

# Statistical Inference in Distributed or Constrained Settings: Techniques and Recipes



Conference on Learning Theory 2021



## What's on the menu?

I. Appetizers

Jayadev

II. MC 1

Jayadev

III. MC 2

Himanshu

IV. DIY Desserts

Clément

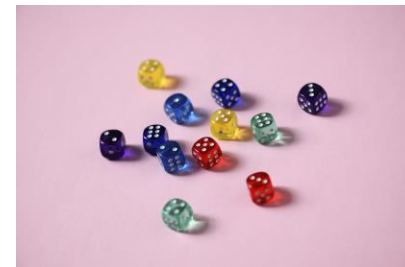
Chefs: Jayadev Acharya, Clément Canonne, Himanshu Tyagi

COLT 2021

# Appetizers

- Statistical Inference
- Distributed / constrained settings
- Problems and examples
- Related work and pointers

# Main Course – I: Discrete distributions



- A puzzle to solve **all** problems under communication constraints
- Lower bounds for interactive estimation for arbitrary channels
  - Tight bounds under communication, privacy as application

# Main Course – II: General distributions

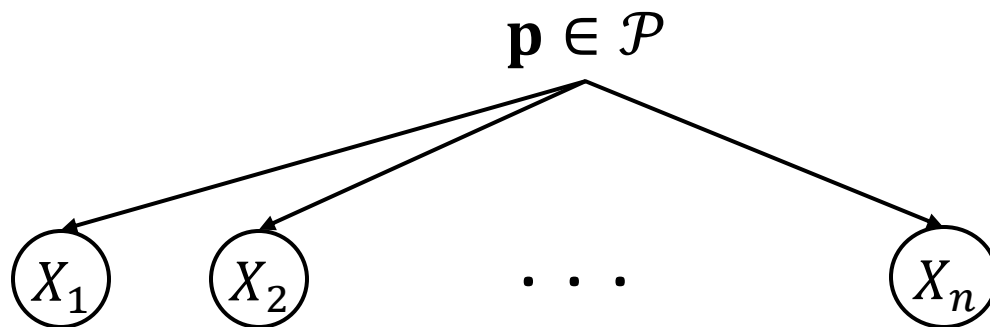
Unified method to prove “interactive” lower bounds

- Discrete, high-dimensional, nonparametric, etc
- Communication, privacy, etc
- General plug-n-play methods

