CSCIT 2021 - Lecture 3

Clément Canonne (University of Sydney)

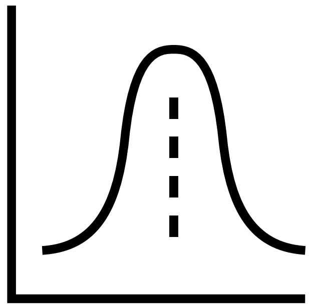
Estimation and hypothesis testing under information constraints

Last lecture: recap

1. Learning and testing discrete distributions: upper bounds
    - Learning, under communication or local privacy (LDP) constraints
    - Testing, under communication or LDP constraints
2. Lower bounds
    - A general bound for learning and testing
    - Application to communication and LDP

Contents of this lecture

1. Estimation for high-dimensional distributions: upper bounds
    - Mean estimation under communication or local privacy (LDP) constraints
2. Lower bounds
    - A general bound for estimation (in the interactive setting)
    - Application to communication and LDP



Caveat: the Roads not Taken

- These 3 lectures cover specifically work from arXiv:1812.11476, arXiv:2007.10976, and arXiv:2010.06562

- There are others! See references at the end

- Only covers "local" constraints — e.g., not central differential privacy. See, for instance, arXiv:2005.00010

Recall: What are ~~learning~~ estimating and testing?

Standard statistical setting: n iid samples from some unknown probability distribution p

Goal: estimate something about p

↳ learn a parameter/functional θ of p

output $\hat{\theta}$ such that

$$\mathbb{E}_p\left[\ell(\hat{\theta}, \theta(p))\right] \leq \varepsilon$$

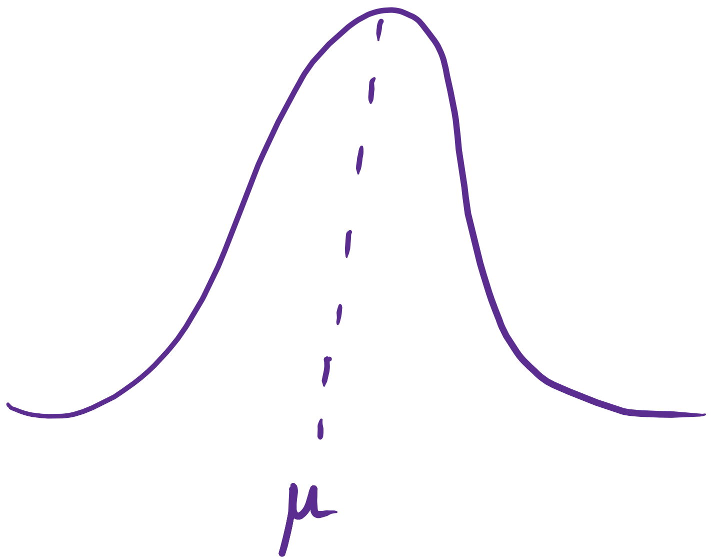
for instance, the mean of p

Recap: what are we learning?

High-dimensional Gaussians (with identity covariance)

dimension d

Learning the mean under ℓ_2^2 loss



$p = \mathcal{N}(\mu, I_d)$

$\in \mathbb{R}^d$

Recap: the "centralised" setting

$d \gg 1$

$\varepsilon \in (0,1]$

Identity-covariance Gaussians

Theorem. Learning the mean of an unknown $\mathcal{N}(\mu, I_d)$ to ℓ_2^2 loss ε^2 has sample complexity $\Theta\left(\frac{d}{\varepsilon^2}\right)$.

Recap: the "centralised" setting

$d \gg 1$

$\varepsilon \in (0,1]$

Identity-covariance Gaussians

Theorem. Learning the mean of an unknown $\mathcal{N}(\mu, I_d)$ to ℓ_2^2 loss ε^2 has sample complexity $\Theta\left(\frac{d}{\varepsilon^2}\right)$.

Proof. For the upper bound, the empirical mean works!

$$\mathbb{E}\left[\|\bar{X}-\mu\|_2^2\right] = \sum_{i=1}^{d} \mathbb{E}\left[(\bar{X}_i-\mu_i)^2\right] = \sum_{i=1}^{d} \mathbb{E}\left[\frac{1}{n^2}\sum_{t,s} (X_{t,i}-\mu_i)(X_{s,i}-\mu_i)\right]$$

$$\bar{X} = \frac{1}{n}\sum_{t=1}^{n} X_t$$

Recap: the "centralised" setting

$d \gg 1$

$\varepsilon \in (0,1]$

Identity-covariance Gaussians

Theorem. Learning the mean of an unknown $\mathcal{N}(\mu, I_d)$ to ℓ_2^2 loss ε^2 has sample complexity $\Theta\left(\frac{d}{\varepsilon^2}\right)$.

