Cookbook: Lower Bounds for Statistical Inference in Distributed and Constrained Settings

Jayadev Acharya, Clement Canonne, **Himanshu Tyagi**

FOCS 2020

Part III: Lower bounds for hypothesis testing

Hypothesis testing

1. Identity testing

Dimension $= k-1$, Accuracy $= \varepsilon$

$X^n := (X_1, \dots, X_n)$: Samples from an unknown $\mathbf{p}$ on $\mathcal{X} = [k]$
$\mathbf{q}$: reference distribution

Design a test $T(X^n)$ such that

$$\Pr(T(X^n) = 0) > 0.9, \text{ if } \mathbf{p} = \mathbf{q}$$

$$\Pr(T(X^n) = 1) > 0.9, \text{ if } \mathrm{d}_{\mathrm{TV}}(\mathbf{p}, \mathbf{q}) > \varepsilon$$

Sample Complexity = $\Theta\left(\frac{\sqrt{k}}{\varepsilon^2}\right)$

Hypothesis testing

2. High-dimensional mean testing

Dimension $= d$, Accuracy $= \varepsilon$

X^n: Samples from an unknown product distribution $\mathbf{p}$ on $\mathcal{X} = \mathbb{R}^d$
θ: mean of $\mathbf{p}$, $i.e.$, $\mathbb{E}[X_1] = \theta$

Design a test $T(X^n)$ such that

$$\Pr(T(X^n) = 0) > 0.9, \text{ if } \theta = 0$$
$$\Pr(T(X^n) = 1) > 0.9, \text{ if } \|\theta\|_2 > \varepsilon$$

Families of interest: Gaussian, Product Bernoulli

Sample Complexity = $\Theta\left(\frac{\sqrt{d}}{\varepsilon^2}\right)$

Hypothesis testing

3. The hide-and-seek problem

(Example for lower bounds, related to sparse mean estimation)

X^n: Samples from a product Bernoulli dist. $\mathbf{p}$ on $\mathcal{X} = \{-1, +1\}^d$

$\mathbb{E}\,[X_i] \in \{\theta_1, \theta_2, \dots, \theta_d\}$ is unkown where $\theta_j = \varepsilon \boldsymbol{e}_j, j \in [d]$

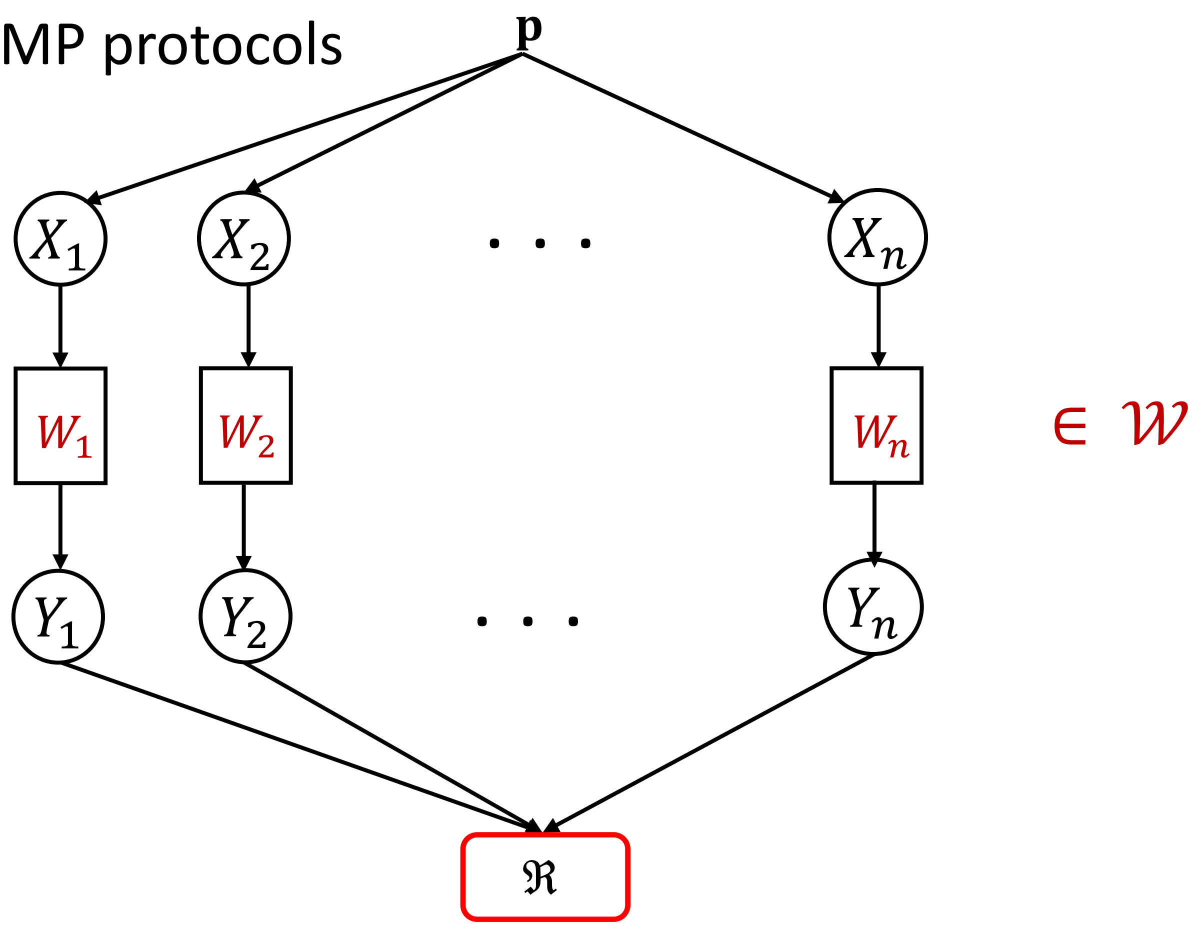
Design a test $T(X^n)$ such that

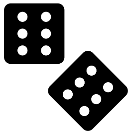
$$\Pr(T(X^n) = j) > 0.9, \text{ if } \mathbb{E}\,[X_1] = \theta_j, j \in [d]$$