- Example of how to apply the lower bounds
- Several exercises

# Statistical Inference

$\mathcal{P}$ : family of distributions over  $\mathcal{X}$



Given  $X^n := (X_1, \dots, X_n)$ : i.i.d. samples from an unknown  $\mathbf{p}$

Solve some inference task about  $\mathbf{p}$

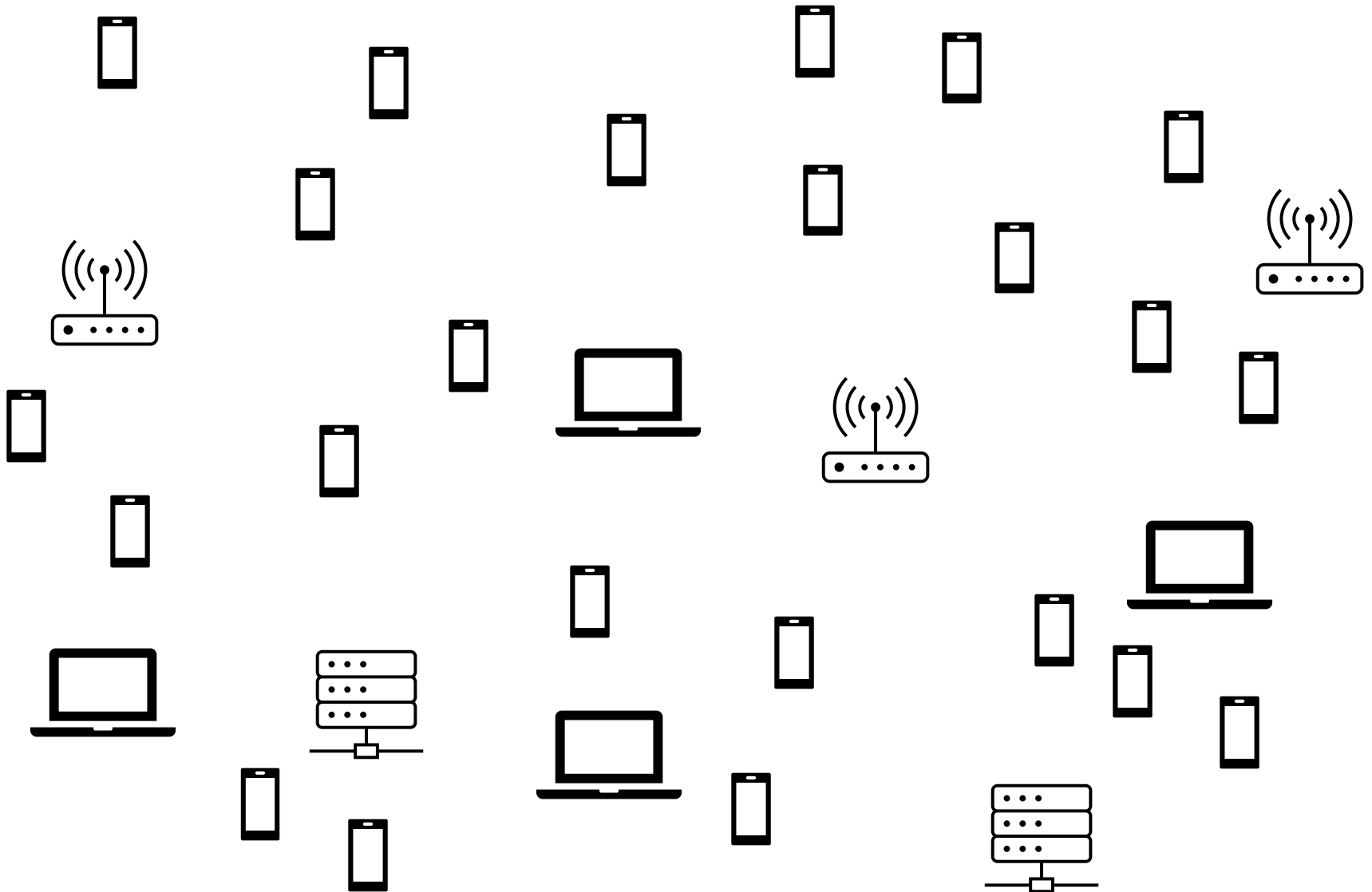
This is inference in central setting

# Information Constraints

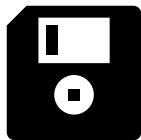


# Distributed or Constrained Settings

No direct access to  $X_i$ s



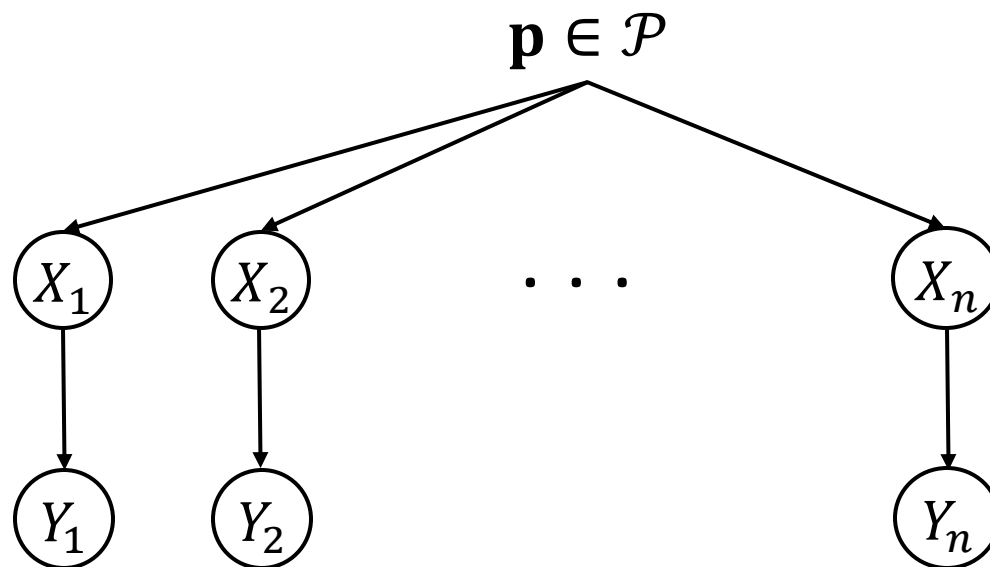
# Techniques and **Statistical Inference** under **constraints**



Local constraints



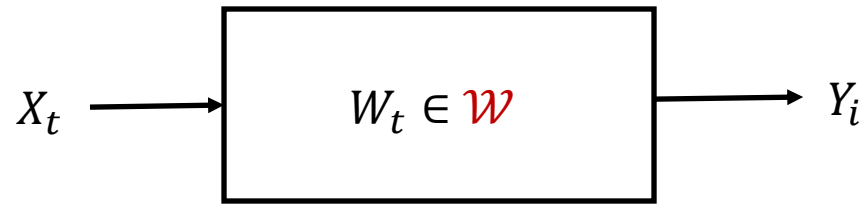
# Statistical Inference



# Modeling the constraints

[ACT20c]

$n$  users, user  $t$  observes  $X_t$  and sends message  $Y_t$

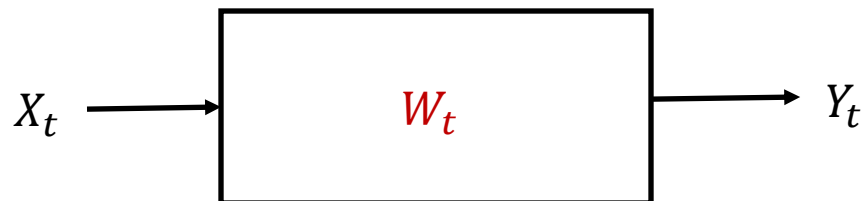


$$W_t(y|x) := \Pr(Y_t = y | X_t = x)$$

$\mathcal{W}$ : a set of **allowed** (randomized) channels  $\Leftrightarrow$  the **constraints**

The algorithm/protocol dictates how user  $t$  chooses  $W_t$  from  $\mathcal{W}$

# Modeling the local information constraints [\[ACT20c\]](#)



When  $X_t \sim \mathbf{p}$

$$\mathbf{p}^{W_t}(y) := \sum_x \mathbf{p}(x) W_t(y|x) = \mathbb{E}[W_t(y|X)]$$

# Example 1: Communication constraints

[Shamir14, HMÖW18, ACT20d...]