Proof. For the upper bound, the empirical mean works!

$$\mathbb{E}\left[\|\bar{X}-\mu\|_2^2\right] = \sum_{i=1}^{d} \mathbb{E}\left[(\bar{X}_i-\mu_i)^2\right] = \sum_{i=1}^{d} \mathbb{E}\Big[\frac{1}{n^2} \underbrace{\sum_t}_{\text{indep}^t} (X_{t,i}-\mu_i)^2\Big] = \frac{d}{n^2}\cdot n \cdot \underbrace{1}_{\sigma^2=1}$$

Recap: the "centralised" setting

$d \gg 1$

$\varepsilon \in (0,1]$

Identity-covariance Gaussians

Theorem. Learning the mean of an unknown $\mathcal{N}(\mu, I_d)$ to ℓ_2^2 loss ε^2 has sample complexity $\Theta\left(\frac{d}{\varepsilon^2}\right)$.

Proof. For the upper bound, the empirical mean works!

$$\mathbb{E}\left[\|\bar{X}-\mu\|_2^2\right] = \sum_{i=1}^{d} \mathbb{E}\left[(\bar{X}_i-\mu_i)^2\right] = \sum_{i=1}^{d} \mathbb{E}\left[\frac{1}{n^2}\sum_t (X_{t,i}-\mu_i)^2\right] = \frac{d}{n} \leq \varepsilon^2$$

$\square$

Now, under constraints?

$d \gg 1$

$\varepsilon \in (0,1]$

Identity-covariance Gaussians

Theorem. Learning the mean of an unknown $\mathcal{N}(\mu, I_d)$ to ℓ_2^2 loss ε^2 under ℓ-bit communication constraints has sample complexity ______.

Theorem. Learning the mean of an unknown $\mathcal{N}(\mu, I_d)$ to ℓ_2^2 loss ε^2 under ρ-local privacy (LDP) constraints has sample complexity ______.

Now, under constraints?

$d \gg 1$

$\varepsilon \in (0,1]$

Identity-covariance Gaussians

Theorem. Learning the mean of an unknown $\mathcal{N}(\mu, I_d)$ to ℓ_2^2 loss ε^2 under ℓ-bit communication constraints has sample complexity $\Theta\left(\frac{d^2}{\ell \varepsilon^2}\right)$.

Theorem. Learning the mean of an unknown $\mathcal{N}(\mu, I_d)$ to ℓ_2^2 loss ε^2 under ϱ-local privacy (LDP) constraints has sample complexity $\Theta\left(\frac{d^2}{\varrho^2 \varepsilon^2}\right)$.

Now, under constraints?

$d \gg 1$

$\varepsilon \in (0,1]$

Identity-covariance Gaussians

Theorem. Learning the mean of an unknown $\mathcal{N}(\mu, I_d)$ to ℓ_2^2 loss ε^2 under ℓ-bit communication constraints has sample complexity $\Theta\left(\frac{d^2}{\ell\varepsilon^2}\right)$. Further, attained by a private-coin protocol.

Theorem. Learning the mean of an unknown $\mathcal{N}(\mu, I_d)$ to ℓ_2^2 loss ε^2 under ϱ-local privacy (LDP) constraints has sample complexity $\Theta\left(\frac{d^2}{\varrho^2\varepsilon^2}\right)$. Further, attained by a private-coin protocol.

Now, under constraints?

$d \gg 1$

$\varepsilon \in (0,1]$

Identity-covariance Gaussians

Some remarks:

* Dependence is $\frac{1}{\ell}$, not $\frac{1}{2^\ell}$
* Generalises to s-sparse mean estimation (but private-coin << interactive)
* For simplicity, will consider related Bernoulli mean estimation:

Now, under constraints?

$d \gg 1$

$\varepsilon \in (0,1]$

Identity-covariance Gaussians

Some remarks:

* Dependence is $\frac{1}{\ell}$, not $\frac{1}{2^\ell}$

* Generalises to s-sparse mean estimation (but private-coin << interactive)

* For simplicity, will consider related Bernoulli mean estimation:

$$p = p_1 \otimes p_2 \otimes \cdots \otimes p_d \quad \text{on } \{\pm 1\}^d$$

$$\mu = \mathbb{E}_p[X] \in [-1,1]^d$$

Now, under constraints?

