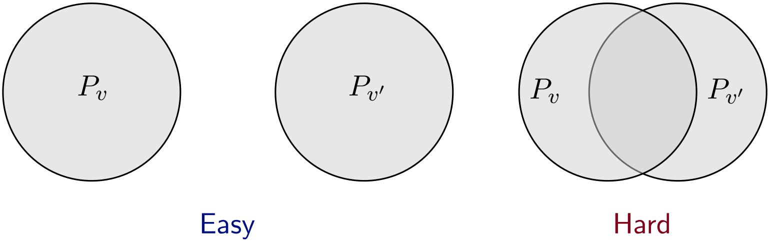
Showing minimax bounds:

- Have possible "true" parameters $\{\theta_v\}$ we want to find
- Distribution P_v associated with each parameter
- Problem is *hard* when $P_v \approx P_{v'}$

General theory

Showing minimax bounds:

- Have possible "true" parameters $\{\theta_v\}$ we want to find
- Distribution P_v associated with each parameter
- Problem is *hard* when $P_v \approx P_{v'}$



Differential privacy and probability distributions

Samples: Z_i are drawn $X_i \to Q \to Z_i$ from *marginal*

$$M_v(Z) := \int Q(Z \mid X = x) dP_v(x)$$

Differential privacy and probability distributions

Samples: Z_i are drawn $X_i \to Q \to Z_i$ from *marginal*

$$M_v(Z) := \int Q(Z \mid X = x) dP_v(x)$$

Strong data processing: If $Q(Z \mid x)/Q(Z \mid x') \le e^{\alpha}$,

$$D_{\rm kl}\left(M_1 \| M_2\right) + D_{\rm kl}\left(M_2 \| M_1\right) \le 4\boldsymbol{(e^{\alpha} - 1)^2} \left\|P_1 - P_2\right\|_{\rm TV}^2$$

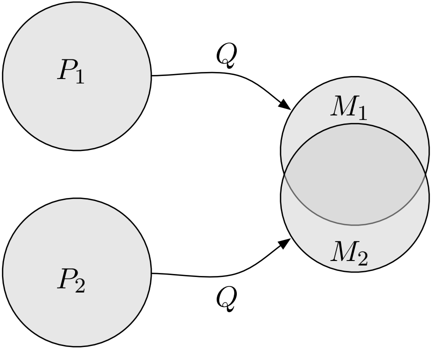
Differential privacy and probability distributions

Samples: Z_i are drawn $X_i \to Q \to Z_i$ from *marginal*

$$M_v(Z) := \int Q(Z \mid X = x) dP_v(x)$$

Strong data processing: If $Q(Z \mid x)/Q(Z \mid x') \le e^{\alpha}$,

$$D_{\mathrm{kl}}(M_1 \| M_2) + D_{\mathrm{kl}}(M_2 \| M_1) \le 4 \boldsymbol{(e^{\alpha} - 1)^2} \|P_1 - P_2\|_{\mathrm{TV}}^2$$



Contraction and lower bounds

Samples: Z_i are drawn $X_i \to Q \to Z_i$ from *marginal*

$$M_v(Z) := \int Q(Z \mid X = x) dP_v(x)$$

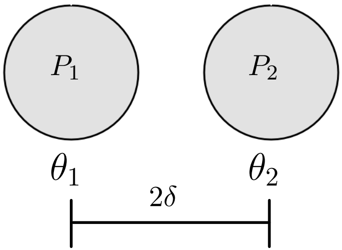
Contraction and lower bounds

Samples: Z_i are drawn $X_i \to Q \to Z_i$ from *marginal*

$$M_v(Z) := \int Q(Z \mid X = x) dP_v(x)$$

Le Cam's Method

- θ_1 and θ_2 are 2δ separated



Contraction and lower bounds

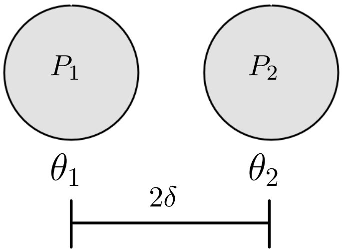
Samples: Z_i are drawn $X_i \to Q \to Z_i$ from *marginal*

$$M_v(Z) := \int Q(Z \mid X = x) dP_v(x)$$

Le Cam's Method

- θ_1 and θ_2 are 2δ separated
- Non-private version:

$$\mathfrak{M}_n(\Theta, (\cdot)^2) \geq \delta^2 \left(1 - \sqrt{n D_{\mathrm{kl}}\left(P_1 \| P_2\right)}\right)$$