Sample Complexity = $O\left(\frac{\log d}{\varepsilon^2}\right)$

Types of protocols

Private-coin SMP protocols

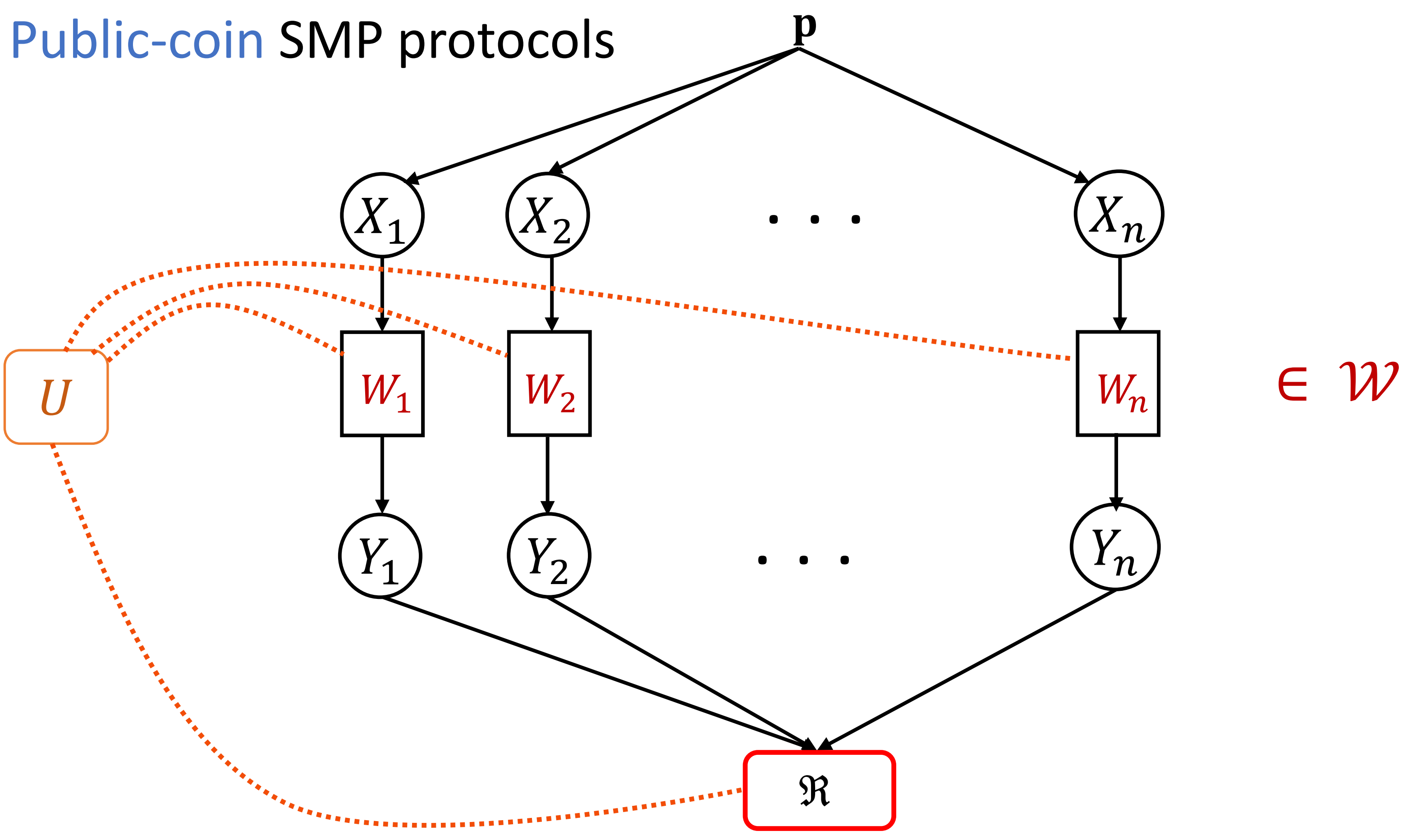


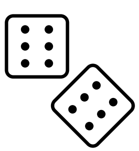


Referee $\mathfrak{R}$ applies test $T(Y^n)$

Types of protocols

Public-coin SMP protocols

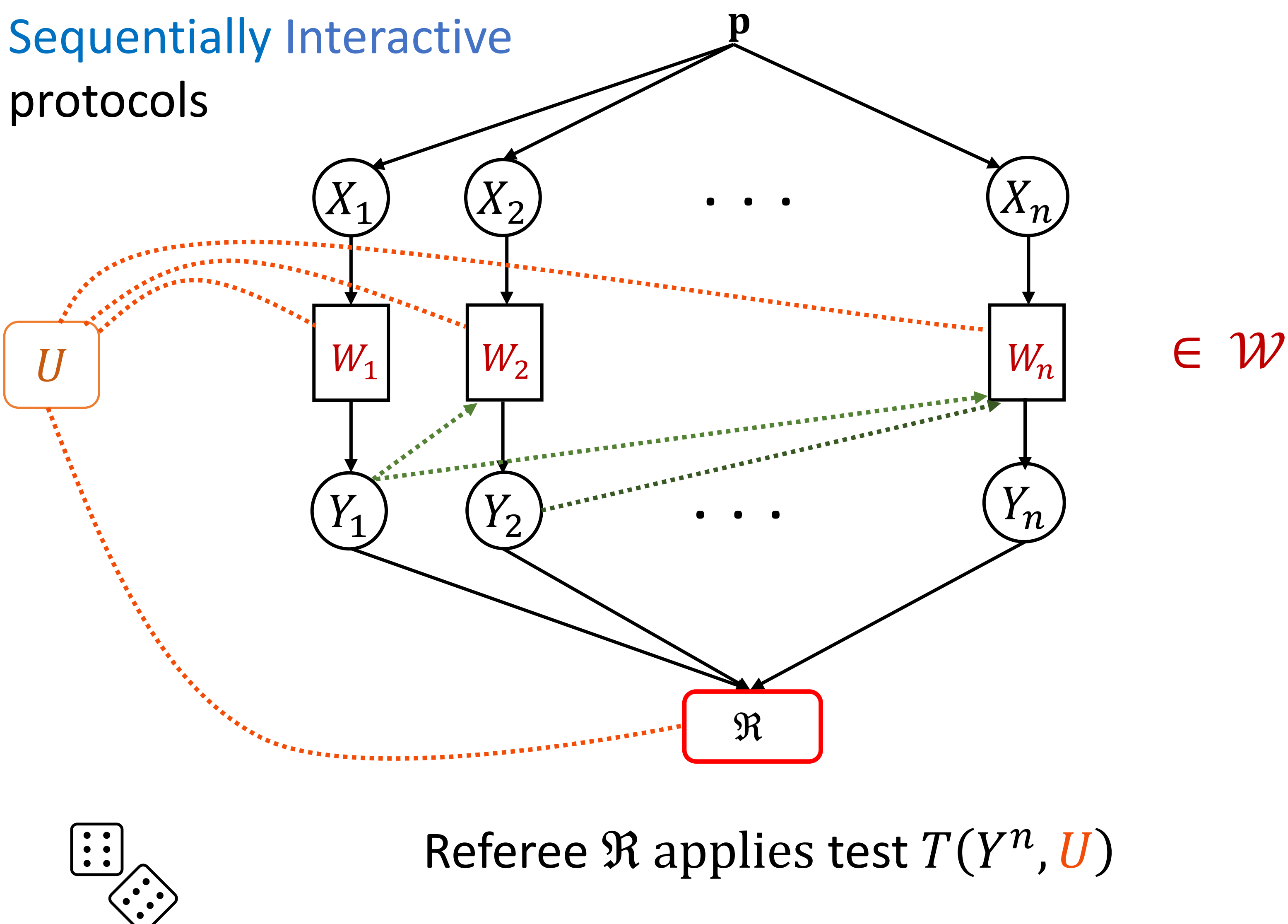




Referee $\mathfrak{R}$ applies test $T(Y^n, U)$

Types of protocols

Sequentially Interactive protocols



Referee $\mathfrak{R}$ applies test $T(Y^n, U)$

Information constrained setting

1. Communication constraints

ℓ-bit communication constrained players:

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

2. Local differential privacy constraints

ϱ-LDP channels

$$\mathcal{W}_\varrho = \left\{W: \max_{\{x,x' \in \mathcal{X}, y \in \mathcal{Y}\}} \frac{W(y|x)}{W(y|x')} \leq e^\varrho\right\}$$

The plan for this hour

Will derive lower bounds for sample complexity of hypothesis testing problems 1-3 under information constraints

1. Decoupled chi-square contraction bound
   - directly handle how chi-square distances between n-fold distributions shrink when samples are passed through channels from $\mathcal{W}$
2. Average mutual information bound
   - relate testing to the average information about each coordinate of the unknown parameter
3. Extensions to high-dimensional mean testing
   - general bounds and difficulties that emerge due to "nonlinear perturbations"

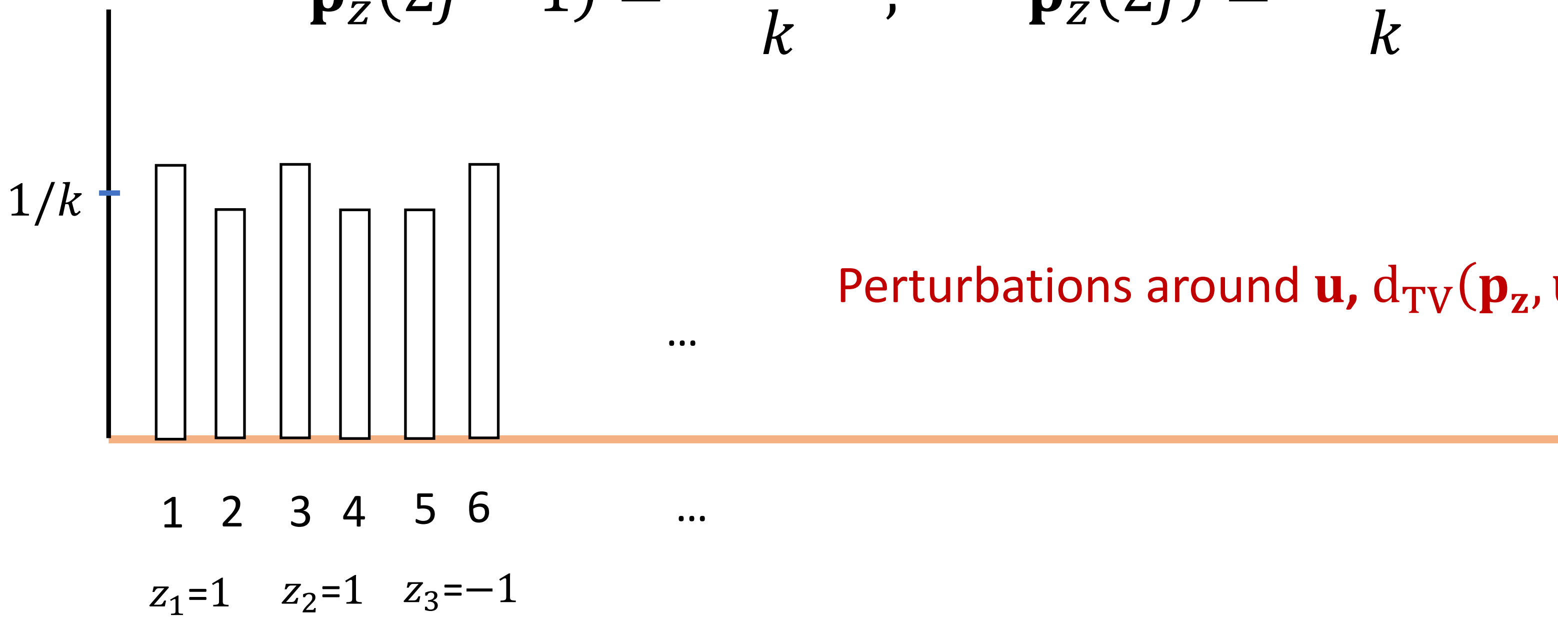
1. The Decoupled Chi-square Contraction Bound

Ingster's method

Use-case: Lower bound for (k, ε)-identity testing [Paninski '08]

Consider the set $\mathcal{P} = \left\{\mathbf{p}_z : z \in \{-1,1\}^{k/2}\right\}$

$$\mathbf{p}_z(2j-1) = \frac{1 + z_j \varepsilon}{k}, \qquad \mathbf{p}_z(2j) = \frac{1 - z_j \varepsilon}{k}$$



Perturbations around $\mathbf{u}$, $\mathrm{d_{TV}}(\mathbf{p_z}, \mathbf{u}) = \varepsilon$

[Paninski '08] L. Paninski, "A coincidence-based test for uniformity given very sparsely sampled discrete data," *IEEE Transactions on Information Theory*, 2008

Ingster's method

Use-case: Lower bound for (k, ε)-identity testing [Paninski '08]

Consider the set $\mathcal{P} = \left\{\mathbf{p}_z : z \in \{-1,1\}^{k/2}\right\}$

$$\mathbf{p}_z(2j-1) = \frac{1 + z_j \varepsilon}{k}, \qquad \mathbf{p}_z(2i) = \frac{1 - z_j \varepsilon}{k}$$

Observation 1**:**

Optimal Bayesian error = $\frac{1}{2}(1 - \mathrm{d}_{\mathrm{TV}}(\mathbf{p}, \mathbf{q}))$

Thus, if a test can distinguish $\mathbf{p}$ and $\mathbf{q}$, then $\mathrm{d}_{\mathrm{TV}}(\mathbf{p}, \mathbf{q}) \geq \mathrm{c}$