$d \gg 1$

$\varepsilon \in (0,1]$

Product Bernoulli distributions

Some remarks:

* Dependence is $\frac{1}{\ell}$, not $\frac{1}{2^\ell}$
* Generalises to s-sparse mean estimation (but private-coin << interactive)
* For simplicity, will consider related Bernoulli mean estimation:

$$\mathbf{p} = p_1 \otimes p_2 \otimes \cdots \otimes p_d \quad \text{on } \{\pm 1\}^d$$

$$\mu = \mathbb{E}_{\mathbf{p}}[X] \in [-1,1]^d$$

How to prove the lower bound?

How to prove the lower bound?

Application of a general lower bound framework (generalising the learning part from lecture 2):

How to prove the lower bound?

Application of a general lower bound framework (generalising the learning part from lecture 2):

1. Come up with family $\{P_z\}_{z \in \{\pm 1\}^d}$ of hard instances

How to prove the lower bound?

Application of a general lower bound framework (generalising the learning part from lecture 2):

$\mathcal{P}$

① Come up with family $\{P_z\}_{z\in\{\pm 1\}^d}$ of hard instances

② Check that $\mathcal{P}$ satisfies 3 (or 4) assumptions

How to prove the lower bound?

Application of a general lower bound framework (generalising the learning part from lecture 2):

$\mathcal{P}$

1. Come up with family $\{p_z\}_{z \in \{\pm 1\}^d}$ of hard instances
2. Check that $\mathcal{P}$ satisfies 3 (or 4) assumptions
3. Get a lower bound against interactive protocols

How to prove the lower bound?

Application of a general lower bound framework (generalising the learning part from lecture 2):

$\mathcal{P}$

1. Come up with family $\{P_z\}_{z\in\{\pm 1\}^d}$ of hard instances
2. Check that $\mathcal{P}$ satisfies 3 (or 4) assumptions
3. Get a lower bound against interactive protocols
4. (Compare the lower bound to the known upper bound(s).)

How to prove the lower bound?

① Come up with family $\{p_z\}_{z\in\{\pm1\}^d}$ of hard instances

How to prove the lower bound?

① Come up with family $\{p_z\}_{z \in \{\pm 1\}^d}$ of hard instances

"Perturbation around uniform": each coordinate has mean $\pm \frac{\varepsilon}{\sqrt{d}}$

$$\mu_z = \frac{\varepsilon}{\sqrt{d}} z \, , \quad z \in \{\pm 1\}^d$$

① $\|\mu_z\|_2 = \varepsilon$

② $p_z = \bigotimes_{i=1}^{d} \mathrm{Rademacher}\left(\frac{\mu_{z,i}+1}{2}\right)$

How to prove the lower bound?

② Check that $\mathcal{P}$ satisfies 3 (or 4) assumptions

Assumption 1. For all $z, z' \in \{\pm 1\}^d$,

$$\ell_2^2(\mu_z, \mu_{z'}) \geq 8\varepsilon^2 \frac{\mathrm{Ham}(z, z')}{d}$$

→ "Additive loss"

How to prove the lower bound?

② Check that $\mathcal{P}$ satisfies 3 (or 4) assumptions

Assumption 1. For all $z, z' \in \{\pm 1\}^d$,

$$\ell_2^2(\mu_z, \mu_{z'}) \geq 8\varepsilon^2 \frac{\mathrm{Ham}(z, z')}{d}$$

→ "Additive loss"

Satisfied $\left(\mu = \mathrm{cst.}\, \frac{\varepsilon}{\sqrt{d}} z\right)$

How to prove the lower bound?

② Check that $\mathcal{P}$ satisfies 3 (or 4) assumptions

Assumption 2. For all $z \in \{\pm 1\}^d$ and $1 \leq i \leq d$, $\exists\, \alpha_{z,i}, \phi_{z,i}$

$$\frac{dP_{z \oplus i}}{dP_z} = 1 + \alpha_{z,i} \phi_{z,i}$$

with $|\alpha_{z,i}| \leq \alpha$.

How to prove the lower bound?

② Check that $\mathcal{P}$ satisfies 3 (or 4) assumptions

Assumption 2. For all $z \in \{\pm 1\}^d$ and $1 \leq i \leq d$, $\exists\, \alpha_{z,i}, \phi_{z,i}$

$$\frac{dp_{z^{\oplus i}}}{dp_z} = 1 + \alpha_{z,i}\phi_{z,i}$$

$z^{\oplus i}$: z with i-th coordinate flipped

with $|\alpha_{z,i}| \leq \alpha$.

indep^t of z, i

How to prove the lower bound?