Contraction and lower bounds

Samples: Z_i are drawn $X_i \to Q \to Z_i$ from *marginal*

$$M_v(Z) := \int Q(Z \mid X = x) dP_v(x)$$

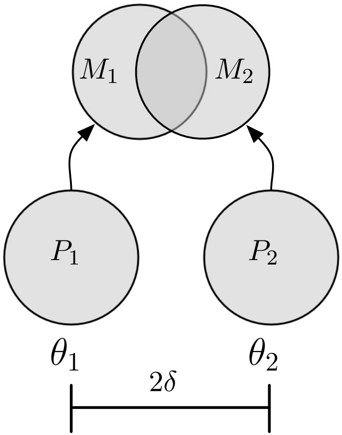
Le Cam's Method

- θ_1 and θ_2 are 2δ separated
- Non-private version:

$$\mathfrak{M}_n(\Theta, (\cdot)^2) \geq \delta^2 \left(1 - \sqrt{n D_{\mathrm{kl}}\left(P_1 \| P_2\right)}\right)$$

- Private version:

$$\mathfrak{M}_n\left(\Theta, (\cdot)^2, \alpha\right) \geq \delta^2 \left(1 - \sqrt{n \alpha^2 \left\| P_1 - P_2 \right\|_{\mathrm{TV}}^2}\right)$$



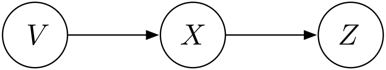
Variational results on privacy and probability distributions

Samples: Z_i are drawn $X_i \to Q \to Z_i$ from *marginal*

$$M_v(Z) := \int Q(Z \mid X = x) dP_v(x)$$

Canonical problem: Nature samples V uniformly from $v = 1, \ldots, K$ and draws

$$X_i \stackrel{\text{i.i.d.}}{\sim} P_v \text{ when } V = v$$



Goal: Find V based on $Z_1, \ldots, Z_n$

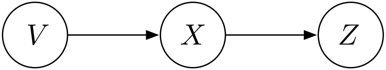
Variational results on privacy and probability distributions

Samples: Z_i are drawn $X_i \to Q \to Z_i$ from *marginal*

$$M_v(Z) := \int Q(Z \mid X = x) dP_v(x)$$

Canonical problem: Nature samples V uniformly from $v = 1, \ldots, K$ and draws

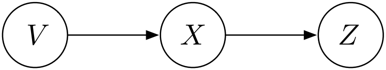
$$X_i \stackrel{\text{i.i.d.}}{\sim} P_v \;\text{ when } V = v$$



Goal: Find V based on $Z_1, \ldots, Z_n$
Difficulty of problem: Saw earlier *mutual information*

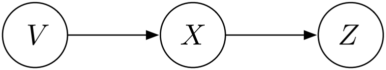
$$I(X_1, \ldots, X_n; V) \quad \mapsto \quad I(Z_1, \ldots, Z_n; V)$$

Fano inequality, lower bounds, contraction



- Have parameters $\theta_1, \ldots, \theta_K$, choose randomly $V \in [K]$
- Sample X_i according to θ_v when $V = v$
- Sample Z_i according to $Q(\cdot \mid X_i)$

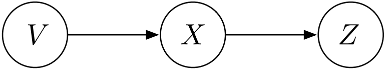
Fano inequality, lower bounds, contraction



- Have parameters $\theta_1, \ldots, \theta_K$, choose randomly $V \in [K]$
- Sample X_i according to θ_v when $V = v$
- Sample Z_i according to $Q(\cdot \mid X_i)$
- **Non-private Fano inequality:**

$$\mathbb{P}(\text{Error}) \geq 1 - \frac{I(X_1, \ldots, X_n; V)}{\log K} - o(1)$$

Fano inequality, lower bounds, contraction



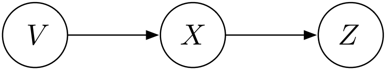
- Have parameters $\theta_1, \ldots, \theta_K$, choose randomly $V \in [K]$
- Sample X_i according to θ_v when $V = v$
- Sample Z_i according to $Q(\cdot \mid X_i)$
- **Non-private Fano inequality:**

$$\mathbb{P}(\text{Error}) \geq 1 - \frac{I(X_1, \ldots, X_n; V)}{\log K} - o(1)$$