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

Each  $X_t$  is mapped to  $\ell$  bits.

Bandwidth  
constraints



# Example 2: Local Differential Privacy (LDP)

[Warner65, EPR03, KLNRS11]

$W: \mathcal{X} \rightarrow \{0,1\}^*$  is  $\varrho$ -**LDP** if  $\forall x, x' \in \mathcal{X}, \forall y$ ,

$$\frac{W(y|x)}{W(y|x')} \leq e^{\varrho} \approx 1 + \varrho$$

$\mathcal{W}_{\varrho} = \{\text{all } \varrho - \text{LDP channels}\}$

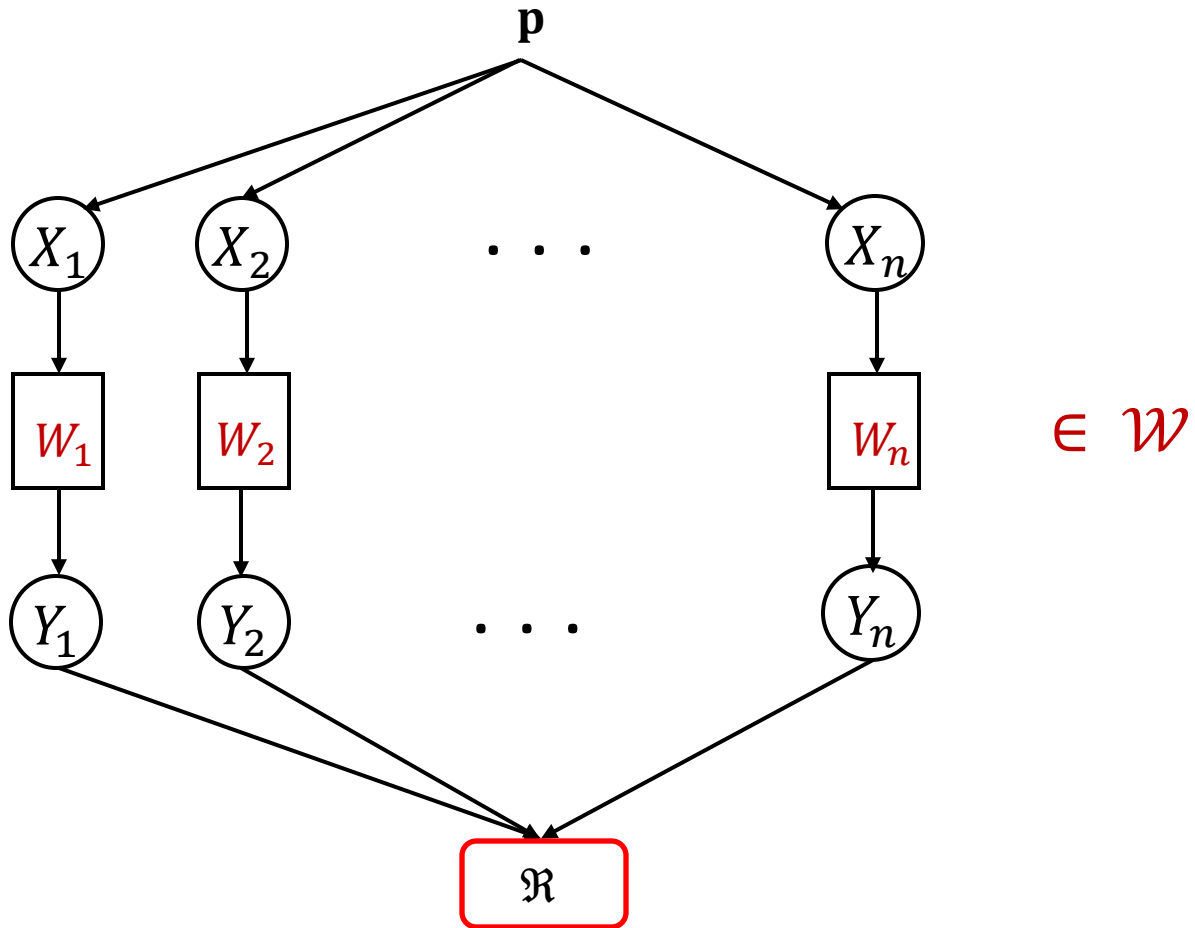
Privacy guarantees even  
“against” the server





