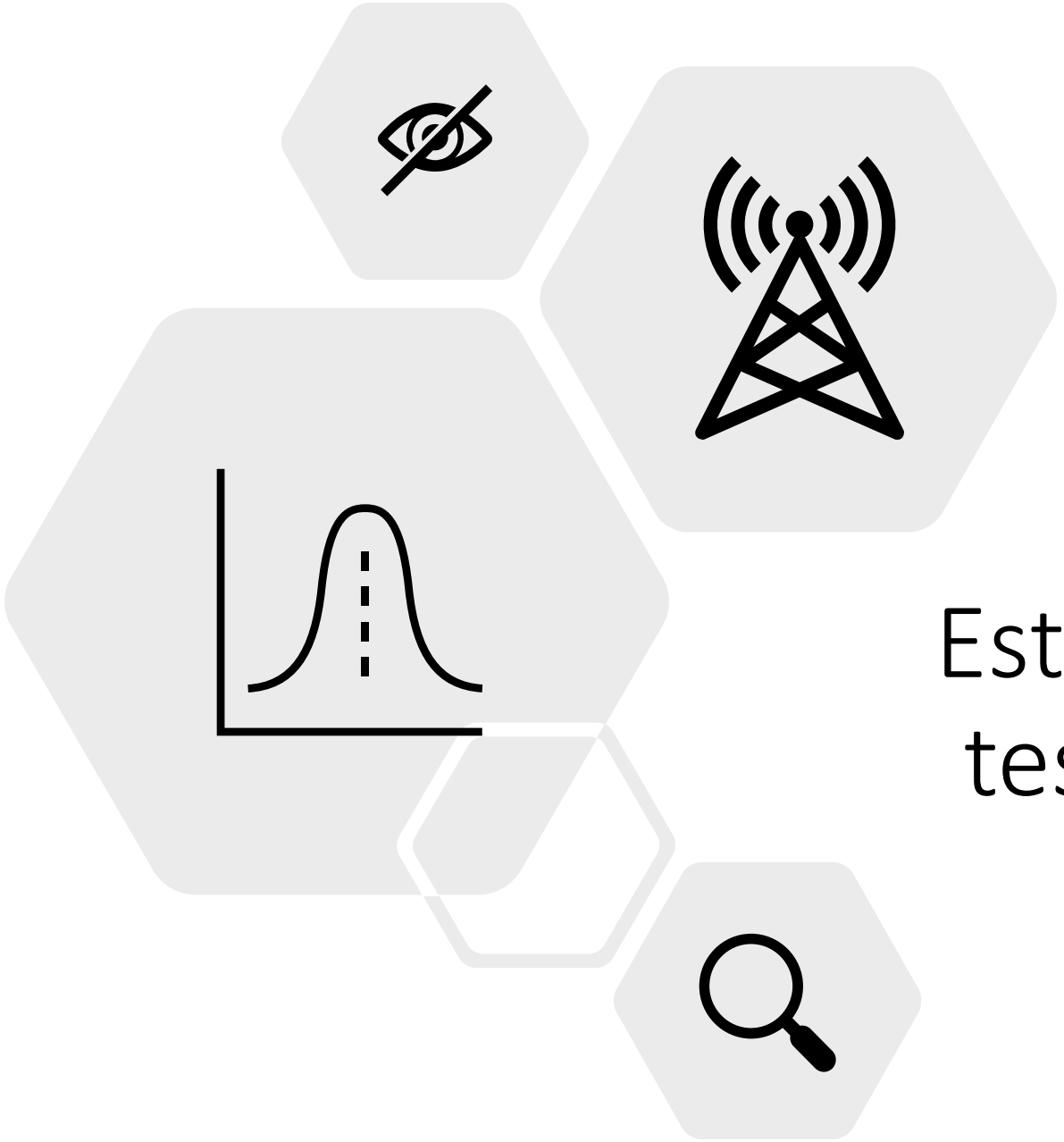


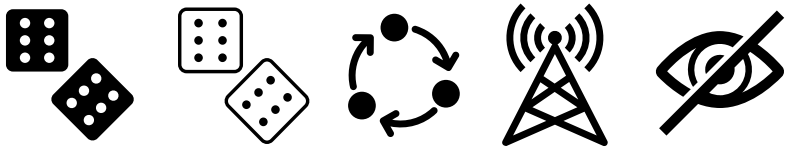
CSCIT 2021 - Lecture 2

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 $X_1, X_2, \dots, X_n \sim p$, one per user

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

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

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

Recall (1)

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

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

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

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

"uniformity testing"

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

u: uniform over $[d]$

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 ε- differential privacy

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

$$\forall y, x, x', \quad W(y|x) \leq e^\varepsilon W(y|x')$$

Can't reveal too much

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

Upper bounds

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 ———— .

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

Upper bounds

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 ———— .

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

Upper bounds

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

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

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

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

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

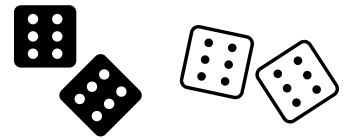
Upper bounds

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

Theorem. Testing if an arbitrary p over $[d]$ is u or has $TV(p,u) > \varepsilon$ 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 $TV(p,u) > \varepsilon$ 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 $TV(p,u) > \varepsilon$ 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 $TV(p,u) > \varepsilon$ 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.