② Check that $\mathcal{P}$ satisfies 3 (or 4) assumptions

Assumption 2. For all $z \in \{\pm 1\}^d$ and $1 \leq i \leq d$, $\exists\, \alpha_{z,i}, \phi_{z,i}$

$$\frac{dP_{z^{\oplus i}}}{dP_z} = 1 + \alpha_{z,i}\phi_{z,i}$$

$z^{\oplus i}$: z with i-th coordinate flipped

with $|\alpha_{z,i}| \leq \alpha$.

indep^t of z, i

$\rightsquigarrow$ "Densities exist"

How to prove the lower bound?

$\gamma = \frac{\varepsilon}{\sqrt{d}}$

② Check that P satisfies 3 (or 4) assumptions

Assumption 2. For all $z \in \{\pm 1\}^d$ and $1 \leq i \leq d$, $\exists\, \alpha_{z,i}, \phi_{z,i}$

$$\frac{dP_{z \oplus i}}{dP_z} = 1 + \alpha_{z,i} \phi_{z,i}$$

with $|\alpha_{z,i}| \leq \alpha$.

Satisfied* with $\alpha = \alpha_{z,i} = \frac{2\gamma}{\sqrt{1-\gamma}}$

and

$\phi_{z,i}(x) = -\frac{x_i z_i - \gamma}{\sqrt{1-\gamma^2}}$

How to prove the lower bound?

$\gamma = \frac{\varepsilon}{\sqrt{d}}$

② Check that $\mathcal{P}$ satisfies 3 (or 4) assumptions

Assumption 3. For all $1 \leq i, j \leq d$ and $z \in \{\pm 1\}^d$,

$$\mathbb{E}_{P_z}\left[\phi_{z,i}(x)\,\phi_{z,j}(x)\right] = \mathbb{1}_{i=j}$$

How to prove the lower bound?

$\gamma = \frac{\varepsilon}{\sqrt{d}}$

② Check that $\mathcal{P}$ satisfies 3 (or 4) assumptions

Assumption 3. For all $1 \le i, j \le d$ and $z \in \{\pm 1\}^d$,

$$\mathop{\mathbb{E}}_{P_z}\left[\phi_{z,i}(x)\,\phi_{z,j}(x)\right] = \mathbb{1}_{i=j}$$

"Orthonormality"

Note that $\mathop{\mathbb{E}}_{P_z}[\phi_{z,i}(x)] = 0$ by Assumption 2.

How to prove the lower bound?

$\gamma = \frac{\varepsilon}{\sqrt{d}}$

② Check that P satisfies 3 (or 4) assumptions

Assumption 3. For all $1 \leq i, j \leq d$ and $z \in \{\pm 1\}^d$,

$$\mathbb{E}_{P_z}\left[\phi_{z,i}(x)\,\phi_{z,j}(x)\right] = \mathbb{1}_{i=j}$$

"Orthonormality"

Satisfied* with $\alpha = \alpha_{z,i} = \frac{2\gamma}{\sqrt{1-\gamma}}$

and

$$\phi_{z,i}(x) = -\frac{x_i z_i - \gamma}{\sqrt{1-\gamma^2}}$$

*This is why we chose to divide $\alpha_{z,i}\,\phi_{z,i}$ this way.

$\gamma = \frac{\varepsilon}{\sqrt{d}}$

Checking why. For $z \in \{\pm 1\}^d$,

$$\forall x \in \{\pm 1\}^d \quad p_z(x) = \frac{1}{2^d} \prod_{i=1}^{d} (1 + \gamma x_i z_i)$$

and so

$$\frac{p_{z \oplus i}}{p_z}(x) = \underbrace{\left(\prod_{j \neq i} \frac{1 + \gamma x_j z_j}{1 + \gamma x_j z_j} \right)}_{=1} \cdot \frac{1 - \gamma x_i z_i}{1 + \gamma x_i z_i} = 1 - \underbrace{\frac{2 \gamma x_i z_i}{1 + \gamma x_i z_i}}_{\text{gives us } \alpha_{z,i} \phi_{z,i}(x)}$$

and

$$\mathbb{E}_{p_z}\left[\left(\frac{-2 \gamma x_i z_i}{1 + \gamma x_i z_i} \right)^2 \right] = [\ldots] = \frac{4 \gamma^2}{1 - \gamma^2} . \longrightarrow \text{gives us } \alpha_{z,i} .$$

How to prove the lower bound?