# The Protocols

# Distributed Statistical Inference



Given  $Y^n := Y_1, \dots, Y_n$ , solve the inference task

# Distributed statistical inference

For  $W^n := W_1, \dots, W_n$ ,

$$\mathbf{p}^{W^n}(Y^n) = \prod_t \mathbf{p}^{W_t}(Y_t)$$

**How to choose  $W_1, W_2, \dots, W_n \in \mathcal{W}$  to minimize  $n$ ?**

# The protocols

## Simultaneous Message Passing (SMP)/Non-interactive schemes

$W_i$ s are chosen simultaneously

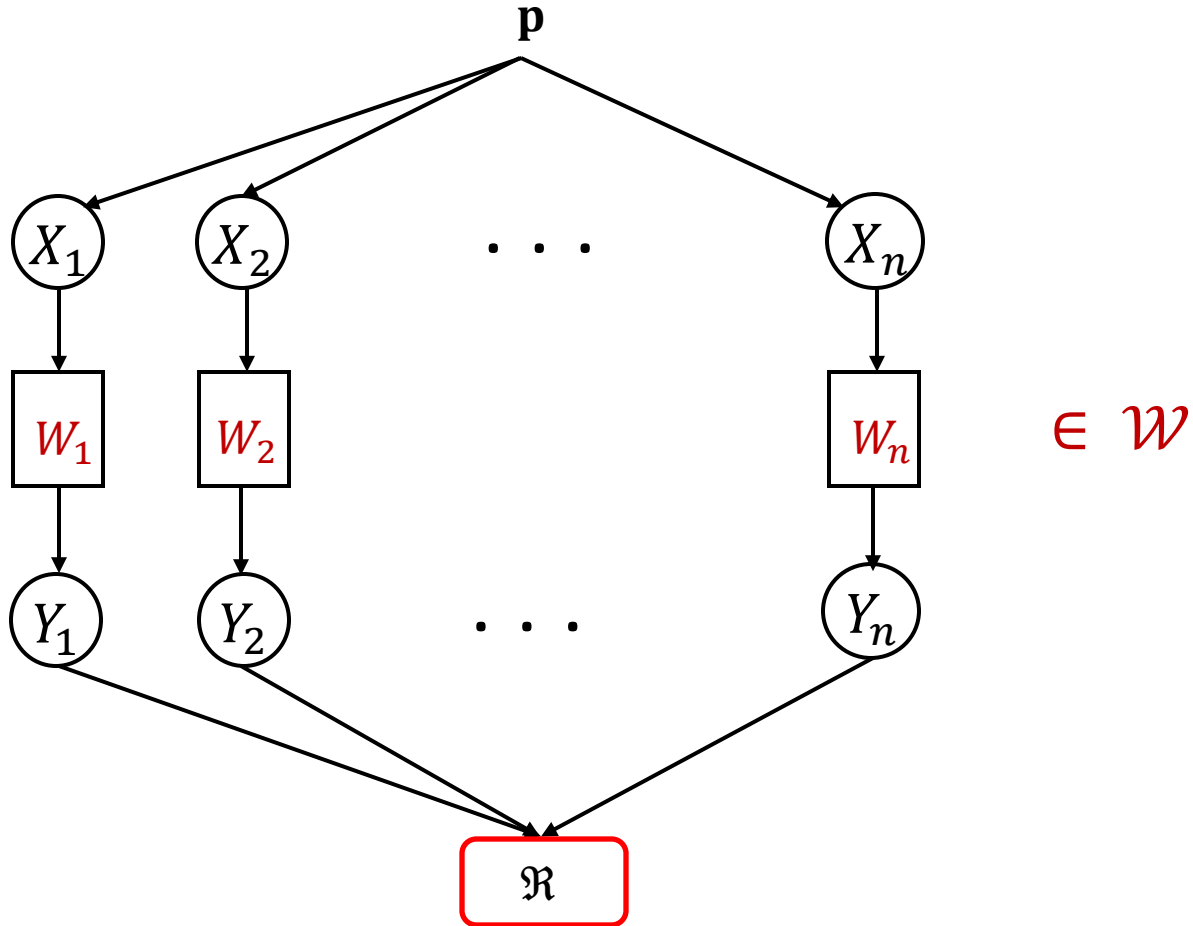
### private-coin SMP (no shared randomness)