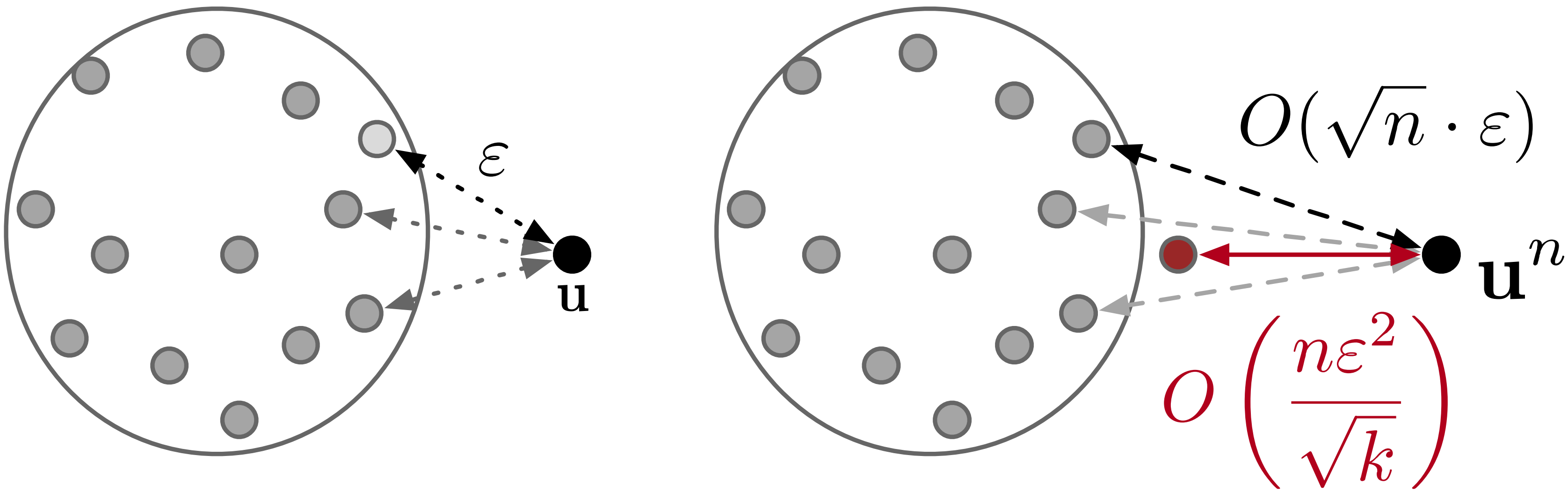
Observation 2**:**

Since our test can distinguish $\mathbf{p}_z^n$ and $\mathbf{u}^n$ for every z,

it can distinguish $\mathbb{E}\,[\mathbf{p}_Z^n]$ and $\mathbf{u}^n$ whereby $\mathrm{d}_{\mathrm{TV}}(\mathbb{E}\,[\mathbf{p}_Z^n], \mathbf{u}^n) \geq \mathrm{c}$

What is the least n needed to get $\mathrm{d}_{\mathrm{TV}}(\mathbb{E}\,[\mathbf{p}_Z^n], \mathbf{u}^n) \geq \mathrm{c}$?

Ingster's method



The mixture $\mathbb{E}\left[\mathbf{p}_Z^n\right]$ is much closer to $\mathbf{u}^n$ than any individual $\mathbf{p}_z^n$

Ingster's method

The mixture $\mathbb{E}\left[\mathbf{p}_Z^n\right]$ is much closer to $\mathbf{u}^n$ than any individual $\mathbf{p}_Z^n$

1. Switch to *chi-square divergence* …

A very quick primer on chi-square divergence

Definition

$$\mathrm{d}_{\chi^2}(\mathbf{p},\mathbf{q}) \stackrel{\text{def}}{=} \sum_x \frac{\left(\mathbf{p}(x)-\mathbf{q}(x)\right)^2}{\mathbf{q}(x)}$$
$$= \mathbb{E}_{\mathbf{q}}\left[\Delta^2\right]$$
$$= \mathbb{E}_{\mathbf{q}}\left[(1+\Delta)^2\right]-1,$$

where $\Delta(x) \stackrel{\text{def}}{=} \frac{(\mathbf{p}(x)-\mathbf{q}(x))}{\mathbf{q}(x)}$ is the normalized change

Property

$$\mathrm{d}_{\mathrm{TV}}(\mathbf{p},\mathbf{q}) = \mathbb{E}_{\mathbf{q}}\left[|\Delta|\right] \leq \sqrt{\mathrm{d}_{\chi^2}(\mathbf{p},\mathbf{q})}$$

Ingster's method

The mixture $\mathbb{E}\left[\mathbf{p}_Z^n\right]$ is much closer to $\mathbf{u}^n$ than any individual $\mathbf{p}_z^n$

1. Switch to *chi-square divergence* …

$$\mathrm{d}_{\mathrm{TV}}(\mathbb{E}\left[\mathbf{p}_Z^n\right], \mathbf{u}^n) \leq \sqrt{\mathrm{d}_{\chi^2}(\mathbb{E}\left[\mathbf{p}_Z^n\right], \mathbf{u}^n)}$$

2. Exploit the uncorrelatedness of Z_i to cancel "contributions" to the distance (the Ingster trick):

Warning: manipulations ahead…

Ingster's method

- $\Delta_Z^n \stackrel{\text{def}}{=} \frac{\mathbf{p}_Z^n - \mathbf{u}^n}{\mathbf{u}^n} \Rightarrow \; 1 + \Delta_Z^n(\boldsymbol{x}) = \prod_{i=1}^n \frac{\mathbf{p}_Z(x_i)}{\mathbf{u}(x_i)} = \prod_{i=1}^n (1 + \Delta_Z(x_i))$
- Z' is an independent copy of Z

(The Decoupling Step)

$$\mathrm{d}_{\chi^2}(\mathbb{E}_Z\,[\mathbf{p}_Z^n], \mathbf{u}^n)$$

$$= \mathbb{E}[(1 + \mathbb{E}_Z[\Delta_Z^n])^2] \;- 1$$

$$= \mathbb{E}[\mathbb{E}_{ZZ'}[(1 + \Delta_Z^n)(1 + \Delta_{Z'}^n)]] \;- 1$$

$$= \mathbb{E}_{ZZ'}[\mathbb{E}[(1 + \Delta_Z^n)(1 + \Delta_{Z'}^n)]] \;- 1$$

Ingster's method

- $\Delta_Z^n \overset{\text{def}}{=} \frac{\mathbf{p}_Z^n - \mathbf{u}^n}{\mathbf{u}^n} \Rightarrow \; 1 + \Delta_Z^n(\boldsymbol{x}) = \prod_{i=1}^n \frac{\mathbf{p}_Z(x_i)}{\mathbf{u}(x_i)} = \prod_{i=1}^n (1 + \Delta_Z(x_i))$
- Z' is an independent copy of Z

$$\mathrm{d}_{\chi^2}(\mathbb{E}_Z\,[\mathbf{p}_Z^n], \mathbf{u}^n) = \mathbb{E}_{ZZ'}[\mathbb{E}[(1+\Delta_Z^n)(1+\Delta_{Z'}^n)]] - 1$$ (decoupling)

$$= \mathbb{E}_{ZZ'}[\prod_{i=1}^n (1 + \mathbb{E}[\,\Delta_Z(X_i)\Delta_{Z'}(X_i)])] - 1$$ (averaging out uncorrelated terms)

$$\leq \mathbb{E}_{ZZ'}\left[\mathrm{e}^{n\mathbb{E}[\,\Delta_Z(X_1)\Delta_{Z'}(X_1)]}\right] - 1$$ (since $1 + t \leq e^t$)