$\gamma = \frac{\varepsilon}{\sqrt{d}}$

② Check that $\mathcal{P}$ satisfies 3 (or 4) assumptions

Assumption 4. $\exists \sigma$ s.t., for all $z \in \{\pm 1\}^d$,

$$\phi_z(X) := (\phi_{z,1}(X), \ldots, \phi_{z,d}(X)) \in \mathbb{R}^d$$

is σ^2-subgaussian with indep^t coordinates when $X \sim p_z$.

How to prove the lower bound?

$\gamma = \frac{\varepsilon}{\sqrt{d}}$

② Check that $\mathcal{P}$ satisfies 3 (or 4) assumptions

Assumption 4. $\exists \sigma$ s.t., for all $z \in \{\pm 1\}^d$,

$$\phi_z(X) := (\phi_{z,1}(X), \ldots, \phi_{z,d}(X)) \in \mathbb{R}^d$$

is σ^2-subgaussian with indep^t coordinates when $X \sim p_z$.

"Subgaussianity"

$\langle u, \phi_z(X) \rangle$ is σ^2-subgaussian for every fixed unit vector u.

How to prove the lower bound?

$\gamma = \frac{\varepsilon}{\sqrt{d}}$

② Check that $\mathcal{P}$ satisfies 3 (or 4) assumptions

Assumption 4. $\exists \sigma$ s.t., for all $z \in \{\pm 1\}^d$,

$$\phi_z(X) := (\phi_{z,1}(X), \ldots, \phi_{z,d}(X)) \in \mathbb{R}^d$$

is σ^2-subgaussian with indep^t coordinates when $X \sim p_z$.

"Subgaussianity"

Satisfied for $\sigma^2 := \frac{1+\gamma}{1-\gamma}$ (Hoeffding's lemma)

How to prove the lower bound?

Application of a general lower bound framework (generalising the learning part from lecture 2):

① Come up with family $\{P_z\}_{z\in\{\pm 1\}^d}$ of hard instances $\mathcal{P}$

② Check that $\mathcal{P}$ satisfies 3 (or 4) assumptions

Satisfied with $\alpha = \alpha_{z,i} = \dfrac{2\gamma}{\sqrt{1-\gamma}}$ $\quad \sigma^2 := \dfrac{1+\gamma}{1-\gamma}$

and

$$\phi_{z,i}(x) = -\frac{x_i z_i - \gamma}{\sqrt{1-\gamma^2}}$$

where $\gamma = \dfrac{\varepsilon}{\sqrt{d}}$

How to prove the lower bound?

③ Get a lower bound against interactive protocols

Theorem. If Π is an interactive protocol using $\mathcal{W}$ with n users under ℓ_2^2 loss satisfying Assumptions 1, 2, 3, then

$$\Omega(1) \leq \frac{n\alpha^2}{d} \cdot \max_{z} \max_{W \in \mathcal{W}} \sum_{y \in \mathcal{Y}} \frac{\mathrm{Var}_{p_z}[W(y \mid X)]}{\mathbb{E}_{p_z}[W(y \mid X)]}$$

How to prove the lower bound?

③ Get a lower bound against interactive protocols

Theorem. If Π is an interactive protocol using $\mathcal{W}$ with n users under ℓ_2^2 loss satisfying Assumptions 1, 2, 3, then

$$\Omega(1) \leq \frac{n\alpha^2}{d} \cdot \max_{z} \max_{W \in \mathcal{W}} \sum_{y \in \mathcal{Y}} \frac{\mathrm{Var}_{p_z}[W(y|X)]}{\mathbb{E}_{p_z}[W(y|X)]}$$

since $W(y|X) \in [0,1]$,
$\mathrm{Var}\, W \leq \mathbb{E}[W^2] \leq \mathbb{E}[W]$

How to prove the lower bound?

③ Get a lower bound against interactive protocols

Theorem. If Π is an interactive protocol using $\mathcal{W}$ with n users under ℓ_2^2 loss satisfying Assumptions 1,2,3, then

$$\Omega(1) \leq \frac{n\alpha^2}{d} \cdot \max_{z} \max_{W \in \mathcal{W}} \sum_{y \in \mathcal{Y}} \frac{\mathrm{Var}_{p_z}[W(y|X)]}{\mathbb{E}_{p_z}[W(y|X)]}$$

and, under 1,2,4,

$$\Omega(1) \leq \frac{n\alpha^2\sigma^2}{d} \max_{z} \max_{W \in \mathcal{W}} H(p_z^W)$$

How to prove the lower bound?