$W_t$ s are chosen independently

$Y_1, Y_2, \dots, Y_n$  are independent

*e.g.*,  $W_1, \dots, W_n$  are fixed

# Private-coin SMP protocols



Noninteractive (“simultaneous message-passing”),  
no common randomness

# The protocols

## Simultaneous Message Passing (SMP)/Non-interactive schemes

$W_i$ s are chosen simultaneously

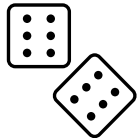
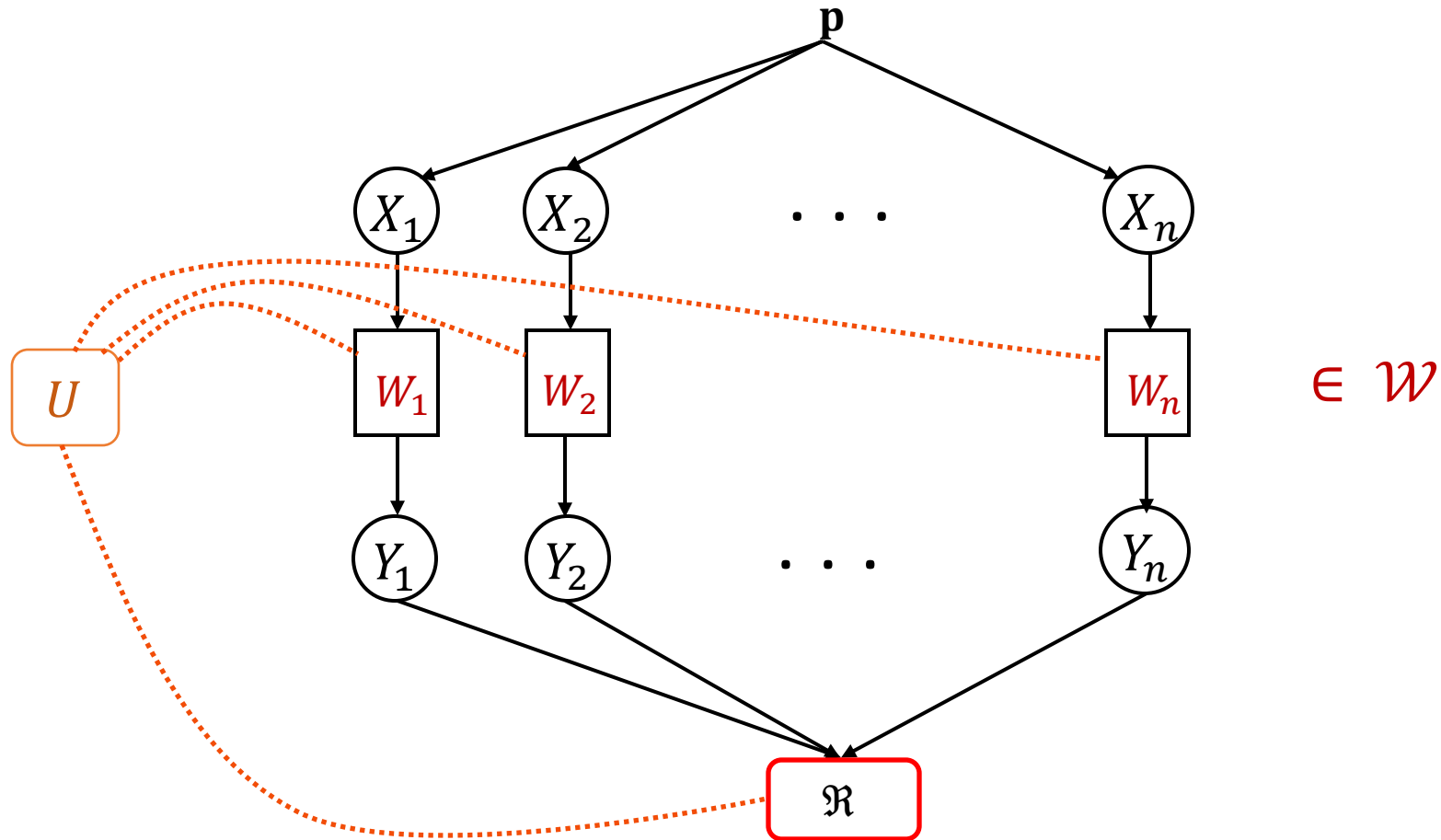
### public-coin SMP (shared randomness)

$U$ : common random string available to all users and referee

$W_t$ s is a function of  $U$

$Y_1, Y_2, \dots, Y_n$  are independent **given**  $U$

# Public-coin SMP protocols



Noninteractive (“simultaneous message-passing”),  
*but* common random seed

# The protocols

## Interactive schemes

$W_i$ s can depend on previous messages

### sequentially interactive protocols

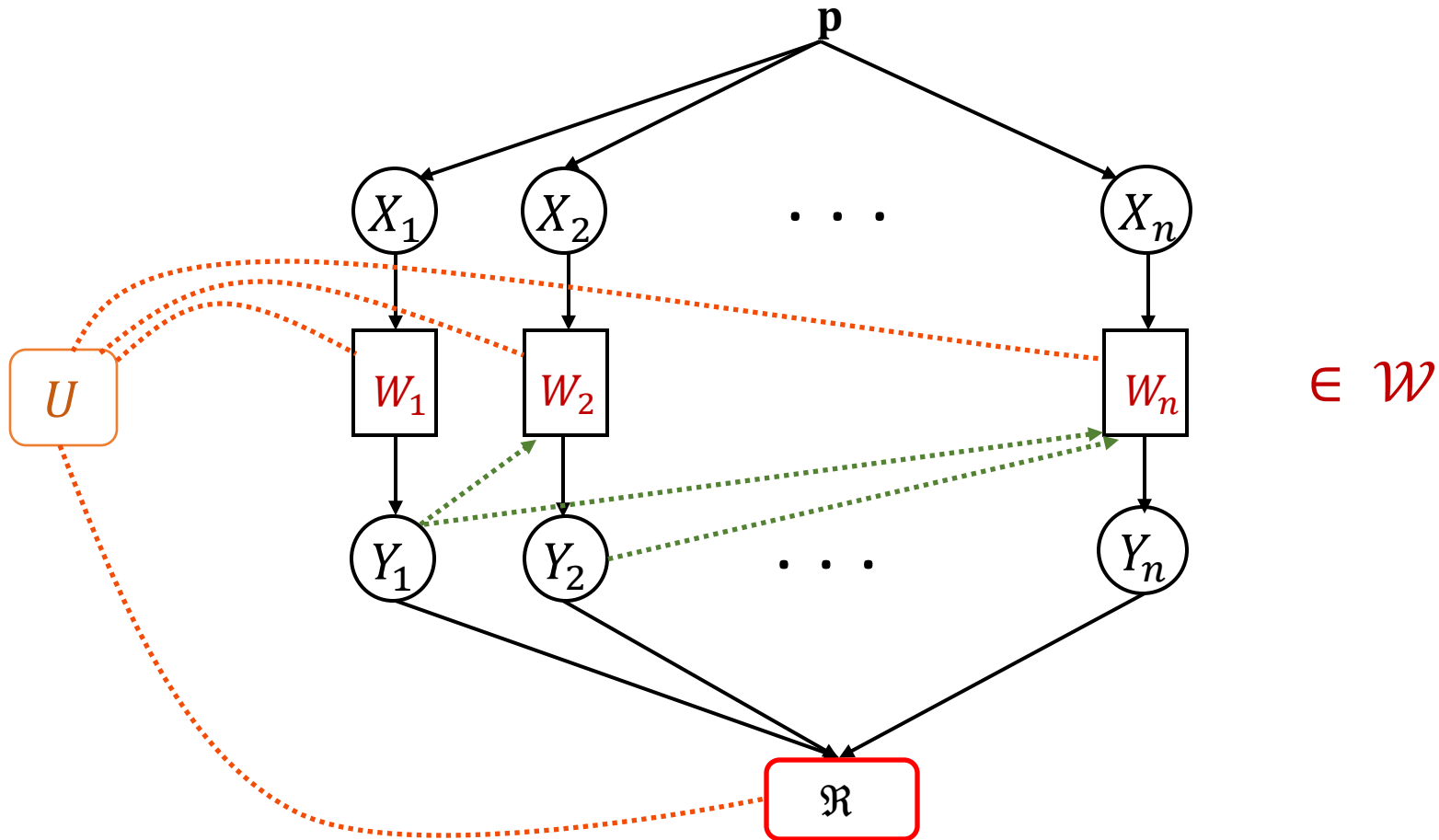
$U$ : common random string available to all users and referee

