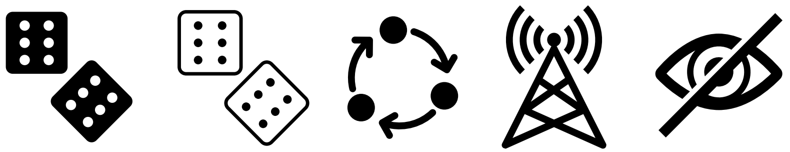
CSCIT 2021 - Lecture 2

Clément Canonne (University of Sydney)

Estimation and hypothesis testing under information constraints

Last lecture: recap

1. What are learning and testing?
2. Baseline: the "centralised" setting
3. Beyond the centralised setting: 3 flavours
   - Private-coin protocols
   - Public-coin protocols
   - Interactive protocols
4. What are information constraints?
   - Two guiding examples: **communication** and **privacy**



Contents of this lecture

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. Learning and testing discrete distributions: upper bounds
   - Learning, under communication or local privacy (LDP) constraints
   - Testing, under communication or LDP constraints

   ] Theorems + proof sketches

2. Lower bounds
   - A general bound for learning and testing
   - Application to communication and LDP

   ] detailed proof

Recall (1)

n iid samples $X_1, X_2, \ldots, X_n \sim p$, one per user

Learning: output $\hat{p}$ s.t. $\mathbb{E}_p[TV(\hat{p}, p)] \leq \varepsilon$

Testing: output $\hat{b} \in \{0,1\}$ s.t.

$$\mathbb{P}\{\hat{b}=1\}\mathbb{1}_{p=u} + \mathbb{P}\{\hat{b}=0\}\mathbb{1}_{TV(p,u)>\varepsilon} \leq \frac{1}{10}$$

Recall (1)

n iid samples $X_1, X_2, \ldots, X_n \sim p$, one per user

over $[d] = \{1, 2, \ldots, d\}$

Learning: output $\hat{p}$ s.t. $\mathbb{E}_p[TV(\hat{p}, p)] \leq \varepsilon$

Testing: output $\hat{b} \in \{0,1\}$ s.t.

$$\mathbb{P}\{\hat{b}=1\} \mathbb{1}_{p=u} + \mathbb{P}\{\hat{b}=0\} \mathbb{1}_{TV(p,u)>\varepsilon} \leq \frac{1}{10}$$

uniform over $[d]$

"uniformity testing"

Recall (2)

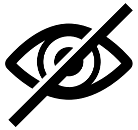
Communication

Each user can only send ℓ bits

$$\mathcal{W}_\ell = \{ W : \mathcal{X} \to \{0,1\}^\ell \}$$

Can't send too much

Local Privacy



Each user requires ϱ- differential privacy

$$\forall W \in \mathcal{W}_\varrho$$

$$\forall y, x, x', \quad W(y \mid x) \leq e^\varrho W(y \mid x')$$

Can't reveal too much

Upper bounds

$d \gg 1$
$\varepsilon \in (0,1]$
$\ell \leq \log_2 d$
$\rho \in (0,1]$

Theorem. Learning an arbitrary p over $[d]$ to TV loss ε under ℓ-bit communication constraints has sample complexity ______.

Theorem. Learning an arbitrary p over $[d]$ to TV loss ε under ρ-local privacy (LDP) constraints has sample complexity ______.

Upper bounds

$d \gg 1$
$\varepsilon \in (0,1]$
$\ell \leq \log_2 d$
$\rho \in (0,1]$

Theorem. Learning an arbitrary p over $[d]$ to TV loss ε under ℓ-bit communication constraints has sample complexity ______ .

Theorem. Learning an arbitrary p over $[d]$ to TV loss ε under ρ-local privacy (LDP) constraints has sample complexity ______ .

Upper bounds

$d \gg 1$
$\varepsilon \in (0,1]$
$\ell \le \log_2 d$
$\rho \in (0,1]$

Theorem. Learning an arbitrary p over $[d]$ to TV loss ε under ℓ-bit communication constraints has sample complexity $O\left(\frac{d^2}{2^\ell \varepsilon^2}\right)$.

Theorem. Learning an arbitrary p over $[d]$ to TV loss ε under ρ-local privacy (LDP) constraints has sample complexity $O\left(\frac{d^2}{\rho^2 \varepsilon^2}\right)$.

Upper bounds

$d \gg 1$
$\varepsilon \in (0,1]$
$\ell \leq \log_2 d$
$\rho \in (0,1]$

Theorem. Learning an arbitrary p over $[d]$ to TV loss ε under ℓ-bit communication constraints has sample complexity $O\left(\frac{d^2}{2^\ell \varepsilon^2}\right)$. Further, this is attained by a private-coin protocol.

Theorem. Learning an arbitrary p over $[d]$ to TV loss ε under ρ-local privacy (LDP) constraints has sample complexity $O\left(\frac{d^2}{\rho^2 \varepsilon^2}\right)$. Further, this is attained by a private-coin protocol.

Upper bounds

Recall: $\frac{d}{\varepsilon^2}$ in the centralised case.

$d \gg 1$
$\varepsilon \in (0,1]$
$\ell \leq \log_2 d$
$\rho \in (0,1]$

Theorem. Learning an arbitrary p over $[d]$ to TV loss ε under ℓ-bit communication constraints has sample complexity $O\left(\frac{d^2}{2^\ell \varepsilon^2}\right)$. Further, this is attained by a private-coin protocol.

Theorem. Learning an arbitrary p over $[d]$ to TV loss ε under ρ-local privacy (LDP) constraints has sample complexity $O\left(\frac{d^2}{\rho^2 \varepsilon^2}\right)$. Further, this is attained by a private-coin protocol.

What about testing?

d? $\sqrt{d}$? d^2? $d^{2/3}$?

$d^{3/4}$? $d^{3/2}$?

Upper bounds

$d \gg 1$
$\varepsilon \in (0,1]$
$\ell \leq \log_2 d$
$\rho \in (0,1]$

Theorem. Testing if an arbitrary p over $[d]$ is u or has $TV(p,u) > \varepsilon$ under ℓ-bit communication constraints has sample complexity ______.