- **Private Fano inequality:**

$$\mathbb{P}(\text{Error}) \geq 1 - \frac{I(\color{red}{Z_1, \ldots, Z_n}; V)}{\log K} - o(1)$$

Fano inequality, lower bounds, contraction



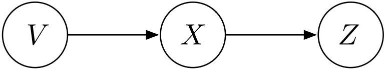
- Have parameters $\theta_1, \ldots, \theta_K$, choose randomly $V \in [K]$
- Sample X_i according to θ_v when $V = v$
- Sample Z_i according to $Q(\cdot \mid X_i)$
- **Non-private Fano inequality:**

$$\mathbb{P}(\text{Error}) \geq 1 - \frac{I(X_1, \ldots, X_n; V)}{\log K} - o(1)$$

- **Private Fano inequality:**

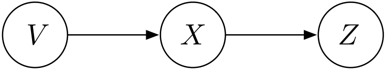
$$\mathbb{P}(\text{Error}) \geq 1 - \frac{\boldsymbol{\alpha^2} I(X_1, \ldots, X_n; V)}{\boldsymbol{d} \log K} - o(1)$$

Variational results on privacy and probability distributions



- Define mixture $\overline{P} = \frac{1}{K} \sum_{v=1}^{K} P_v$

Variational results on privacy and probability distributions

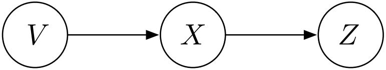


- Define mixture $\overline{P} = \frac{1}{K} \sum_{v=1}^{K} P_v$

Mutual information contraction: For any non-interactive α-locally private channel Q,

$$I(Z_1, \ldots, Z_n; V) \leq n\boldsymbol{(e^{\alpha} - 1)^2} \sup_S \frac{1}{K} \sum_{v=1}^{K} \left(P_v(S) - \overline{P}(S)\right)^2$$

Variational results on privacy and probability distributions

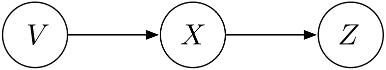


- Define mixture $\overline{P} = \frac{1}{K}\sum_{v=1}^{K} P_v$

Mutual information contraction: For any non-interactive α-locally private channel Q,

$$I(Z_1, \ldots, Z_n; V) \le n\boldsymbol{(e^{\alpha} - 1)^2} \quad \underbrace{\sup_S \frac{1}{K}\sum_{v=1}^{K} \left(P_v(S) - \overline{P}(S)\right)^2}_{\text{Dimension-dependent total variation}}$$

Variational results on privacy and probability distributions



- Define mixture $\overline{P} = \frac{1}{K} \sum_{v=1}^{K} P_v$

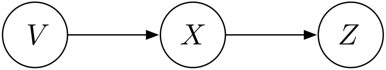
Mutual information contraction: For any non-interactive α-locally private channel Q,

$$I(Z_1, \ldots, Z_n; V) \le n \boldsymbol{(e^{\alpha} - 1)^2} \underbrace{\sup_S \frac{1}{K} \sum_{v=1}^{K} \left(P_v(S) - \overline{P}(S)\right)^2}_{\text{Dimension-dependent total variation}}$$

What happens? Roughly

$$n \sup_S \frac{1}{K} \sum_{v=1}^{K} \left(P_v(S) - \overline{P}(S)\right)^2 \approx \frac{1}{d} I(X_1, \ldots, X_n; V)$$

Variational results on privacy and probability distributions



- Define mixture $\overline{P} = \frac{1}{K} \sum_{v=1}^{K} P_v$

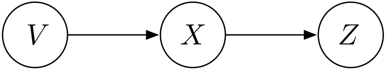
Mutual information contraction: For any non-interactive α-locally private channel Q,

$$I(Z_1, \ldots, Z_n; V) \le n \boldsymbol{(e^{\alpha} - 1)^2} \quad \underbrace{\sup_S \frac{1}{K} \sum_{v=1}^{K} \left(P_v(S) - \overline{P}(S)\right)^2}_{\text{Dimension-dependent total variation}}$$

What happens? Roughly

$$n \sup_S \frac{1}{K} \sum_{v=1}^{K} \left(P_v(S) - \overline{P}(S)\right)^2 \approx \frac{1}{d} \underbrace{I(X_1, \ldots, X_n; V)}_{\text{Classical}}$$