Noting that $\mathbb{E}[\,\Delta_Z(X_1)\Delta_{Z'}(X_1)] = \frac{2\varepsilon^2}{k}\sum_{j=1}^{k/2} Z_j Z_j'$ and using Hoeffding's bound

$$\boxed{\mathrm{d}_{\chi^2}(\mathbb{E}\,[\mathbf{p}_Z^n], \mathbf{u}^n) \leq e^{\frac{n^2\varepsilon^4}{k}} - 1}$$

Ingster's method (as used in [Paninski'08])

The mixture $\mathbb{E}\left[\mathbf{p}_Z^n\right]$ is much closer to $\mathbf{u}^n$ than any individual $\mathbf{p}_Z^n$

1. Switch to *chi-square divergence* …

$$\mathrm{d_{TV}}(\mathbb{E}\left[\mathbf{p}_Z^n\right], \mathbf{u}^n) \leq \sqrt{\mathrm{d}_{\chi^2}(\mathbb{E}\left[\mathbf{p}_Z^n\right], \mathbf{u}^n)}$$

2. Exploit the uncorrelatedness of Z_i to cancel "contributions" to the distance (the Ingster trick):

$$\mathrm{d}_{\chi^2}(\mathbb{E}\left[\mathbf{p}_Z^n\right], \mathbf{u}^n) \leq e^{\frac{n^2\varepsilon^4}{k}} - 1$$

whereby $e^{n^2\varepsilon^4/k} \geq \log(1+c) \Rightarrow$ $n \geq \Omega\left(\frac{\sqrt{k}}{\varepsilon^2}\right)$

Take away 1: Summary of Ingster's method

- The mixture is much closer to $\mathbf{u}^n$ than individual $\mathbf{p}_Z^n$ (which are all at distance $\sqrt{n}\,\varepsilon$)

$$\mathrm{d_{TV}}(\mathbb{E}\,[\mathbf{p}_Z^n], \mathbf{u}^n) \leq \sqrt{e^{\frac{n^2\varepsilon^4}{k}} - 1} \approx \sqrt{n}\,\varepsilon \,.\, \frac{\sqrt{n}\,\varepsilon}{\sqrt{k}}$$

Smaller than 1 if $n < k/\varepsilon^2$

- The quadratic form of d_{χ^2} is useful to handle mixtures

Ingster's method in the information constrained setting

Lower bounds for identity testing

- <u>Notation.</u>
  - Channels $W^n = W_1 \otimes \; ... \otimes \; W_n$
  - $\mathbf{p}^{W^n}$ the output distrib. for W^n when the input distrib. is $\mathbf{p}^n$
  - For a p.s.d. matrix A with eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_m$, recall

$$\| A \|_{\mathrm{F}} = \sqrt{\lambda_1^2 + \lambda_2^2 + \cdots + \lambda_m^2}$$

$$\| A \|_{*} = \lambda_1 + \lambda_2 + \cdots + \lambda_m$$

$$\| A \|_{\mathrm{OP}} = \max_{\mathrm{i}} \lambda_i$$

Ingster's method in the information constrained setting

Lower bounds for identity testing

- <u>An observation.</u> For *public-coin protocols*:
  If we can resolve the mixture vs uniform problem,
  we can derandomize and resolve it using a constant U

- Earlier we saw
  $\mathrm{d}_{\chi^2}\left(\mathbb{E}\,[\mathbf{p}_Z^{W^n}], \mathbf{u}^{W^n}\right) = \mathbb{E}_{ZZ'}[\prod_{i=1}^{n}(1 + \mathbb{E}[\,\Delta_Z^{W_i}(X_i)\Delta_{Z'}^{W_i}(X_i)])] - 1$

Ingster's method in the information constrained setting

We start at the last expression:

$$\mathrm{d}_{\chi^2}\left(\mathbb{E}\,[\mathbf{p}_Z^{W^n}], \mathbf{u}^{W^n}\right) = \mathbb{E}_{ZZ'}[\prod_{i=1}^{n}(1 + \mathbb{E}[\,\Delta_Z^{W_i}(X_i)\Delta_{Z'}^{W_i}(X_i)])] - 1$$

The key observation.

$$\mathbb{E}\left[\,\Delta_Z^{W_i}(X_i)\Delta_{Z'}^{W_i}(X_i)\right] = \frac{2\varepsilon^2}{k} Z^T H(W_i) Z'$$

where $H(W)$ is $\frac{k}{2} \times \frac{k}{2}$ matrix with (j_1, j_2) entry given by

$$\sum_y \frac{\big(W(y|2j_1 - 1) - W(y|2j_1)\big)\big(W(y|2j_2 - 1) - W(y|2j_2)\big)}{\sum_x W(y|x)}$$

Ingster's method in the information constrained setting

$$\mathrm{d}_{\chi^2}\left(\mathbb{E}\left[\mathbf{p}_Z^{W^n}\right], \mathbf{u}^{W^n}\right) = \mathbb{E}_{ZZ'}\left[\prod_{i=1}^{n}\left(1 + \frac{2\varepsilon^2}{k} Z^T H(W_i) Z'\right)\right] - 1$$

$$\leq \mathbb{E}_{ZZ'}\left[e^{\frac{2\varepsilon^2}{k} Z^T \left(\sum_{i=1}^{n} H(W_i)\right) Z'}\right] - 1$$

$$= \mathbb{E}_{ZZ'}\left[e^{\frac{2n\varepsilon^2}{k} Z^T \overline{H} Z'}\right] - 1, \text{ where } \overline{H} = \frac{1}{n}\sum_i H(W_i)$$

Using a decoupling bound (for Rademacher chaos),

$$\mathrm{d}_{\chi^2}\left(\mathbb{E}\left[\mathbf{p}_Z^{W^n}\right], \mathbf{u}^{W^n}\right) \lesssim \frac{n^2 \varepsilon^4}{k^2} \cdot \|\overline{H}\|_F^2 \leq \frac{n^2 \varepsilon^4}{k^2} \max_{W \in \mathcal{W}} \|H(W)\|_F^2$$

which implies that $n \geq \Omega\left(\frac{\sqrt{k}}{\varepsilon^2} \cdot \frac{\sqrt{k}}{\max_{W \in \mathcal{W}} \|H(W)\|_F}\right)$

Chi-square contraction due to information constraints

Private-coin protocols

Ingster's method applied to private-coin identity testing

- For public-coin protocol, we "derandomized" in the first step. Perhaps a better bound can be obtained if minimize over the choice of $\left\{\mathbf{p}_z, z \in \{-1,1\}^{k/2}\right\}$

- But this approach cannot work for public-coin protocols because, heuristically, the shared randomness allows the protocol to "align" to the difficult case *(formally, the choice of channels used can depend on the difficult case)*

However, this can be done for private-coin protocols!