Theorem. Testing if an arbitrary p over $[d]$ is u or has $TV(p,u) > \varepsilon$ under ρ-local privacy (LDP) constraints has sample complexity ______.

Upper bounds

$d \gg 1$
$\varepsilon \in (0,1]$
$\ell \leq \log_2 d$
$\rho \in (0,1]$

Theorem. Testing if an arbitrary p over $[d]$ is u or has $\mathrm{TV}(p,u) > \varepsilon$ under ℓ-bit communication constraints has sample complexity $O\left(\frac{d^{3/2}}{2^{\ell}\varepsilon^2}\right)$ (private-coin)

Theorem. Testing if an arbitrary p over $[d]$ is u or has $\mathrm{TV}(p,u) > \varepsilon$ under ρ-local privacy (LDP) constraints has sample complexity $O\left(\frac{d^{3/2}}{\rho^2\varepsilon^2}\right)$ (private-coin)

Upper bounds

$d \gg 1$
$\varepsilon \in (0,1]$
$\ell \leq \log_2 d$
$\rho \in (0,1]$

Theorem. Testing if an arbitrary p over $[d]$ is u or has $\mathrm{TV}(p,u) > \varepsilon$ under ℓ-bit communication constraints has sample complexity $O\left(\frac{d^{3/2}}{2^{\ell}\varepsilon^2}\right)$ (private-coin) and $O\left(\frac{d}{2^{\ell/2}\varepsilon^2}\right)$ (public-coin).

Theorem. Testing if an arbitrary p over $[d]$ is u or has $\mathrm{TV}(p,u) > \varepsilon$ under ρ-local privacy (LDP) constraints has sample complexity $O\left(\frac{d^{3/2}}{\rho^2\varepsilon^2}\right)$ (private-coin) and $O\left(\frac{d}{\rho^2\varepsilon^2}\right)$ (public-coin).

Upper bounds

Recall: $\frac{\sqrt{d}}{\varepsilon^2}$ in the centralised case.

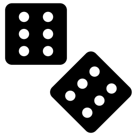
$d \gg 1$
$\varepsilon \in (0,1]$
$\ell \leq \log_2 d$
$\rho \in (0,1]$

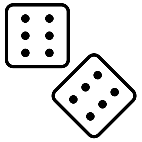
Theorem. Testing if an arbitrary p over $[d]$ is u or has $\mathrm{TV}(p,u) > \varepsilon$ under ℓ-bit communication constraints has sample complexity $O\left(\frac{d^{3/2}}{2^{\ell}\varepsilon^2}\right)$ (private-coin) and $O\left(\frac{d}{2^{\ell/2}\varepsilon^2}\right)$ (public-coin).

Theorem. Testing if an arbitrary p over $[d]$ is u or has $\mathrm{TV}(p,u) > \varepsilon$ under ρ-local privacy (LDP) constraints has sample complexity $O\left(\frac{d^{3/2}}{\rho^2\varepsilon^2}\right)$ (private-coin) and $O\left(\frac{d}{\rho^2\varepsilon^2}\right)$ (public-coin).

Upper bounds

Proof. If time allows.

1. "Simulate-and-Infer" 

2. "Domain Compression" 

General, useful primitives.

Lower bounds

Can we do better?

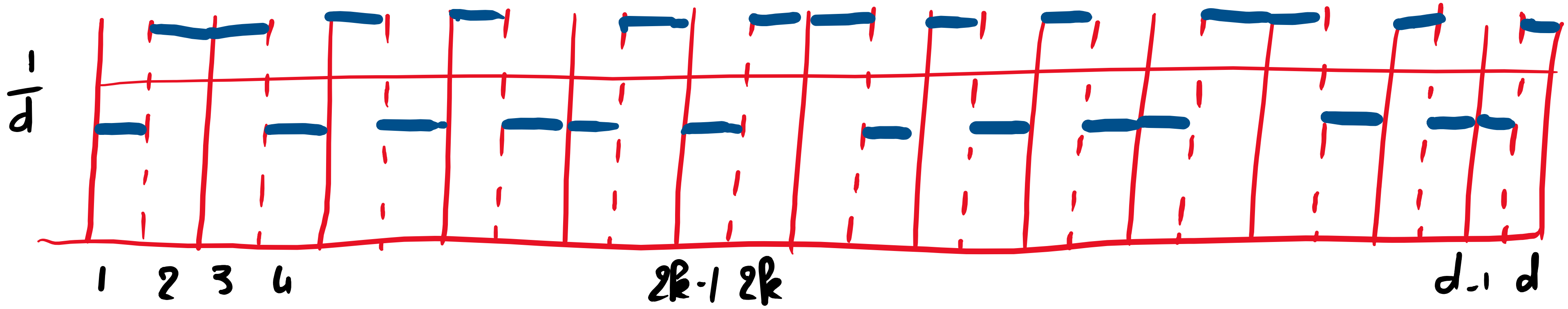
Lower bounds

Can we do better? No.

(But how to prove it?)

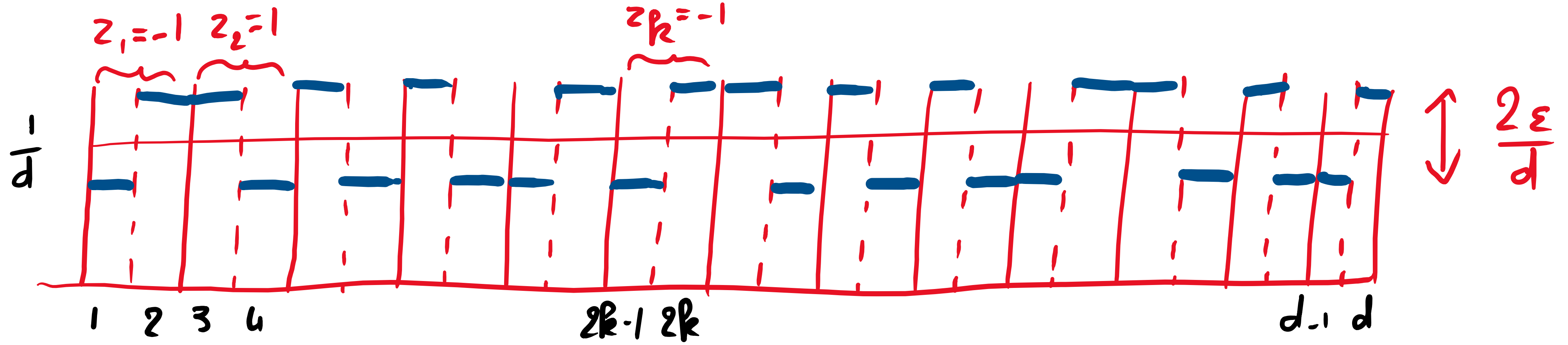
Let's start with a collection of hard instances $\mathcal{P} = \{p_z\}_{z \in \{\pm 1\}^{d/2}}$:

$$p_z = \frac{1}{d}\left(1+\varepsilon z_1, 1-\varepsilon z_1, 1+\varepsilon z_2, 1-\varepsilon z_2, \ldots, 1+\varepsilon z_{\frac{d}{2}}, 1-\varepsilon z_{\frac{d}{2}}\right)$$



Let's start with a collection of hard instances $\mathcal{P} = \{p_z\}_{z \in \{\pm 1\}^{d/2}}$:

$$p_z = \frac{1}{d}\Big(\underbrace{1+\varepsilon z_1, 1-\varepsilon z_1}_{\text{sum to } 2}, \underbrace{1+\varepsilon z_2, 1-\varepsilon z_2}_{\text{sum to } 2}, \ldots, \underbrace{1+\varepsilon z_{\frac{d}{2}}, 1-\varepsilon z_{\frac{d}{2}}}_{\text{sum to } 2}\Big)$$



Let's start with a collection of hard instances $\mathcal{P} = \{p_z\}_{z \in \{\pm 1\}^{d/2}}$:

$$p_z = \frac{1}{d}\left(1+2\varepsilon z_1,\ 1-2\varepsilon z_1,\ 1+2\varepsilon z_2,\ 1-2\varepsilon z_2,\ \ldots,\ 1+2\varepsilon z_{\frac{d}{2}},\ 1-2\varepsilon z_{\frac{d}{2}}\right)$$

Note that $\mathrm{TV}(p_z, u) = \varepsilon$, and $\mathrm{TV}(p_z, p_{z'}) = \frac{2\varepsilon}{d} \cdot \mathrm{Ham}(z, z')$

Let's start with a collection of hard instances $\mathcal{P} = \{p_z\}_{z \in \{\pm 1\}^{d/2}}$:

$$p_z = \frac{1}{d}\left(1+2\varepsilon z_1, 1-2\varepsilon z_1, 1+2\varepsilon z_2, 1-2\varepsilon z_2, \ldots, 1+2\varepsilon z_{\frac{d}{2}}, 1-2\varepsilon z_{\frac{d}{2}}\right)$$

Note that $\mathrm{TV}(p_z, u) = \varepsilon$, and $\mathrm{TV}(p_z, p_{z'}) = \frac{2\varepsilon}{d} \cdot \mathrm{Ham}(z, z')$

useful for testing

useful for learning

Fix $\mathcal{W}$ (constraints). For $W \in \mathcal{W}$, $W : [d] \to \mathcal{Y}$, $X \sim p$ induces a distribution on $\mathcal{Y}$:

$$p^W(y) = \mathbb{E}_{X \sim p}\left[W(y \mid X)\right] \qquad \forall y \in \mathcal{Y}$$

Fix $\mathcal{W}$ (constraints). For $W \in \mathcal{W}$, $W : [d] \to \mathcal{Y}$, $X \sim p$ induces a distribution on $\mathcal{Y}$:

$$p^W(y) = \mathbb{E}_{X \sim p}[W(y \mid X)] \qquad \forall y \in \mathcal{Y}$$

Fix any (interactive) protocol w/ n users under constraints $\mathcal{W}$, with message space $\mathcal{Y}$.

Inputs $X_1, \ldots, X_n \sim p$ (iid) $\longrightarrow$ induced distribution on $\mathcal{Y}^n$

(not a product distribution)

Fix $\mathcal{W}$ (constraints). For $W \in \mathcal{W}$, $W : [d] \to \mathcal{Y}$, $X \sim p$ induces a distribution on $\mathcal{Y}$:

$$p^W(y) = \mathbb{E}_{X \sim p}[W(y \mid X)] \qquad \forall y \in \mathcal{Y}$$

Fix any (interactive) protocol w/ n users under constraints $\mathcal{W}$, with message space $\mathcal{Y}$.

Inputs $X_1, \ldots, X_n \sim p$ (iid) $\longrightarrow$ induced distribution on $\mathcal{Y}^n$

p^{Y^n} depends on p, and the protocol (and thus $\mathcal{W}$)

We will take a uniform prior on Z: $Z_1, \ldots, Z_{d/2}$ iid. ± 1.

Our goal:

① Lower bound $\sum_{i=1}^{d/2} I(Z_i ; Y^n)$ for both learning and testing

We will take a uniform prior on Z: $Z_1, \ldots, Z_{d/2}$ iid. ± 1.

Our goal:

① Lower bound $\sum_{i=1}^{d/2} I(Z_i ; Y^n)$ for both learning and testing

note: not $I(Z; Y^n)$!

"Assouad-type bound"

Le Cam's method

We will take a uniform prior on Z: $Z_1, \dots, Z_{d/2}$ iid. ± 1.

Our goal:

① Lower bound $\sum_{i=1}^{d/2} I(Z_i ; Y^n)$ for both learning and testing

"Assouad-type bound" — Le Cam's method

② Upper bound $\sum_{i=1}^{d/2} I(Z_i ; Y^n)$ for both learning and testing

We will take a uniform prior on Z: $Z_1, \ldots, Z_{d/2}$ iid. ± 1.

Our goal:

① Lower bound $\sum_{i=1}^{d/2} I(Z_i; Y^n)$ for both learning and testing

"Assouad-type bound" — Le Cam's method

② Upper bound $\sum_{i=1}^{d/2} I(Z_i; Y^n)$ for both learning and testing

as a function of $n, \varepsilon, d, \mathcal{W}$

We will take a uniform prior on Z : $Z_1, \dots, Z_{d/2}$ iid. ± 1.