**for**  $t = 1, \dots, n$

$W_t$  is a function of  $(U, Y^{t-1})$



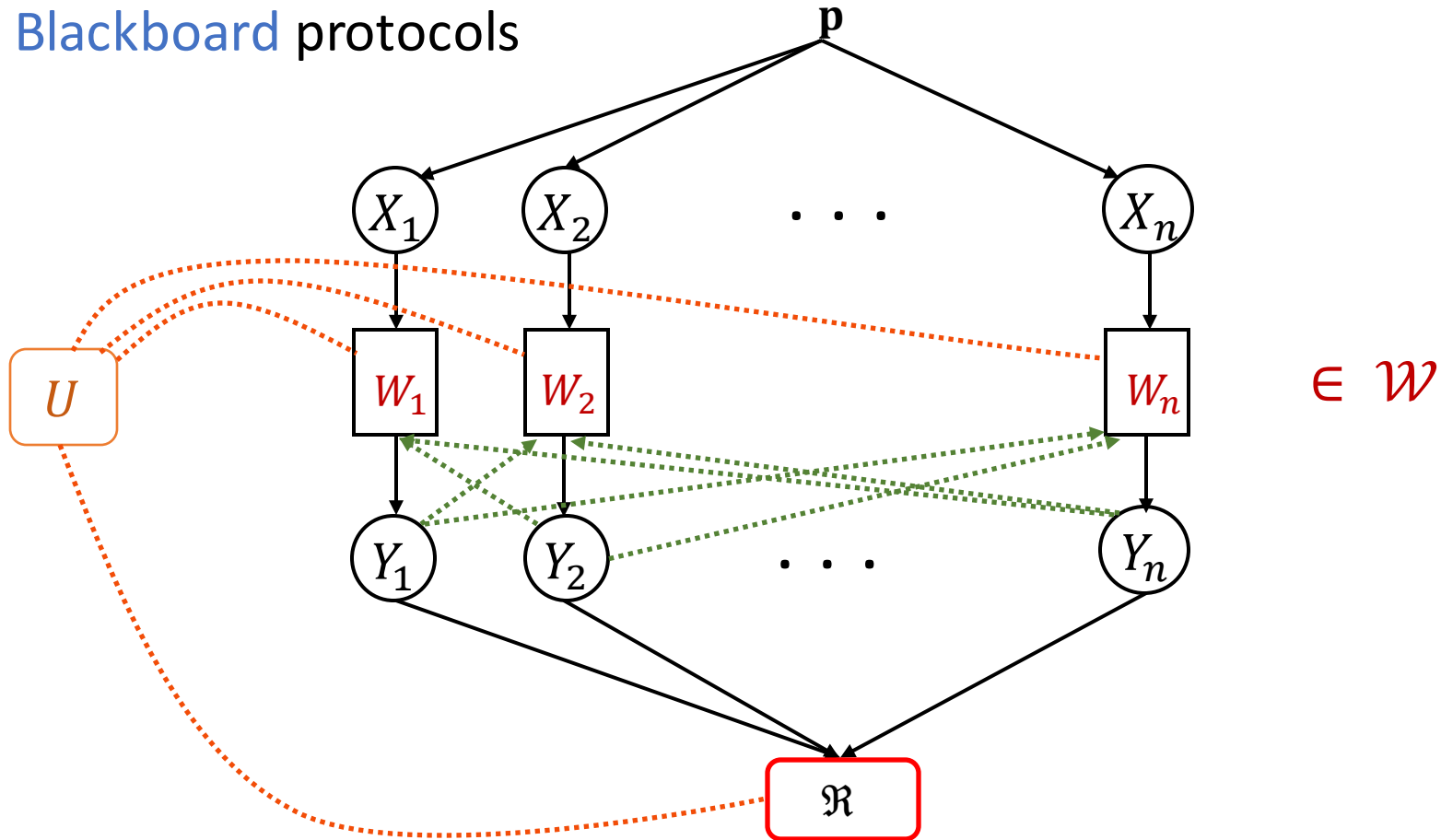
# Sequentially Interactive protocols



Interactive (“one-pass, sequential”),  
and common random seed

# Types of protocols

## Blackboard protocols



Fully interactive (“many passes”),  
and common random seed

# Types of protocols

Each of these models is **at least as powerful** as the previous

private-coin  $\preceq$  public-coin  $\preceq$  sequentially interactive  $\preceq$  blackboard

Each has its pros and cons (both in theory *and* practice), and may require different techniques to analyze.