Private-coin protocols

Ingster's method applied to private-coin identity testing

We choose $Z = VY$, where Y is Rademacher vector as before and V is a $\frac{k}{2} \times \frac{k}{4}$ matrix chosen to make the family the most challenging for W^n

$$\mathrm{d}_{\chi^2}\left(\mathbb{E}\left[\mathbf{p}_Z^{W^n}\right], \mathbf{u}^{W^n}\right) \le \mathbb{E}_{ZZ'}\left[e^{\frac{2n\varepsilon^2}{k} Z^T \overline{H} Z'}\right] - 1 = \mathbb{E}_{YY'}\left[e^{\frac{2n\varepsilon^2}{k} Y^T V^T \overline{H} V Y'}\right] - 1$$

$$\approx \frac{n^2\varepsilon^4}{k^2} \| V^T \overline{H} V \|_F^2$$

Choose V so that it picks the smallest $\frac{k}{4}$ eigenvalues of p.s.d. matrix $\overline{H}$ to get

$$\mathrm{d}_{\chi^2}\left(\mathbb{E}\left[\mathbf{p}_Z^{W^n}\right], \mathbf{u}^{W^n}\right) \lesssim \frac{n^2\varepsilon^4}{k^2} \frac{\|\overline{H}\|_*^2}{k} \le \frac{n^2\varepsilon^4}{k^3} \max_{W \in \mathcal{W}} \| H(W) \|_*^2$$

which implies that $n \ge \Omega\left(\frac{\sqrt{k}}{\varepsilon^2} \cdot \frac{k}{\max_{W \in \mathcal{W}} \|H(W)\|_*}\right)$

Chi-square contraction due to information constraints

Take away 2: SMP chi-square contraction

- We can bound the contraction in chi-square divergences between mixture and the uniform using Ingster's method

- We get more restrictive bounds for private-coin protocols:

Sample-complexity lower bounds for identity testing

Public-coin protocols: $\Omega\left(\frac{\sqrt{k}}{\varepsilon^2} \cdot \frac{\sqrt{k}}{\max_{W \in \mathcal{W}} \|H(W)\|_F}\right)$

Private-coin protocols: $\Omega\left(\frac{\sqrt{k}}{\varepsilon^2} \cdot \frac{k}{\max_{W \in \mathcal{W}} \|H(W)\|_*}\right)$

2. The average information bound for interactive testing

Relating testing to average information

- Assouad's method implies that the difficulty of the learning problem is related to the average information $\frac{2}{k}\sum_i I(Z_i \wedge Y^n)$
- Interestingly, we will now see that even the difficulty of the testing problem can be related to the same quantity

Abbreviate $\mathbf{q}^{Y^n} = \mathbb{E}\left[\mathbf{p}_Z^{W^n}\right]$ and $\mathbf{u}^{Y^n} = \mathbf{u}^{W^n}$

Step 1. Chain rule in KL divergence before switching to chi-square

$$2\mathrm{d}_{\mathrm{TV}}\left(\mathbf{q}^{Y^n} \parallel \mathbf{u}^{Y^n}\right)^2 \le \mathrm{D}\left(\mathbf{q}^{Y^n} \parallel \mathbf{u}^{Y^n}\right)$$

$$= \sum_{t=1}^n \mathbb{E}_{\mathbf{q}^{Y^{t-1}}}\left[\mathrm{D}\left(\mathbf{q}^{Y_t|Y^{t-1}} \parallel \mathbf{u}^{Y_t|Y^{t-1}}\right)\right]$$

$$\le \sum_{t=1}^n \mathbb{E}_{\mathbf{q}^{Y^{t-1}}}\left[\mathrm{d}_{\chi^2}\left(\mathbf{q}^{Y_t|Y^{t-1}} \parallel \mathbf{u}^{Y_t|Y^{t-1}}\right)\right]$$

Relating testing to average information

- Step 1 gives $2\mathrm{d}_{\mathrm{TV}}\left(\mathbf{q}^{Y^n} \parallel \mathbf{u}^{Y^n}\right)^2 \leq \sum_{t=1}^{n} \mathbb{E}_{\mathbf{q}^{Y^{t-1}}}\left[\mathrm{d}_{\chi^2}\left(\mathbf{q}^{Y_t|Y^{t-1}} \parallel \mathbf{u}^{Y_t|Y^{t-1}}\right)\right]$

Step 2. Bringing-in the Channel Information matrix H

- Recall that for Paninski's construction:

$$\mathbf{q}_{X_t|Y^{t-1}}(2j-1) = \frac{1+\varepsilon\mathbb{E}[Z_j|Y^{t-1}]}{k}; \quad \mathbf{q}_{X_t|Y^{t-1}}(2j) = \frac{1-\varepsilon\mathbb{E}[Z_j|Y^{t-1}]}{k}, j \in \left[\frac{k}{2}\right]$$

$$\mathrm{d}_{\chi^2}\left(\mathbf{q}^{Y_t|Y^{t-1}} \parallel \mathbf{u}^{Y_t|Y^{t-1}}\right)$$

$$= \frac{\varepsilon^2}{k} \sum_y \frac{\left(\sum_i \mathbb{E}[Z_i|Y^{t-1}]\left(W^{Y^{t-1}}(y|2i-1) - W^{Y^{t-1}}(y|2i)\right)\right)^2}{\sum_x W^{Y^{t-1}}(y|x)}$$

$$= \frac{\varepsilon^2}{k} \mathbb{E}[Z|Y^{t-1}]^T H\left(W^{Y^{t-1}}\right)\mathbb{E}[Z|Y^{t-1}]$$

Relating testing to average information

- Step 1 gives $2\mathrm{d}_{\mathrm{TV}}\left(\mathbf{q}^{Y^n} \parallel \mathbf{u}^{Y^n}\right)^2 \leq \sum_{t=1}^{n} \mathbb{E}_{\mathbf{q}^{Y^{t-1}}}\left[\mathrm{d}_{\chi^2}\left(\mathbf{q}^{Y_t|Y^{t-1}} \parallel \mathbf{u}^{Y_t|Y^{t-1}}\right)\right]$
- Step 2 gives $\mathrm{d}_{\chi^2}\left(\mathbf{q}^{Y_t|Y^{t-1}} \parallel \mathbf{u}^{Y_t|Y^{t-1}}\right) = \frac{\varepsilon^2}{k}\mathbb{E}[Z|Y^{t-1}]^T H\left(W^{Y^{t-1}}\right)\mathbb{E}[Z|Y^{t-1}]$

Step 3. "Channel Alignment" Bound

$$\mathbb{E}[Z|Y^{t-1}]^T H\left(W^{Y^{t-1}}\right)\mathbb{E}[Z|Y^{t-1}]$$

$$\leq \parallel H\left(W^{Y^{t-1}}\right) \parallel_{\mathrm{OP}} \cdot \parallel \mathbb{E}[Z|Y^{t-1}] \parallel_2^2$$

$$\leq \max_{W \in \mathcal{W}} \parallel H(W) \parallel_{\mathrm{OP}} \cdot \parallel \mathbb{E}[Z|Y^{t-1}] \parallel_2^2$$

Finally, the average information bound for testing ...