Our goal:

① Lower bound $\sum_{i=1}^{d/2} I(Z_i ; Y^n)$ for both learning and testing

"Assouad-type bound"

Le Cam's method

② Upper bound $\sum_{i=1}^{d/2} I(Z_i ; Y^n)$ for both learning and testing *

as a function of $n, \varepsilon, d, \mathcal{W}$

③ Put things together to get a LB on n.

Let's do first ① + ② + ③ for learning

(step ② will be reused for testing)

Step ①.

Learning: For Z uniform and Y^n transcript of learning protocol,

w/ accuracy $\frac{\varepsilon}{20}$, say

$$\frac{1}{d} \sum_{k=1}^{d/2} I(Z_k ; Y^n) = \Omega(1)$$

Step ①.

Learning: For Z uniform and Y^n transcript of learning protocol,

$$\frac{1}{d} \sum_{k=1}^{d/2} I(Z_k ; Y^n) = \Omega(1)$$

Proof. Given $\hat{p} = \hat{p}(Y^n)$, let $\hat{Z} := \operatorname{argmin}_z TV(p_z, \hat{p})$. Then

$$TV(p_{\hat{Z}}, p_Z) \leq TV(p_{\hat{Z}}, \hat{p}) + TV(\hat{p}, p_Z) \leq 2\,TV(\hat{p}, p_Z)$$

and, taking $\mathbb{E}$,

$$\frac{2\varepsilon}{d} \sum_{k=1}^{d/2} \mathbb{P}\{\hat{Z}_k \neq Z_k\} \leq 2\, \mathbb{E}[TV(\hat{p}, p_Z)] \leq 2 \cdot \frac{\varepsilon}{20}$$

Step ①.

Learning: For Z uniform and Y^n transcript of learning protocol,

$$\frac{1}{d}\sum_{k=1}^{d/2} I(Z_k; Y^n) = \Omega(1)$$

Proof. Given $\hat{p} = \hat{p}(Y^n)$, let $\hat{Z} := \operatorname{argmin}_z \mathrm{TV}(p_z, \hat{p})$. Then

$$\frac{2\varepsilon}{d}\mathrm{Ham}(\hat{Z}, Z) \to \mathrm{TV}(p_{\hat{Z}}, p_Z) \leq \mathrm{TV}(p_{\hat{Z}}, \hat{p}) + \mathrm{TV}(\hat{p}, p_Z) \leq 2\,\mathrm{TV}(\hat{p}, p_Z)$$

and, taking $\mathbb{E}$,

$$\frac{2\varepsilon}{d}\sum_{k=1}^{d/2} \mathbb{P}\{\hat{Z}_k \neq Z_k\} \leq 2\,\mathbb{E}[\mathrm{TV}(\hat{p}, p_Z)] \underset{\text{learning protocol}}{\leq} 2 \cdot \frac{\varepsilon}{20}$$

Step ①.

Learning: For Z uniform and Y^n transcript of learning protocol,

$$\frac{1}{d}\sum_{k=1}^{d/2} I(Z_k;Y^n) = \Omega(1)$$

Proof. So $\frac{1}{d}\sum_k \mathbb{P}\{\hat{Z}_k \neq Z_k\} \leq \frac{1}{10}$. Now, $Z_k - Y^n - \hat{Z}_k$, so

$$I(Z_k;Y^n) \geq I(Z_k;\hat{Z}_k) \geq 1 - h(\mathbb{P}\{Z_k \neq \hat{Z}_k\})$$

Step ①.

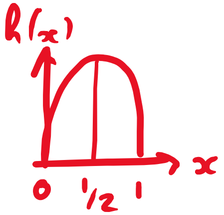
Learning: For Z uniform and Y^n transcript of learning protocol,

$$\frac{1}{d}\sum_{k=1}^{d/2} I(Z_k;Y^n) = \Omega(1)$$

Proof. So $\frac{1}{d}\sum_k \mathbb{P}\{\hat{Z}_k \neq Z_k\} \leq \frac{1}{10}$. Now, $Z_k - Y^n - \hat{Z}_k$, so

$$I(Z_k;Y^n) \overset{\text{DPI}}{\geq} I(Z_k;\hat{Z}_k) \overset{\text{Fano}}{\geq} 1 - h(\mathbb{P}\{Z_k \neq \hat{Z}_k\})$$

binary entropy



Step ①.

Learning: For Z uniform and Y^n transcript of learning protocol,

$$\frac{1}{d}\sum_{k=1}^{d/2} I(Z_k;Y^n) = \Omega(1)$$

Proof. So $\frac{2}{d}\sum_k \mathbb{P}\{\hat{Z}_k \neq Z_k\} \leq \frac{1}{5}$ ⊛. Now, $Z_k — Y^n — \hat{Z}_k$, so

$$I(Z_k;Y^n) \geq I(Z_k;\hat{Z}_k) \geq 1 - h(\mathbb{P}\{Z_k \neq \hat{Z}_k\})$$

and so

concavity

$$\frac{2}{d}\sum_{k=1}^{d/2} I(Z_k;Y^n) \geq 1 - \frac{2}{d}\sum_k h(\mathbb{P}\{Z_k \neq \hat{Z}_k\}) \geq 1 - h\left(\frac{2}{d}\sum_k \mathbb{P}\{Z_k \neq \hat{Z}_k\}\right) \overset{⊛}{\geq} 1 - h(1/5) \approx 0.3$$

$\square$

Step ② For $1 \leq i \leq \frac{d}{2}$, consider the partial mixtures

$$P_{+i}^{Y^n} := \mathbb{E}_Z\left[P_Z^{Y^n} \mid Z_i = +1\right] = \frac{2}{2^{d/2}} \sum_{z : z_i = 1} P_z^{Y^n}$$

(same for $P_{-i}^{Y^n}$)

Step ② For $1 \leq i \leq \frac{d}{2}$, consider the partial mixtures

$$p_{+i}^{Y^n} := \mathbb{E}_z\left[p_z^{Y^n} \mid z_i = +1\right] = \frac{2}{2^{d/2}} \sum_{z : z_i = 1} p_z^{Y^n}$$