Variational results on privacy and probability distributions



- Define mixture $\overline{P} = \frac{1}{K} \sum_{v=1}^{K} P_v$

Mutual information contraction: For any non-interactive α-locally private channel Q,

$$I(Z_1, \ldots, Z_n; V) \le n\boldsymbol{(e^{\alpha} - 1)^2} \quad \underbrace{\sup_S \frac{1}{K} \sum_{v=1}^{K} \big(P_v(S) - \overline{P}(S)\big)^2}_{\text{Dimension-dependent total variation}}$$

What happens? Roughly

$$I(Z_1, \ldots, Z_n; V) \le \frac{\alpha^2}{d} \underbrace{I(X_1, \ldots, X_n; V)}_{\text{Classical}}$$

Summary

High level results:

- Formal minimax framework for *local* differential privacy

Summary

High level results:

- Formal minimax framework for *local* differential privacy
- Two main theorems bound distances between probability distributions as function of privacy

Summary

High level results:

- Formal minimax framework for *local* differential privacy
- Two main theorems bound distances between probability distributions as function of privacy
  - Pairwise contraction: Le Cam's method
  - Mutual information contraction: Fano's method

Summary

High level results:

- Formal minimax framework for *local* differential privacy
- Two main theorems bound distances between probability distributions as function of privacy
  - Pairwise contraction: Le Cam's method
  - Mutual information contraction: Fano's method

Extensions and other conclusions:

- In essentially any problem, effective number of samples

$$n \mapsto n\alpha^2$$

- In d-dimensional problems, effective number of samples

$$n \mapsto \frac{n\alpha^2}{d}$$

Summary

High level results:

- Formal minimax framework for *local* differential privacy
- Two main theorems bound distances between probability distributions as function of privacy
  - Pairwise contraction: Le Cam's method
  - Mutual information contraction: Fano's method

Extensions and other conclusions:

- In essentially any problem, effective number of samples

$$n \mapsto n\alpha^2$$

- In d-dimensional problems, effective number of samples

$$n \mapsto \frac{n\alpha^2}{d}$$

- Rates for regression, multinomial estimation, convex optimization
- Dimension-dependent effects: High-dimensional problems impossible (no logarithmic dependence on dimension)

Summary

High level results:

- Formal minimax framework for *local* differential privacy
- Two main theorems bound distances between probability distributions as function of privacy
  - Pairwise contraction: Le Cam's method
  - Mutual information contraction: Fano's method

Extensions and other conclusions:

- In essentially any problem, effective number of samples

$$n \mapsto n\alpha^2$$

- In d-dimensional problems, effective number of samples

$$n \mapsto \frac{n\alpha^2}{d}$$

- Rates for regression, multinomial estimation, convex optimization
- Dimension-dependent effects: High-dimensional problems impossible (no logarithmic dependence on dimension)
- Identification of optimal mechanism requires geometric understanding

Summary

High level results:

- Formal minimax framework for *local* differential privacy
- Two main theorems bound distances between probability distributions as function of privacy
  - Pairwise contraction: Le Cam's method
  - Mutual information contraction: Fano's method

Summary

High level results:

- Formal minimax framework for *local* differential privacy
- Two main theorems bound distances between probability distributions as function of privacy
  - Pairwise contraction: Le Cam's method
  - Mutual information contraction: Fano's method
- Trade-offs between privacy and statistical utility

Summary

High level results:

- Formal minimax framework for *local* differential privacy
- Two main theorems bound distances between probability distributions as function of privacy
  - Pairwise contraction: Le Cam's method
  - Mutual information contraction: Fano's method
- Trade-offs between privacy and statistical utility
- Many open problems:

Summary

High level results:

- Formal minimax framework for *local* differential privacy
- Two main theorems bound distances between probability distributions as function of privacy
  - Pairwise contraction: Le Cam's method
  - Mutual information contraction: Fano's method

- Trade-offs between privacy and statistical utility
- Many open problems:
  - Other models of privacy?
  - Non-local notions of privacy?
  - Privacy without knowing statistical objective?

Summary

High level results:

- Formal minimax framework for *local* differential privacy
- Two main theorems bound distances between probability distributions as function of privacy
  - Pairwise contraction: Le Cam's method
  - Mutual information contraction: Fano's method
- Trade-offs between privacy and statistical utility
- Many open problems:
  - Other models of privacy?
  - Non-local notions of privacy?
  - Privacy without knowing statistical objective?

Pre/post-print online: "Local privacy and statistical minimax rates." D., Jordan, & Wainwright (2013). *arXiv:1302.3203 [stat.TH]*