The average information bound for testing

Till now we have:

$$2 d_{TV}\left(\mathbf{q}^{Y^n} \parallel \mathbf{u}^{Y^n}\right)^2 \leq \frac{\varepsilon^2}{k} \cdot \max_{W \in \mathcal{W}} \parallel H(W) \parallel_{OP} \cdot \sum_{t=1}^{n} \mathbb{E}\left[\parallel \mathbb{E}[Z|Y^{t-1}] \parallel_2^2\right]$$

<u>An observation.</u>

For a random variable V taking values in $\{-1, +1\}$,

$$2 d_{TV}\left(\mathbf{q}^{Y^n} \parallel \mathbf{u}^{Y^n}\right)^2 \leq c \cdot \varepsilon^2 \max_{W \in \mathcal{W}} \parallel H(W) \parallel_{OP} \cdot \sum_{t=1}^{n} \frac{2}{k} \sum_{i} I(Z_i \wedge Y^{t-1})$$

$$1 - H(V) = D(P_V \parallel P_U) \geq \frac{\ln 2}{2} \mathbb{E}[V^2]$$

Therefore,

$$\mathbb{E}[\parallel \mathbb{E}[Z|Y^{t-1}] \parallel_2^2] = \sum_i \mathbb{E}[\mathbb{E}[Z_i|Y^{t-1}]^2] \leq \frac{2}{\ln 2} \sum_i I(Z_i \wedge Y^{t-1})$$

Bounding the average information

The average information bound for testing:

$$2\mathrm{d}_{\mathrm{TV}}\left(\mathbf{q}^{\mathrm{Y^n}} \parallel \mathbf{u}^{\mathrm{Y^n}}\right)^2 \le c \cdot \varepsilon^2 \max_{\mathrm{W}\in\mathcal{W}} \parallel \mathrm{H(W)} \parallel_{\mathrm{OP}} \cdot \sum_{t=1}^{n} \frac{2}{k} \sum_i I(Z_i \wedge Y^{t-1})$$

Earlier we saw:

$$\frac{2}{k} \sum_i I(Z_i \wedge Y^{t-1}) \le c \cdot (t-1) \cdot \frac{\varepsilon^2}{k^2} \max_{\mathrm{W}\in\mathcal{W}} \parallel \mathrm{H(W)} \parallel_*$$

which gives

$$\boxed{2\mathrm{d}_{\mathrm{TV}}\left(\mathbf{q}^{\mathrm{Y^n}} \parallel \mathbf{u}^{\mathrm{Y^n}}\right)^2 \le c.\frac{n^2\varepsilon^4}{k^2} \max_{\mathrm{W}\in\mathcal{W}} \parallel \mathrm{H(W)} \parallel_{\mathrm{OP}} \cdot \max_{\mathrm{W}\in\mathcal{W}} \parallel \mathrm{H(W)} \parallel_*}$$

whereby

$$n \ge \Omega\left(\frac{k}{\varepsilon^2 \sqrt{\max_{\mathrm{W}\in\mathcal{W}} \parallel \mathrm{H(W)} \parallel_{\mathrm{OP}} \cdot \max_{\mathrm{W}\in\mathcal{W}} \parallel \mathrm{H(W)} \parallel_*}}\right)$$

Take away 3: All chi-square contraction bounds

- Lower bounds for identity testing under information constraints

$$\| \mathcal{W} \| \stackrel{\text{def}}{=} \max_{W \in \mathcal{W}} \| H(W) \|$$

Classic	Private-coin SMP	Public-coin SMP	Sequentially Interactive
$\Omega\left(\frac{\sqrt{k}}{\varepsilon^2}\right)$	$\Omega\left(\frac{k^{3/2}}{\varepsilon^2 \| \mathcal{W} \|_*}\right)$	$\Omega\left(\frac{k}{\varepsilon^2 \| \mathcal{W} \|_{\mathrm{F}}}\right)$	$\Omega\left(\frac{k}{\varepsilon^2 \sqrt{\| \mathcal{W} \|_{OP} \| \mathcal{W} \|_*}}\right)$

- For the sequentially interactive lower bound:
  - Can be improved, in general, using the same recipe
  - We can find an example of constraints where interaction helps

Application: Identity testing for different $\mathcal{W}$

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

$\| \mathcal{W}_\ell \|_F \le \sqrt{2^\ell}$, $\| \mathcal{W}_\ell \|_* \le 2^\ell$, $\| \mathcal{W}_\ell \|_{OP} \le 2$

Private-coin	Public-coin	Sequentially Interactive
$\Omega\left(\frac{k^{3/2}}{\varepsilon^2 2^\ell}\right)$	$\Omega\left(\frac{k}{\varepsilon^2 \sqrt{2^\ell}}\right)$	$\Omega\left(\frac{k}{\varepsilon^2 \sqrt{2^\ell}}\right)$

- **These bounds will be seen to be tight**
- **Interaction doesn't help, but public coins do**

Private-coin	Public-coin	Sequentially Interactive
$\Omega\left(\frac{k^{3/2}}{\varepsilon^2 \rho^2}\right)$	$\Omega\left(\frac{k}{\varepsilon^2 \rho^2}\right)$	$\Omega\left(\frac{k}{\varepsilon^2 \rho^2}\right)$

3. High-dimensional mean testing (under communication constraints)

General chi-square bounds for public-coin SMP

1. Chi-square bound (we didn't see it earlier, but it's easy)

$$\mathrm{D}\left(\mathbb{E}\,[\mathbf{p}_Z^{W^n}] \parallel \mathbf{p}^{W^n}\right) \le \mathbb{E}_Z\left[\mathrm{D}\left(\mathbf{p}_Z^{W^n} \parallel \mathbf{p}^{W^n}\right)\right] = \sum_i \mathbb{E}_Z\left[\mathrm{D}\left(\mathbf{p}_Z^{W_i} \parallel \mathbf{p}^{W_i}\right)\right]$$

which upon bounding divergence with d_{χ^2} gives

$$\mathrm{D}\left(\mathbb{E}\,[\mathbf{p}_Z^{W^n}] \parallel \mathbf{p}^{W^n}\right) \le n \cdot \max_{W \in \mathcal{W}_\ell} \mathbb{E}_Z\left[\sum_y \frac{\mathbb{E}_X[\delta_Z(X) W(y|X)]^2}{\mathbb{E}_X[W(y|X)]}\right]$$