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

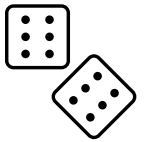
Theorem. Testing if an arbitrary p over $[d]$ is u or has $TV(p,u) > \varepsilon$ under ℓ-bit communication constraints has sample complexity $O\left(\frac{d^{3/2}}{2^\ell \varepsilon^2}\right)$ (private-coin) and $O\left(\frac{d}{2^{\ell/2}\varepsilon^2}\right)$ (public-coin).

Theorem. Testing if an arbitrary p over $[d]$ is u or has $TV(p,u) > \varepsilon$ 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?

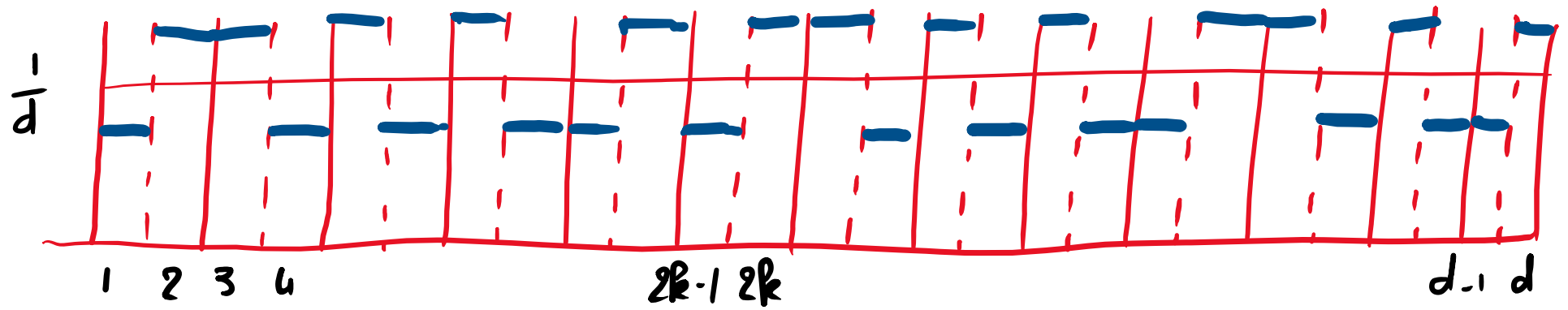
Lower bounds

Can we do better? No.

(But how to prove it?)

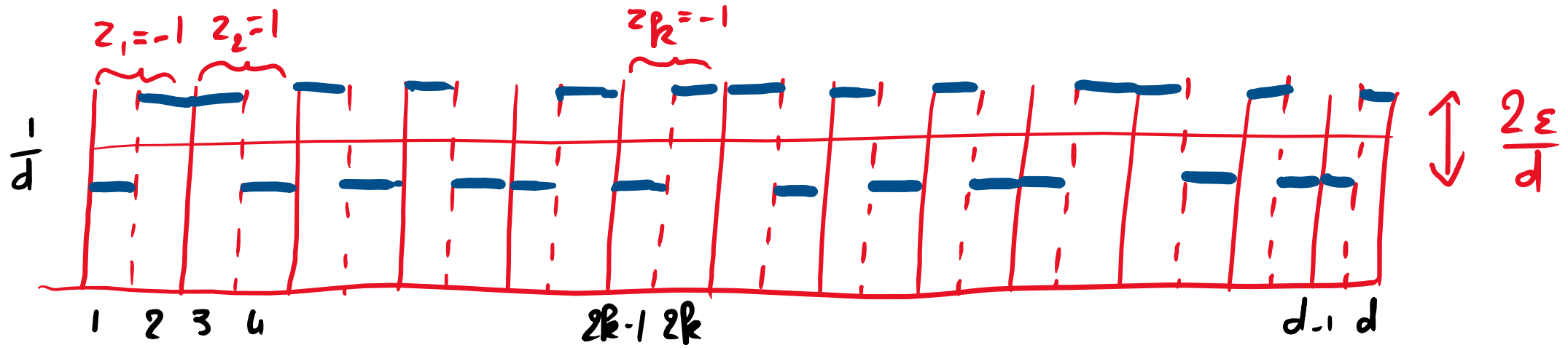
Let's start with a collection of hard instances $\mathcal{P} = \{p_z\}_{z \in \{\pm 1\}^{d/2}}$:

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



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

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



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

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

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

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

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

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

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 $\mathcal{Y}$:

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

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

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

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

Inputs $X_1, \dots, X_n \sim p$ (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 $\mathcal{Y}$:

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

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

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

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

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

Our goal:

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

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: $Z_1, \dots, Z_{d/2}$ 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, ε, d, 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, ε, d, 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,

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

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

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

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

and, taking $\mathbb{E}$,

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

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 $\hat{Z} := \operatorname{argmin}_z TV(p_z, \hat{p})$. Then

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

$\frac{2\varepsilon}{d}\,\mathrm{Ham}(\hat{Z}, Z)$

and, taking $\mathbb{E}$,

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

learning protocol

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.