(same for $p_{-i}^{Y^n}$)

and let $$q^{Y^n} := \mathbb{E}_z\left[p_z^{Y^n}\right] = \frac{1}{2}\left(p_{+i}^{Y^n} + p_{-i}^{Y^n}\right)$$

Step ② For $1 \leq i \leq \frac{d}{2}$, consider the partial mixtures

$$P_{+i}^{Y^n} := \mathbb{E}_Z\left[P_Z^{Y^n} \mid Z_i = +1\right] = \frac{2}{2^{d/2}} \sum_{Z : Z_i = 1} P_Z^{Y^n}$$

(same for $P_{-i}^{Y^n}$)

and let $q^{Y^n} := \mathbb{E}_Z\left[P_Z^{Y^n}\right] = \frac{1}{2}\left(P_{+i}^{Y^n} + P_{-i}^{Y^n}\right)$

Then

$$I(Z_i; Y^n) = \frac{1}{2}\left(KL\left(P_{+i}^{Y^n} \| q^{Y^n}\right) + KL\left(P_{-i}^{Y^n} \| q^{Y^n}\right)\right)$$

$$\leq \frac{1}{4}\left(KL\left(P_{+i}^{Y^n} \| P_{-i}^{Y^n}\right) + KL\left(P_{-i}^{Y^n} \| P_{+i}^{Y^n}\right)\right)$$

$$\leq \frac{1}{4}\left(\mathbb{E}\left[KL\left(P_Z^{Y^n} \| P_{Z \oplus i}^{Y^n}\right) \mid Z_i = +1\right] + \mathbb{E}\left[KL\left(P_Z^{Y^n} \| P_{Z \oplus i}^{Y^n}\right) \mid Z_i = -1\right]\right)$$

Step ② For $1 \le i \le \frac{d}{2}$, consider the partial mixtures

$$P_{+i}^{Y^n} := \mathbb{E}_Z\left[P_Z^{Y^n} \mid Z_i = +1\right] = \frac{2}{2^{d/2}} \sum_{z : z_i = 1} P_z^{Y^n}$$

(same for $P_{-i}^{Y^n}$)

and let $q^{Y^n} := \mathbb{E}_Z\left[P_Z^{Y^n}\right] = \frac{1}{2}\left(P_{+i} + P_{-i}\right)$

Then

$$I(Z_i ; Y^n) = \frac{1}{2}\left(KL\left(P_{+i}^{Y^n} \| q^{Y^n}\right) + KL\left(P_{-i}^{Y^n} \| q^{Y^n}\right)\right)$$

def^n: $I(X;Y) = \mathbb{E}_X\left[KL(P_{Y|X} \| P_Y)\right]$

$$\le \frac{1}{4}\left(KL\left(P_{+i}^{Y^n} \| P_{-i}^{Y^n}\right) + KL\left(P_{-i}^{Y^n} \| P_{+i}^{Y^n}\right)\right)$$

← joint convexity

$$\le \frac{1}{4}\left(\mathbb{E}\left[KL\left(P_Z^{Y^n} \| P_{Z^{\oplus i}}^{Y^n}\right) \mid Z_i = +1\right] + \mathbb{E}\left[KL\left(P_Z^{Y^n} \| P_{Z^{\oplus i}}^{Y^n}\right) \mid Z_i = -1\right]\right)$$

$Z^{\oplus i} = Z$ with i^{th} bit flipped

Step ② For $1 \leq i \leq \frac{d}{2}$,

$Z^{\oplus i} = Z$ with i^{th} bit flipped

$$I(Z_i ; Y^n) \leq \frac{1}{2} \mathbb{E}_Z \left[KL\left(P_Z^{Y^n} \| P_{Z^{\oplus i}}^{Y^n} \right) \right]$$

$$= \frac{1}{2} \mathbb{E}_Z \left[\sum_{t=1}^{n} \mathbb{E}_{P_Z^{Y^{t-1}}} \left[KL\left(P_Z^{Y^t | Y^{t-1}} \| P_{Z^{\oplus i}}^{Y^t | Y^{t-1}} \right) \right] \right]$$

Step ② For $1 \le i \le \frac{d}{2}$,

$Z^{\oplus i} = Z$ with i^{th} bit flipped

$$I(Z_i ; Y^n) \leq \frac{1}{2} \mathop{\mathbb{E}}_{Z}\left[KL\left(P_Z^{Y^n} \,\|\, P_{Z^{\oplus i}}^{Y^n}\right)\right]$$

$$= \frac{1}{2} \mathop{\mathbb{E}}_{Z}\left[\sum_{t=1}^{n} \mathop{\mathbb{E}}_{P_Z^{Y^{t-1}}} \left[KL\left(P_Z^{Y^t | Y^{t-1}} \,\|\, P_{Z^{\oplus i}}^{Y^t | Y^{t-1}}\right)\right]\right]$$

no dependence on i

Chain rule for KL

Step 2   For $1 \le i \le \frac{d}{2}$,

$Z^{\oplus i} = Z$ with i^{th} bit flipped

$$I(Z_i ; Y^n) \le \frac{1}{2} \mathbb{E}_Z \left[KL\left(P_Z^{Y^n} \| P_{Z^{\oplus i}}^{Y^n}\right)\right]$$

$$= \frac{1}{2} \mathbb{E}_Z \left[\sum_{t=1}^{n} \mathbb{E}_{P_Z^{Y^{t-1}}} \left[KL\left(P_Z^{Y^t | Y^{t-1}} \| P_{Z^{\oplus i}}^{Y^t | Y^{t-1}}\right)\right]\right]$$

$$\le \frac{1}{2} \sum_{t=1}^{n} \mathbb{E}_Z \mathbb{E}_{P_Z^{Y^{t-1}}} \left[\chi^2\left(P_Z^{Y^t | Y^{t-1}} \| P_{Z^{\oplus i}}^{Y^t | Y^{t-1}}\right)\right] \quad (KL \le \chi^2)$$