# The Problems

Parameter/density  
estimation

Goodness-of-fit /  
Hypothesis testing

**Sample complexity:** smallest  $n$  to solve the task

# Example 1: Discrete distributions

$$\mathcal{P} = \Delta_d: \text{distbs on } [d] := \{1 \dots d\}$$

**Goal:** output  $\hat{\mathbf{p}}$  such that

$$\mathbb{E}[\text{TV}(\hat{\mathbf{p}}, \mathbf{p})] \leq \varepsilon$$

Sample complexity =  $\Theta\left(\frac{d}{\varepsilon^2}\right)$   
(without constraints)

**q:** a reference distribution

**Goal:** Test

$$\mathbf{p} = \mathbf{q} \text{ vs } \text{TV}(\mathbf{p}, \mathbf{q}) > \varepsilon$$

Sample complexity =  $\Theta\left(\frac{\sqrt{d}}{\varepsilon^2}\right)$   
(without constraints)

$$\text{TV}(\mathbf{p}, \mathbf{q}) := \sup_{S \subseteq [k]} (\mathbf{p}(S) - \mathbf{q}(S)) = \frac{1}{2} \ell_1(\mathbf{p}, \mathbf{q})$$

## Example 2: High dimensional distributions

$$\mathcal{P} = \{\mathcal{N}(\boldsymbol{\mu}, \mathbf{I}_d) : \boldsymbol{\mu} \in \mathbb{R}^d\}$$

**Goal:** output  $\hat{\boldsymbol{\mu}}$  such that

$$\mathbb{E}[|\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}|_2^2] \leq \varepsilon^2$$

Sample complexity =  $\Theta\left(\frac{d}{\varepsilon^2}\right)$   
(without constraints)

**Goal:** Test

$$\boldsymbol{\mu} = \mathbf{0} \text{ vs } |\boldsymbol{\mu}|_2 > \varepsilon$$

Sample complexity =  $\Theta\left(\frac{\sqrt{d}}{\varepsilon^2}\right)$   
(without constraints)

\*detecting signal vs noise

Other families: Bernoulli product

# Research goals

Establish sample complexity bounds for ...

- Different  $\mathcal{W}$ s
- Estimation/Testing/other properties
- Private-coin SMP/public-coin SMP/interactive
- Discrete/high-dimensional/non-parametric

Already a bit too much ... each interesting in its own right ... !



# For example ... discrete distribution testing

$\mathcal{W}_q$ , [AminJosephMao'20, BerrettButucea'20, AcharyaCanonneLiuSunTyagi'20]:

Private-coin SMP  $\ll$  public-coin SMP  $\approx$  SMP/interactive

$\mathcal{W}_\ell$ , [AcharyaCanonneLiuSunTyagi'20]:

Private-coin SMP  $\ll$  public-coin SMP  $\approx$  SMP/interactive

General  $\mathcal{W}$ , [AcharyaCanonneLiuSunTyagi'20]:

Private-coin SMP  $\ll$  public-coin SMP  $\ll$  SMP/interactive

Similarly for Gaussian mean testing ... [AcharyaCanonneTyagi'20, SzaboVuursteenVanZanten'20]

Parameter/density  
estimation

~~Goodness of fit/  
Hypothesis testing~~

Part 3 of tutorial ([link](#))

Learn about Ingster's method from HT!

Establishing tight results for SMP protocols generally easier ... why?

$Y_1, \dots, Y_n$  independent (given  $U$ )

See general discussion in

[ACLST20] J. Acharya, C. Canonne, Y. Liu, Z. Sun, H. Tyagi, “Interactive inference under information constraints” *arXiv: 2007.10976 (in submission)*

# Methods to establish interactive lower bounds

1. Cramer-Rao/van Trees inequality [\[BarnesHanOzgur19, BarnesChenOzgur20, SarbuZaidi21\]](#)
  - Unified results for  $\Delta_d, \mathcal{B}_d, \mathcal{G}_d$
  - Results hold for  $\ell_2$  loss
2. Strong Data Processing + Assouad's method [\[BravermanGardMaNguyenWoodruff16, DuchiRogers19\]](#)
  - Lower bounds for  $\mathcal{B}_d, \mathcal{G}_d$  under  $\ell_2$  loss
  - Naturally extends to other  $\ell_p$  loss functions
3. Chi-squared contractions + Assouad's method [\[AcharysCanonneLiuSunTyagi20, AcharyaCanonneSunTyagi20\]](#)
  - Unified bounds for  $\Delta_d, \mathcal{B}_d, \mathcal{G}_d$
  - Works under  $\ell_p$  for  $p \geq 1$
  - For arbitrary channels