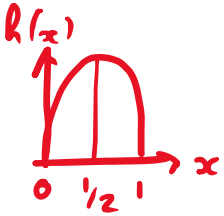
Learning: For Z uniform and Y^n transcript of learning protocol,

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

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

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

and so

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

□

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

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

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

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

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

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

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

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

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

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

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

Then

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

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

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

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

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

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

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

Then

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

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

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

← joint convexity

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

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

Step ② For $1 \le i \le \frac{d}{2}$,

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

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

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

Step ② For $1 \le i \le \frac{d}{2}$,

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

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

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

Chain rule for KL

$P_Z^{Y^{t-1}}$ ← no dependence on i

Step ② For $1 \le i \le \frac{d}{2}$,

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

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

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

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

Step ② For $1 \le i \le \frac{d}{2}$,

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

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

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

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

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

So... what now?

Key observation: $\forall y,$

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

Follows from our construct° + expression of P_z^W

Key observation: $\forall y$,

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

Follows from our construct° + expression of P_Z^W

Using this,

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

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

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

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

$$I(Z_i; Y^n) \leq \text{cst} \frac{\varepsilon^2}{d} \sum_{t=1}^{n} \mathbb{E}_Z \mathbb{E}_{P_2^{Y^{t-1}}} \sum_y \frac{\left(W^{Y^{t-1}}(y|2i-1) - W^{Y^{t-1}}(y|2i)\right)^2}{\sum_x W^{Y^{t-1}}(y|x)}$$

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

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

$$\sum_{i=1}^{d/2} I(Z_i; Y^n) \leq \frac{\text{cst}\,\varepsilon^2}{d} \sum_{t=1}^{n} \mathbb{E}_Z \mathbb{E}_{P_2^{Y^{t-1}}} \sum_{i=1}^{d/2} \sum_y \frac{\left(W^{Y^{t-1}}(y|2i-1) - W^{Y^{t-1}}(y|2i)\right)^2}{\sum_x W^{Y^{t-1}}(y|x)}$$

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

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

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

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

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

$$\sum_{i=1}^{d/2} I(Z_i; Y^n) \lesssim \frac{\text{cst}\,\varepsilon^2}{d} \sum_{t=1}^{n} \mathbb{E}_Z \mathbb{E}_{P_2^{Y^{t-1}}} \mathrm{Tr}[H(W^{t-1})]$$

$$\lesssim \frac{\text{cst}\,\varepsilon^2}{d} \sum_{t=1}^{n} \sup_{W \in \mathcal{W}} \mathrm{Tr}[H(W)]$$

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

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

$$\sum_{i=1}^{d/2} I(Z_i; Y^n) \leq \frac{\text{cst}\,\varepsilon^2}{d} \sum_{t=1}^{n}{}' \mathbb{E}_Z \mathbb{E}_{P_Z^{Y^{t-1}}} \mathrm{Tr}[H(W^{t-1})]$$

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

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

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

Step ②

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

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

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

Step ②

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

(useful for testing)

Learning

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 ③ $\forall 1 \le t \le n$,

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

In particular, for $t = n$

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

What about testing?

Step ①: Le Cam.

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

What about testing?

Step ①: Le Cam.

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

What about testing?

Step ①: Le Cam.

$$\Omega(1) \leq TV\left(\mathbb{E}_z[p_z^{Y^n}], u^{Y^n}\right)^2 \overset{\text{(Pinsker)}}{\leq} KL\left(\mathbb{E}_z[p_z^{Y^n}] \,\|\, u^{Y^n}\right)$$

(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.

$$\Omega(1) \le \mathrm{TV}\left(\mathbb{E}_z[P_z^{Y^n}], u^{Y^n}\right)^2 \overset{\text{(Pinsker)}}{\le} \mathrm{KL}\left(\mathbb{E}_z[P_z^{Y^n}] \,\|\, u^{Y^n}\right)$$

(q^{Y^n})

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

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

What about testing ?

Step ①: Le Cam.

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

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

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

What about testing ?

Step ①: Le Cam.

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

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

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

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

What did we show?

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

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

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

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

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 $\mathcal{P} = \{p_z\}_{z \in [-1,1]^{d/2}}$:

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

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

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

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

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

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

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

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

Testing

Long story short: get

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

for private-coin; and

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

for public-coin.

Hölder:

$$\|H(w)\|_F^2 \le \|H(w)\|_{op} \|H(w)\|_*$$

ℓ_2^2 $\quad$ ℓ_∞ $\quad$ ℓ_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 $\|H(\mathcal{W}_\ell)\|_F \simeq \|H(\mathcal{W}_\ell)\|_* \simeq \|H(\mathcal{W}_\ell)\|_{op} \simeq \ell^2$

. Communication $\|H(\mathcal{W}_\ell)\|_F \simeq 2^\ell$ $\|H(\mathcal{W}_\ell)\|_* \simeq \|H(\mathcal{W}_\ell)\|_{op} \simeq 2^{\ell/2}$

Immediately proves the LB!

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