Step ② For $1 \leq i \leq \frac{d}{2}$,

$Z^{\oplus i} = Z$ with i^{th} bit flipped

$$
\begin{aligned}
I(Z_i ; Y^n) &\leq \frac{1}{2} \mathop{\mathbb{E}}_{Z}\left[KL\left(p_Z^{Y^n} \,\|\, p_{Z^{\oplus i}}^{Y^n}\right)\right] \\
&= \frac{1}{2} \mathop{\mathbb{E}}_{Z}\left[\sum_{t=1}^{n} \mathop{\mathbb{E}}_{Y^{t-1} \sim p_Z} \left[KL\left(p_Z^{Y^t | Y^{t-1}} \,\|\, p_{Z^{\oplus i}}^{Y^t | Y^{t-1}}\right)\right]\right] \\
&\leq \frac{1}{2} \sum_{t=1}^{n} \mathop{\mathbb{E}}_{Z} \mathop{\mathbb{E}}_{Y^{t-1} \sim p_Z} \left[\chi^2\left(p_Z^{Y^t | Y^{t-1}} \,\|\, p_{Z^{\oplus i}}^{Y^t | Y^{t-1}}\right)\right] \qquad (KL \leq \chi^2) \\
&= \frac{1}{2} \sum_{t=1}^{n} \mathop{\mathbb{E}}_{Z} \mathop{\mathbb{E}}_{Y^{t-1} \sim p_Z} \left[\sum_{y} \frac{\left(\mathbb{P}_{p_Z}[Y_t = y \mid Y^{t-1}] - \mathbb{P}_{p_{Z^{\oplus i}}}[Y_t = y \mid Y^{t-1}]\right)^2}{\mathbb{P}_{p_{Z^{\oplus i}}}[Y_t = y \mid Y^{t-1}]}\right]
\end{aligned}
$$

So... what now?

Key observation: $\forall y$,

$$\mathbb{P}_{P_z}\left[Y_t = y \mid Y^{t-1}\right] = \mathbb{P}_{P_{z\oplus i}}\left[Y_t = y \mid Y^{t-1}\right] + \frac{4\varepsilon}{d} z_i \left(W^{Y^{t-1}}(y \mid 2i-1) - W^{Y^{t-1}}(y \mid 2i)\right)$$

Follows from our construct° + expression of P_z^W

Key observation: $\forall y$,

$$\mathbb{P}_{P_Z}\left[Y_t = y \mid Y^{t-1}\right] = \mathbb{P}_{P_{Z \oplus i}}\left[Y_t = y \mid Y^{t-1}\right] + \frac{4\varepsilon}{d} Z_i \left(W^{Y^{t-1}}(y|2i-1) - W^{Y^{t-1}}(y|2i)\right)$$

Follows from our construct° + expression of P_Z^W

Using this,

$$I(Z_i ; Y^n) \leq \frac{cst\, \varepsilon^2}{d} \sum_{t=1}^{n} \mathbb{E}_Z \mathbb{E}_{Y^{t-1} \sim P_Z} \sum_y \frac{\left(W^{Y^{t-1}}(y|2i-1) - W^{Y^{t-1}}(y|2i)\right)^2}{\sum_x W^{Y^{t-1}}(y|x)}$$

also using $\mathbb{P}_{P_{Z \oplus i}}\left[Y_t = y \mid Y^{t-1}\right] \geq \frac{1-2\varepsilon}{d} \sum_x W^{Y^{t-1}}(y|x)$ for the denominator.

Define, for $W \in \mathcal{W}$, the matrix $H(W)$ by

$$H(W)_{ij} = \sum_y \frac{\big(W(y|2i-1) - W(y|2i)\big)\big(W(y|2j-1) - W(y|2j)\big)}{\sum_x' W(y|x)} \qquad i,j \in [d/2]$$

$$I(Z_i; Y^n) \leq \mathrm{cst}\,\frac{\varepsilon^2}{d} \sum_{t=1}^{n} \mathbb{E}_Z\, \mathbb{E}_{P_Z^{Y^{t-1}}} \sum_y \frac{\left(W^{Y^{t-1}}(y|2i-1) - W^{Y^{t-1}}(y|2i)\right)^2}{\sum_x W^{Y^{t-1}}(y|x)}$$

Define, for $W \in \mathcal{W}$, the matrix $H(W)$ by

$$H(W)_{ij} = \sum_y \frac{\left(W(y|2i-1) - W(y|2i)\right)\left(W(y|2j-1) - W(y|2j)\right)}{\sum_x' W(y|x)} \qquad i,j \in [d/2]$$

$$\sum_{i=1}^{d/2} I(Z_i ; Y^n) \leq \mathrm{cst}\,\frac{\varepsilon^2}{d} \sum_{t=1}^{n} \mathbb{E}_Z\, \mathbb{E}_{Y^{t-1} \sim p_Z} \sum_{i=1}^{d/2} \sum_y \frac{\left(W^{Y^{t-1}}(y|2i-1) - W^{Y^{t-1}}(y|2i)\right)^2}{\sum_x W^{Y^{t-1}}(y|x)}$$

Define, for $W \in \mathcal{W}$, the matrix $H(W)$ by

$$H(W)_{ij} = \sum_y \frac{\big(W(y|2i-1) - W(y|2i)\big)\big(W(y|2j-1) - W(y|2j)\big)}{\sum_x W(y|x)} \qquad i,j \in [d/2]$$

$$\sum_{i=1}^{d/2} I(Z_i ; Y^n) \leq \frac{cst\, \varepsilon^2}{d} \sum_{t=1}^{n} \mathbb{E}_Z \mathbb{E}_{Y^{t-1} \sim P_Z} \mathrm{Tr}\left[H(W^{t-1})\right]$$

Define, for $W \in \mathcal{W}$, the matrix $H(W)$ by

$$H(W)_{ij} = \sum_y \frac{\big(W(y|2i-1) - W(y|2i)\big)\big(W(y|2j-1) - W(y|2j)\big)}{\sum_x' W(y|x)} \qquad i,j \in [d/2]$$

$$\sum_{i=1}^{d/2} I(Z_i ; Y^n) \leq \frac{cst\, \varepsilon^2}{d} \sum_{t=1}^{n} \mathbb{E}_Z \, \mathbb{E}_{Y^{t-1} \sim P_Z} \operatorname{Tr}\left[H(W^{t-1})\right]$$

$$\leq \frac{cst\, \varepsilon^2}{d} \sum_{t=1}^{n} \sup_{W \in \mathcal{W}} \operatorname{Tr}\left[H(W)\right]$$