③ Get a lower bound against interactive protocols

Theorem. If Π is an interactive protocol using $\mathcal{W}$ with n users under ℓ_2^2 loss satisfying Assumptions 1, 2, 3, then

$$\Omega(1) \leq \frac{n\alpha^2}{d} \cdot \max_{z} \max_{W\in\mathcal{W}} \sum_{y\in\mathcal{Y}} \frac{\mathrm{Var}_{p_z}[W(y\mid X)]}{\mathbb{E}_{p_z}[W(y\mid X)]}$$

and, under 1, 2, 4,

entropy

$$\Omega(1) \leq \frac{n\alpha^2\sigma^2}{d} \max_{z} \max_{W\in\mathcal{W}} H(p_z^W)$$

induced distribution on $\mathcal{Y}$ (by p_z and W)

How to prove the lower bound?

③ Get a lower bound against interactive protocols

Theorem. If Π is an interactive protocol using $\mathcal{W}$ with n users under ℓ_2^2 loss satisfying Assumptions 1, 2, 3, then

$$\Omega(1) \le \frac{n\alpha^2}{d} \cdot \varrho^2 \qquad \text{for privacy } \mathcal{W}_\varrho$$

and, under 1, 2, 4,

$$\Omega(1) \le \frac{n\alpha^2\sigma^2}{d} \ell \qquad \text{for communication } \mathcal{W}_\ell$$

How to prove the lower bound?

③ Get a lower bound against interactive protocols

Theorem. If Π is an interactive protocol using $\mathcal{W}$ with n users under ℓ_2^2 loss satisfying Assumptions 1, 2, 3, then

$$\Omega(1) \leq \frac{n\varepsilon^2}{d^2} \cdot \rho^2$$ for privacy $\mathcal{W}_\rho$

recalling $\alpha \simeq \frac{\varepsilon^2}{d}$

and, under 1, 2, 4,

$$\Omega(1) \leq \frac{n\,\varepsilon^2}{d^2}\,\ell$$ for communication $\mathcal{W}_\ell$

Now, under constraints?

Product Bernoulli distributions

$d \gg 1$

$\varepsilon \in (0,1]$

Theorem. Learning the mean of an unknown product Bernoulli to ℓ_2^2 loss ε^2 under ℓ-bit communication constraints has sample complexity $\Theta\left(\frac{d^2}{\ell \varepsilon^2}\right)$.

Theorem. Learning the mean of an unknown product Bernoulli to ℓ_2^2 loss ε^2 under ρ-local privacy (LDP) constraints has sample complexity $\Theta\left(\frac{d^2}{\rho^2 \varepsilon^2}\right)$.

Further reading

2021 (6)

On Learning Parametric Distributions from Quantized Samples. Septimia Sarbu; and Abdellatif Zaidi. In *Proceedings of the 2021 IEEE International Symposium on Information Theory (ISIT'21)*, June 2021.
bibtex

Inference Under Information Constraints III: Local Privacy Constraints. Jayadev Acharya; Clément L. Canonne; Cody Freitag; Ziteng Sun; and Himanshu Tyagi. *IEEE J. Sel. Areas Inf. Theory*, 2(1): 253–267. 2021.
bibtex

Unified lower bounds for interactive high-dimensional estimation under information constraints. Jayadev Acharya; Clément L. Canonne; Ziteng Sun; and Himanshu Tyagi. *CoRR*, abs/2010.06562. 2021.
bibtex

Information-constrained optimization: can adaptive processing of gradients help?. Jayadev Acharya; Clément L. Canonne; Prathamesh Mayekar; and Himanshu Tyagi. *CoRR*, abs/2104.00979. 2021.
bibtex

Optimal Rates for Nonparametric Density Estimation under Communication Constraints. Jayadev Acharya; Clément L. Canonne; Aditya Vikram Singh; and Himanshu Tyagi. *CoRR*, abs/2107.10078. 2021.
bibtex

Local Differential Privacy Is Equivalent to Contraction of E_γ-Divergence. Shahab Asoodeh; Maryam Aliakbarpour; and Flávio P. Calmon. *CoRR*, abs/2102.01258. 2021.
bibtex

2020 (12)

Interactive Inference under Information Constraints. Jayadev Acharya; Clément L. Canonne; Yuhan Liu; Ziteng Sun; and Himanshu Tyagi. *CoRR*, abs/2007.10976. 2020.
Paper bibtex