2. Decoupled chi-square bound (Ingster's method)

$$\mathrm{d}_{\chi^2}\left(\mathbb{E}\,[\mathbf{p}_Z^{W^n}], \mathbf{p}^{W^n}\right) \le$$

$$\max_{W^n} \mathbb{E}_{ZZ'}\left[e^{\sum_{i=1}^n \sum_y \frac{\mathbb{E}_X[\delta_Z(x) w_i(y|X)]\,\mathbb{E}_X[\delta_{Z'}(x) w_i(y|X)]}{\mathbb{E}_X[w(y|X)]}}\right] - 1$$

Hide-and-seek for public-coin SMP

$\mathbf{p}$ prod Bernoulli dist. on $\mathcal{X} = \{-1,+1\}^d$ with mean $\mathbf{0}$

$\mathbf{p}_z, z \in [d]$, prod Bernoulli dist. on $\mathcal{X} = \{-1,+1\}^d$ with mean $\varepsilon \boldsymbol{e}_z$

$\boldsymbol{\delta}_z(x) = \varepsilon x_z$ ("linear perturbation")

For the chi-square contraction bound:

$$\mathbb{E}_Z\left[\sum_y \frac{\mathbb{E}_X[\delta_Z(X)W(y|X)]^2}{\mathbb{E}_X[W(y|X)]}\right] = \frac{\varepsilon^2}{d}\sum_y \frac{\mathbb{E}_X[XW(y|X)]^2}{\mathbb{E}_X[W(y|X)]}$$

Hide-and-seek for public-coin SMP

$$\sum_y \mathbb{E}_Z\left[\frac{\mathbb{E}_X[\delta_Z(X)W(y|X)]^2}{\mathbb{E}_X[W(y|X)]}\right] = \frac{\varepsilon^2}{d}\sum_y \frac{\mathbb{E}_X[XW(y|X)]^2}{\mathbb{E}_X[W(y|X)]}$$

A measure change bound

(similar to Talagrand's Gaussian transportation inequality)

For random vector X as above (or Gaussian) and $a\colon \mathcal{X} \to [0,1]$,

$$\frac{\mathbb{E}[Xa(X)]^2}{\mathbb{E}[a(X)]^2} \leq 2\,\mathbb{E}\left[\frac{a(X)}{\mathbb{E}[a(X)]}\log\frac{a(X)}{\mathbb{E}[a(X)]}\right]$$

Proof uses Gibbs variational formula and additivity of divergence

Chi-square bound ⇒

$$\mathrm{D}\left(\mathbb{E}\,[\mathbf{p}_Z^{W^n}] \parallel \mathbf{p}^{W^n}\right) \leq c \cdot \frac{n\,\varepsilon^2}{d} \cdot \max_{\mathrm{W}}\ \mathrm{H}(\mathbb{E}[\mathrm{W}(\cdot\,|X)]) \leq c \cdot \frac{n\,\varepsilon^2}{d} \cdot \ell$$

Hide-and-seek for sequentially interactive

We used:

$$D\left(\mathbb{E}\left[\mathbf{p}_Z^{W^n}\right] \parallel \mathbf{p}^{W^n}\right) \le n \cdot \max_{W \in \mathcal{W}} \mathbb{E}_Z\left[\sum_y \frac{\mathbb{E}_X[\delta_Z(X) W(y|X)]^2}{\mathbb{E}_X[W(y|X)]}\right]$$

Even for sequentially interactive protocol, we can show

$$\mathbb{E}_Z\left[D\left(\mathbf{p}_Z^{W^n} \parallel \mathbf{p}^{W^n}\right)\right] \le \mathbb{E}_Z\left[\sum_i \mathbb{E}_{Y^{i-1}}\left[D\left(\mathbf{p}_Z^{Y_i|Y^{i-1}} \parallel \mathbf{p}^{Y_i|Y^{i-1}}\right)\right]\right]$$

$$\le n\, \mathbb{E}_Z\left[\max_{W \in \mathcal{W}_\ell} \sum_y \frac{\mathbb{E}_X[\delta_Z(X) W(y|X)]^2}{\mathbb{E}_X[W(y|X)]}\right]$$

But our previous bound requires us to take average over Z before taking the max

Alternatively, we can derive an average information bound for this case as well [Shamir '14]

High-dimensional mean testing

$\mathbf{p}$ Gaussian distribution $\mathcal{N}(0, \mathbb{I}_d)$

$\mathbf{p}_z, z \in \{-1,+1\}^d$, Gaussian distribution $\mathcal{N}\left(\frac{\varepsilon}{\sqrt{d}} z, \mathbb{I}_d\right)$

The main difficulty: nonlinear perturbation (in x)

$$\boldsymbol{\delta}_z(x) = e^{-\varepsilon^2/2} \, e^{\frac{\varepsilon}{\sqrt{d}}\langle x,z\rangle} - 1$$

But we can still derive a partial bound (using the chi-square bound)

High-dimensional mean testing

$$\boldsymbol{\delta}_z(x) = e^{-\varepsilon^2/2}\, e^{\frac{\varepsilon}{\sqrt{d}}\langle x,z\rangle} - 1$$

Chi-square divergence bound for Gaussian mean testing

For $\ell \le \frac{\sqrt{d}}{\varepsilon^2}$, $\mathbb{E}_Z\left[\sum_y \frac{\mathbb{E}_X[\delta_Z(X)W(y|X)]^2}{\mathbb{E}_X[W(y|X)]}\right] \le \mathcal{O}\left(\max\left\{\frac{\varepsilon^2\ell}{d}, \frac{\varepsilon^4\ell^2}{d}\right\}\right)$

- It is tight for constant ℓ or small enough ε
- The proof is tedious, uses level-k inequalities instead of our earlier Talagrand-type bound
- Does not work for interactive protocols – we need to take expectation over Z and cannot handle

$$\mathbb{E}_Z\left[\max_W \sum_y \frac{\mathbb{E}_X[\delta_Z(X)W(y|X)]^2}{\mathbb{E}_X[W(y|X)]}\right]$$

In conclusion

- Bounds seen
  - the chi-square contraction bounds for SMP protocols
  - the average information bound for sequentially interactive protocols
- The decoupled chi-square contraction bound obtained using Ingster's method shows separation of private- and public-coin protocols for identity testing
- The average information bound can be used to obtain a family of channels where interaction helps for testing
- Only partial results available for high-dimensional mean testing – the basic approach extends, but difficulty handling the resulting terms