Define, for $W \in \mathcal{W}$, the matrix $H(W)$ by

$$H(W)_{ij} = \sum_y \frac{(W(y|2i-1) - W(y|2i))(W(y|2j-1) - W(y|2j))}{\sum_x' W(y|x)} \qquad i,j \in [d/2]$$

$$\sum_{i=1}^{d/2} I(Z_i ; Y^n) \leq \frac{cst\, \varepsilon^2}{d} \sum_{t=1}^{n} \mathbb{E}_Z\, \mathbb{E}_{Y^{t-1} \sim P_Z} \mathrm{Tr}[H(W^{t-1})]$$

$$\leq \frac{cst\, \varepsilon^2}{d}\, n . \sup_{W \in \mathcal{W}} \mathrm{Tr}[H(W)]$$

Define, for $W \in \mathcal{W}$, the matrix $H(W)$ by

$$H(W)_{ij} = \sum_y \frac{\big(W(y|2i-1) - W(y|2i)\big)\big(W(y|2j-1) - W(y|2j)\big)}{\sum_x' W(y|x)} \qquad i,j \in [d/2]$$

Step 2

$$\frac{\Omega(1)}{d} \sum_{i=1}^{d/2} I(Z_i ; Y^n) \leq \frac{n\varepsilon^2}{d^2} \sup_{W \in \mathcal{W}} \operatorname{Tr}[H(W)]$$

Define, for $W \in \mathcal{W}$, the matrix $H(W)$ by

$$H(W)_{ij} = \sum_y \frac{(W(y|2i-1) - W(y|2i))(W(y|2j-1) - W(y|2j))}{\sum_x' W(y|x)} \qquad i,j \in [d/2]$$

Step 2

$$\frac{\Omega(1)}{d} \sum_{i=1}^{d/2} I(Z_i ; Y^t) \leq \frac{t\,\varepsilon^2}{d^2} \sup_{W \in \mathcal{W}} \mathrm{Tr}[H(W)] \qquad \left(\text{useful for testing}\right)$$

Learning

Define, for $W \in \mathcal{W}$, the matrix $H(W)$ by

$$H(W)_{ij} = \sum_y \frac{\left(W(y|2i-1) - W(y|2i)\right)\left(W(y|2j-1) - W(y|2j)\right)}{\sum_x W(y|x)} \qquad i,j \in [d/2]$$

Step ③ $\forall\, 1 \leq t \leq n$,

$$\Omega(1) \leq \frac{t\,\varepsilon^2}{d} \sup_{W \in \mathcal{W}} \operatorname{Tr}[H(W)]$$

In particular, for $t = n$

$$\boxed{n = \Omega\left(\frac{d^2}{\varepsilon^2 \sup_{W \in \mathcal{W}} \operatorname{Tr}[H(W)]}\right)}$$

What about testing?

Step ①: Le Cam.

$$\Omega(1) \leq \mathrm{TV}\left(\mathbb{E}_z\left[P_z^{Y^n}\right], P_u^{Y^n}\right)^2$$

What about testing ?

Step ①: Le Cam.

(Pinsker)

$$\Omega(1) \leq TV\left(\mathbb{E}_z\left[P_z^{Y^n}\right], P_u^{Y^n}\right)^2 \leq KL\left(\mathbb{E}_z\left[P_z^{Y^n}\right] \| P_u^{Y^n}\right)$$

What about testing?

Step ①: Le Cam.

$$\Omega(1) \leq TV\left(\mathbb{E}_{z}[P_z^{Y^n}], u^{Y^n}\right)^2 \leq KL\left(\mathbb{E}_{z}[P_z^{Y^n}] \,\|\, u^{Y^n}\right) \quad \text{(Pinsker)}$$

q^{Y^n}

$$\leq \sum_{t=1}^{n} \mathbb{E}_{q^{Y^{t-1}}}\left[KL\left(q^{Y_t|Y^{t-1}} \,\|\, u^{Y_t|Y^{t-1}}\right)\right] \quad \text{(chain rule)}$$

What about testing?

Step ①: Le Cam.

$$\Omega(1) \leq TV\left(\mathbb{E}_Z[p_Z^{Y^n}], u^{Y^n}\right)^2 \overset{\text{(Pinsker)}}{\leq} KL\left(\underbrace{\mathbb{E}_Z[p_Z^{Y^n}]}_{q^{Y^n}} \,\|\, u^{Y^n}\right)$$

$$\leq \sum_{t=1}^{n} \mathbb{E}_{q^{Y^{t-1}}}\left[KL\left(q^{Y_t|Y^{t-1}} \,\|\, u^{Y_t|Y^{t-1}}\right)\right] \qquad \text{(chain rule)}$$

$$\leq \sum_{t=1}^{n} \frac{cst.\,\varepsilon^2}{d} \sup_{W\in\mathcal{W}} \|H(W)\|_{op} \cdot \sum_{i=1}^{d/2} I(Z_i; Y^t) \qquad \text{(key lemma)}$$

What about testing?

Step ①: Le Cam.

$$\Omega(1) \leq \sum_{t=1}^{n} \frac{\text{cst.}\,\varepsilon^2}{d} \sup_{W\in\mathcal{W}} \|H(W)\|_{op} \cdot \sum_{i=1}^{d/2} I(Z_i;Y^t)$$ (key lemma)

$$\leq \text{cst.}\,\frac{\varepsilon^2}{d} \sup_{W\in\mathcal{W}} \|H(W)\|_{op} \sum_{t=1}^{n} \frac{t\,\varepsilon^2}{d} \sup_{W\in\mathcal{W}} \mathrm{Tr}[H(W)]$$ (we just proved it!) Step ②

$$\leq \text{cst.}\,\frac{\varepsilon^4 n^2}{d^2} \sup_{W\in\mathcal{W}} \|H(W)\|_{op} \sup_{W\in\mathcal{W}} \mathrm{Tr}[H(W)]$$

What about testing?