Fisher Information Under Local Differential Privacy. Leighton Pate Barnes; Wei-Ning Chen; and Ayfer Özgür. *IEEE J. Sel. Areas Inf. Theory*, 1(3): 645–659. 2020.
bibtex

Geometric Lower Bounds for Distributed Parameter Estimation under Communication Constraints. Yanjun Han; Ayfer Özgür; and Tsachy Weissman. *ArXiv e-prints*, abs/1802.08417v3. September 2020.
bibtex

Inference under information constraints I: Lower bounds from chi-square contraction. Jayadev Acharya; Clément L. Canonne; and Himanshu Tyagi. *IEEE Trans. Inform. Theory*, 66(12): 7835–7855. 2020. Preprint available at arXiv:abs/1812.11476.
Paper doi bibtex

Inference Under Information Constraints II: Communication Constraints and Shared Randomness. Jayadev Acharya; Clément L. Canonne; and Himanshu Tyagi. *IEEE Trans. Inf. Theory*, 66(12): 7856–7877. 2020.
bibtex

Domain Compression and its Application to Randomness-Optimal Distributed Goodness-of-Fit. Jayadev Acharya; Clément L. Canonne; Yanjun Han; Ziteng Sun; and Himanshu Tyagi. In *COLT*, volume 125, of *Proceedings of Machine Learning Research*, pages 3–40, 2020. PMLR
bibtex

Distributed Signal Detection under Communication Constraints. Jayadev Acharya; Clément L. Canonne; and Himanshu Tyagi. In *COLT*, volume 125, of *Proceedings of Machine Learning Research*, pages 41–63, 2020. PMLR
bibtex

Lecture notes on: Information-theoretic methods for high-dimensional statistics. Yihong Wu. 2020.
Paper bibtex

Lower bounds for learning distributions under communication constraints via fisher information. Leighton Pate Barnes; Yanjun Han; and Ayfer Özgür. *J. Mach. Learn. Res.*, 21: Paper No. 236, 30. 2020.
bibtex

Private Identity Testing for High-Dimensional Distributions. Clément L. Canonne; Gautam Kamath; Audra McMillan; Jonathan Ullman; and Lydia Zakynthinou. In *Advances in Neural Information Processing Systems 33*, 2020. Preprint available at arXiv:abs/1905.11947
bibtex

Locally private non-asymptotic testing of discrete distributions is faster using interactive mechanisms. Thomas Berrett; and Cristina Butucea. In *NeurIPS*, 2020.
bibtex

Local differential privacy: elbow effect in optimal density estimation and adaptation over Besov ellipsoids. Cristina Butucea; Amandine Dubois; Martin Kroll; and Adrien Saumard. *Bernoulli*, 26(3): 1727–1764. 2020.
Paper doi bibtex

2019 (5)

Locally Private Gaussian Estimation. Matthew Joseph; Janardhan Kulkarni; Jieming Mao; and Steven Z. Wu. In H. Wallach; H. Larochelle; A. Beygelzimer; F.; E. Fox; and R. Garnett., editor(s), *Advances in Neural Information Processing Systems 32*, pages 2984–2993. Curran Associates, Inc., 2019.
bibtex

Fisher Information for Distributed Estimation under a Blackboard Communication Protocol. Leighton P. Barnes; Yanjun Han; and Ayfer Özgür. In *ISIT*, pages 2704–2708, 2019. IEEE
bibtex

Lower Bounds for Locally Private Estimation via Communication Complexity. John Duchi; and Ryan Rogers. In Alina Beygelzimer; and Daniel Hsu., editor(s), *Proceedings of the Thirty-Second Conference on Learning Theory*, volume 99, of *Proceedings of Machine Learning Research*, pages 1161–1191, Phoenix, USA, June 2019. PMLR
bibtex

Hadamard Response: Estimating Distributions Privately, Efficiently, and with Little Communication. Jayadev Acharya; Ziteng Sun; and Huanyu Zhang. In Kamalika Chaudhuri; and Masashi Sugiyama., editor(s), *Proceedings of Machine Learning Research*, volume 89, pages 1120–1129, 16–18 Apr 2019. PMLR
Paper bibtex

Communication and Memory Efficient Testing of Discrete Distributions. Ilias Diakonikolas; Themis Gouleakis; Daniel M. Kane; and Sankeerth Rao. In *COLT*, volume 99, of *Proceedings of Machine Learning Research*, pages 1070–1106, 2019. PMLR
bibtex

2018 (4)