Step ①: Le Cam.

$$\Omega(1) \leq \sum_{t=1}^{n} \frac{\text{cst.}\,\varepsilon^2}{d} \sup_{W\in\mathcal{W}} \|H(W)\|_{op} \cdot \sum_{i=1}^{d/2} I(Z_i;Y^n) \quad \text{(key lemma)}$$

$$\leq \text{cst.}\,\frac{\varepsilon^2}{d} \sup_{W\in\mathcal{W}} \|H(W)\|_{op} \sum_{t=1}^{n} \frac{t\varepsilon^2}{d} \sup_{W\in\mathcal{W}} \mathrm{Tr}[H(W)] \quad \text{(Step ②)}$$

$$\leq \text{cst.}\,\frac{\varepsilon^4 n^2}{d^2} \sup_{W\in\mathcal{W}} \|H(W)\|_{op} \sup_{W\in\mathcal{W}} \mathrm{Tr}[H(W)]$$

let us call this $\|H(\mathcal{W})\|_{op}$ $\qquad$ $\|H(\mathcal{W})\|_{*}$

What did we show?

For interactive protocols under constraint $\mathcal{W}$

to each $W \in \mathcal{W}$ corresponds a psd matrix $H(W)$

Learning: $n = \Omega\left(\dfrac{d^2}{\varepsilon^2 \|H(\mathcal{W})\|_*}\right)$

Testing: $n = \Omega\left(\dfrac{d}{\varepsilon^2 \sqrt{\|H(\mathcal{W})\|_* \|H(\mathcal{W})\|_{op}}}\right)$

where $\|H(\mathcal{W})\| := \sup_{W \in \mathcal{W}} \|H(W)\|$

What about the $\Omega(k^{3/2})$ private-coin lower bound?

Are interactive and public-coin the same?

Let's start with a collection of hard instances $\mathcal{P} = \{p_z\}_{z \in [-1,1]^{d/2}}$:

$$p_z = \frac{1}{d}\left(1+c\varepsilon z_1,\ 1-c\varepsilon z_1,\ 1+c\varepsilon z_2,\ 1-c\varepsilon z_2,\ \ldots,\ 1+c\varepsilon z_{\frac{d}{2}},\ 1-c\varepsilon z_{\frac{d}{2}}\right)$$

(for some cst $c>0$) along with a prior ζ on $[-1,1]^{d/2}$.

Want: $\mathbb{P}_{z\sim\zeta}\{\mathrm{TV}(p_z,u)>\varepsilon\} \geq \Omega(1)$.

Let's start with a collection of hard instances $\mathcal{P} = \{p_z\}_{z \in [-1,1]^{d/2}}$:

$$p_z = \frac{1}{d}\left(1+c\varepsilon z_1, 1-c\varepsilon z_1, 1+c\varepsilon z_2, 1-c\varepsilon z_2, \ldots, 1+c\varepsilon z_{\frac{d}{2}}, 1-c\varepsilon z_{\frac{d}{2}}\right)$$

(for some cst $c>0$) along with a prior ζ on $[-1,1]^{d/2}$.

Want: $\mathbb{P}_{Z\sim\zeta}\{TV(p_Z, u) > \varepsilon\} \gtrsim \Omega(1)$.

For instance, Z u.a.r. on $\{\pm 1\}^{d/2}$.

Testing

Long story short: get

$$n = \Omega\left(\frac{d^{3/2}}{\varepsilon^2 \|H(\lambda)\|_*}\right)$$

for private-coin; and

$$n = \Omega\left(\frac{d}{\varepsilon^2 \|H(\lambda)\|_F}\right)$$

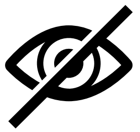
for public-coin.

Hölder : $\|H(w)\|_F^2 \leq \|H(w)\|_{op} \, \|H(w)\|_*$

$\ell_2^2 \qquad \ell_\infty \qquad \ell_1$

More details, discussion, full proofs :

- 📝 ***Inference under Information Constraints I: Lower Bounds from Chi-Square Contraction.*** Jayadev Acharya, Clément L. Canonne, and Himanshu Tyagi (IEEE Trans. Inf. Theory, 2020). arXiv:1812.11476
- 📝 ***Interactive Inference under Information Constraints.*** Jayadev Acharya, Clément L. Canonne, Yuhan Liu, Ziteng Sun, and Himanshu Tyagi (ISIT, 2021). arXiv:2007.10976



To conclude: what about communication and privacy, again?

To conclude: what about communication and privacy, again?

Easy exercise:

- LDP $$\|H(\mathcal{W}_e)\|_F \asymp \|H(\mathcal{W}_e)\|_* \asymp \|H(\mathcal{W}_e)\|_{op} \asymp \varepsilon^2$$

- Communication $$\|H(\mathcal{W}_e)\|_F^2 \asymp \|H(\mathcal{W}_e)\|_* \asymp 2^\ell \qquad \|H(\mathcal{W}_e)\|_{op} \asymp 1$$

Immediately proves the LBs!

Testing

private-coin $\quad n = \Omega\left(\dfrac{d^{3/2}}{\varepsilon^2 \|H(w)\|_*}\right)$

public-coin $\quad n = \Omega\left(\dfrac{d}{\varepsilon^2 \|H(w)\|_F}\right)$

interactive $\quad n = \Omega\left(\dfrac{d}{\varepsilon^2 \sqrt{\|H(w)\|_* \|H(w)\|_{op}}}\right)$

Testing

private-coin $\quad n = \Omega\left(\dfrac{d^{3/2}}{\varepsilon^2 \|H(w)\|_*}\right)$ $\quad 2^\ell$ or ℓ^2

public-coin $\quad n = \Omega\left(\dfrac{d}{\varepsilon^2 \|H(w)\|_F}\right)$ $\quad \sqrt{2^\ell}$ or ℓ^2

interactive $\quad n = \Omega\left(\dfrac{d}{\varepsilon^2 \sqrt{\|H(w)\|_* \|H(w)\|_{op}}}\right)$ $\quad \sqrt{2^\ell \cdot 1}$ or $\sqrt{\ell^2 \cdot \ell^2}$

Recap: this lecture

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 ☑

Next lecture:

Learning high-dimensional distributions under

those information constraints