Geometric Lower Bounds for Distributed Parameter Estimation under Communication Constraints. Yanjun Han; Ayfer Özgür; and Tsachy Weissman. In *Proceedings of the 31st Conference on Learning Theory, COLT 2018*, volume 75, of *Proceedings of Machine Learning Research*, pages 3163–3188, 2018. PMLR The arXiv (v3) version from 2020 corrects some issues and includes more results.
bibtex

Distributed Statistical Estimation of High-Dimensional and Non-parametric Distributions. Yanjun Han; Pritam Mukherjee; Ayfer Özgür; and Tsachy Weissman. In *Proceedings of the 2018 IEEE International Symposium on Information Theory (ISIT'18)*, pages 506–510, 2018.
bibtex

Minimax optimal procedures for locally private estimation. John C. Duchi; Michael I. Jordan; and Martin J. Wainwright. *J. Amer. Statist. Assoc.*, 113(521): 182–201. 2018.
bibtex

Optimal schemes for discrete distribution estimation under locally differential privacy. Min Ye; and Alexander Barg. *IEEE Trans. Inform. Theory*, 64(8): 5662–5676. 2018.
Paper doi bibtex

2017 (1)

Information-theoretic lower bounds on Bayes risk in decentralized estimation. Aolin Xu; and Maxim Raginsky. *IEEE Transactions on Information Theory*, 63(3): 1580–1600. 2017.
bibtex

2016 (1)

Communication lower bounds for statistical estimation problems via a distributed data processing inequality. Mark Braverman; Ankit Garg; Tengyu Ma; Huy L. Nguyen; and David P. Woodruff. In *Symposium on Theory of Computing Conference, STOC'16*, pages 1011–1020, 2016. ACM
bibtex

2014 (2)

On Communication Cost of Distributed Statistical Estimation and Dimensionality. Ankit Garg; Tengyu Ma; and Huy L. Nguyen. In *Advances in Neural Information Processing Systems 27*, pages 2726–2734, 2014.
bibtex

Fundamental limits of online and distributed algorithms for statistical learning and estimation. Ohad Shamir. In *Advances in Neural Information Processing Systems 27*, pages 163–171, 2014.
bibtex

2013 (1)

Information-theoretic lower bounds for distributed statistical estimation with communication constraints. Yuchen Zhang; John Duchi; Michael I. Jordan; and Martin J. Wainwright. In *Advances in Neural Information Processing Systems 26*, pages 2328–2336, 2013.
bibtex

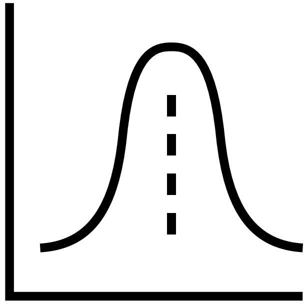
2009 (1)

Information-theoretic limits on sparsity recovery in the high-dimensional and noisy setting. Martin J. Wainwright. *IEEE Trans. Inform. Theory*, 55(12): 5728–5741. 2009.
Paper doi bibtex

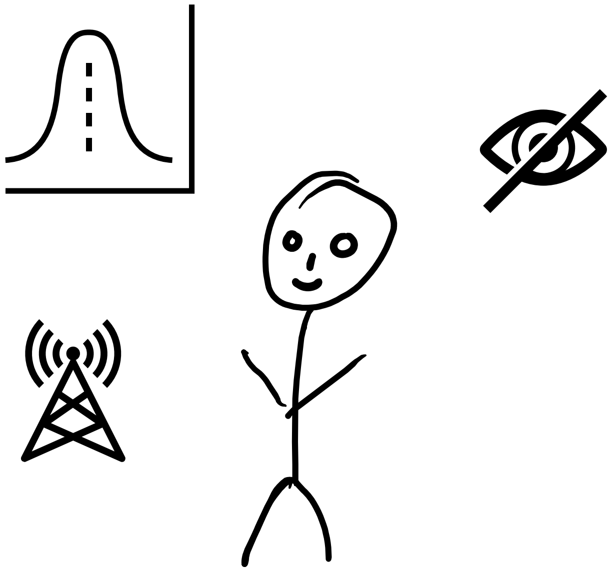
Bibliography available here

Recap: this lecture

1. Estimation for high-dimensional distributions: upper bounds ☑
    - Mean estimation under communication or local privacy (LDP) constraints ☑
2. Lower bounds ☑
    - A general bound for estimation (in the interactive setting) ☑
    - Application to communication and LDP ☑



The End



More: arXiv:2010.06562