# Pointers

Part 2 of tutorial at FOCS'20 ([link](#))

Cramer-Rao/van Trees inequality

Strong Data Processing + Assouad's method

# Next two parts ...

- Discrete distributions
  - Simulate and infer for upper bounds
  - Lower bounds
- Unified method for general distributions and channel families

# Part 2: Discrete Distributions

# Discrete distribution estimation

$\mathcal{P} = \Delta_d$ : distbs on  $[d] := \{1 \dots d\}$

**Goal:** output  $\hat{\mathbf{p}}$  such that

$$\mathbb{E}[\text{TV}(\hat{\mathbf{p}}, \mathbf{p})] \leq \varepsilon$$

Sample complexity =  $\Theta\left(\frac{d}{\varepsilon^2}\right)$  (without constraints)



# Empirical distribution works - DIY

$X_1, \dots, X_n \sim \mathbf{p}$ ,  $N_x := \# \text{ times } x \text{ appears}$

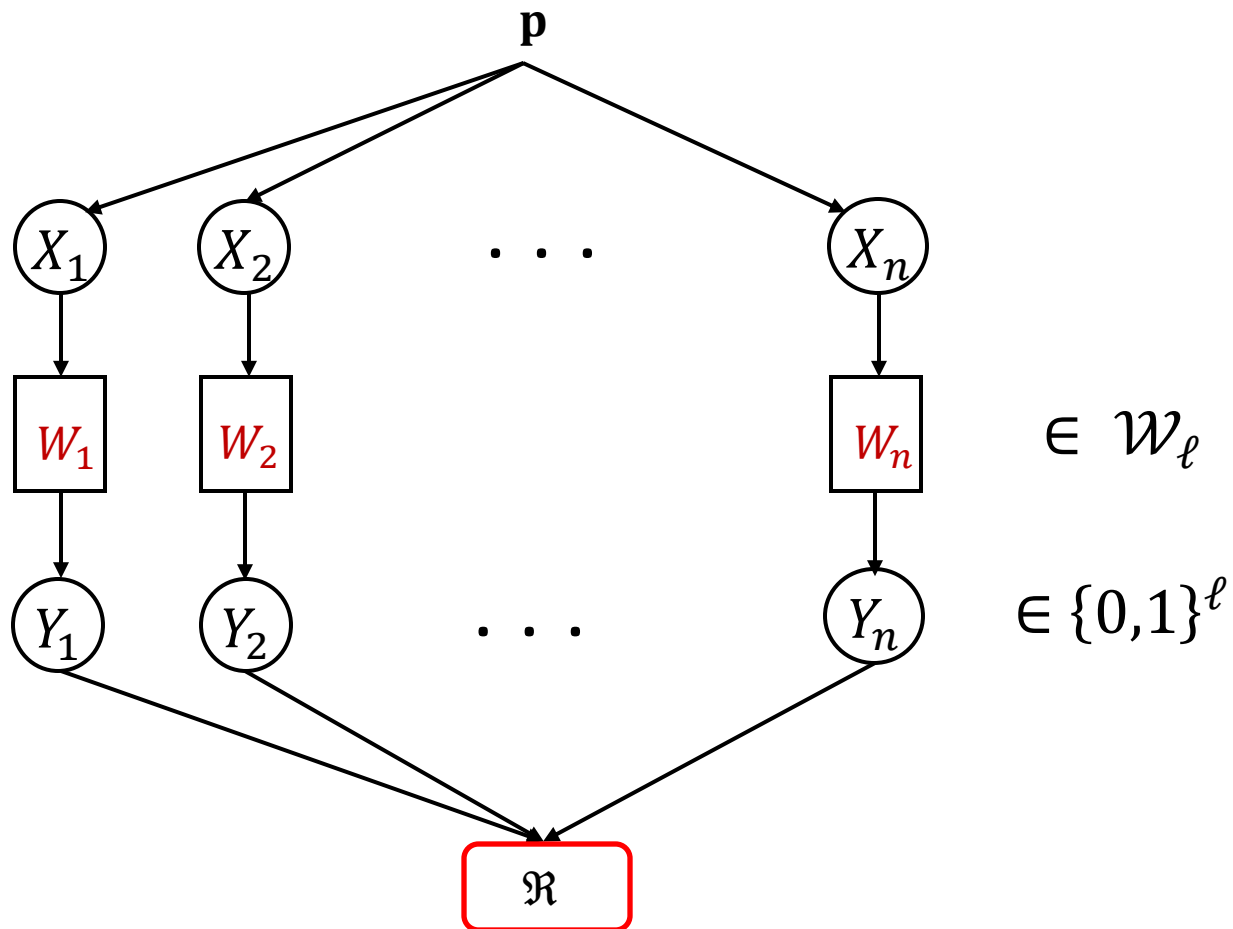
Empirical distribution:  $\hat{\mathbf{p}}(x) = N_x/n$

$N_x \sim \text{Bin}(n, \mathbf{p}(x))$

$$\mathbb{E} \left[ (\hat{\mathbf{p}}(x) - \mathbf{p}(x))^2 \right] = \frac{\mathbf{p}(x)(1 - \mathbf{p}(x))}{n} \Rightarrow \mathbb{E}[\ell_2^2(\hat{\mathbf{p}}, \mathbf{p})] \leq \frac{1}{n}$$

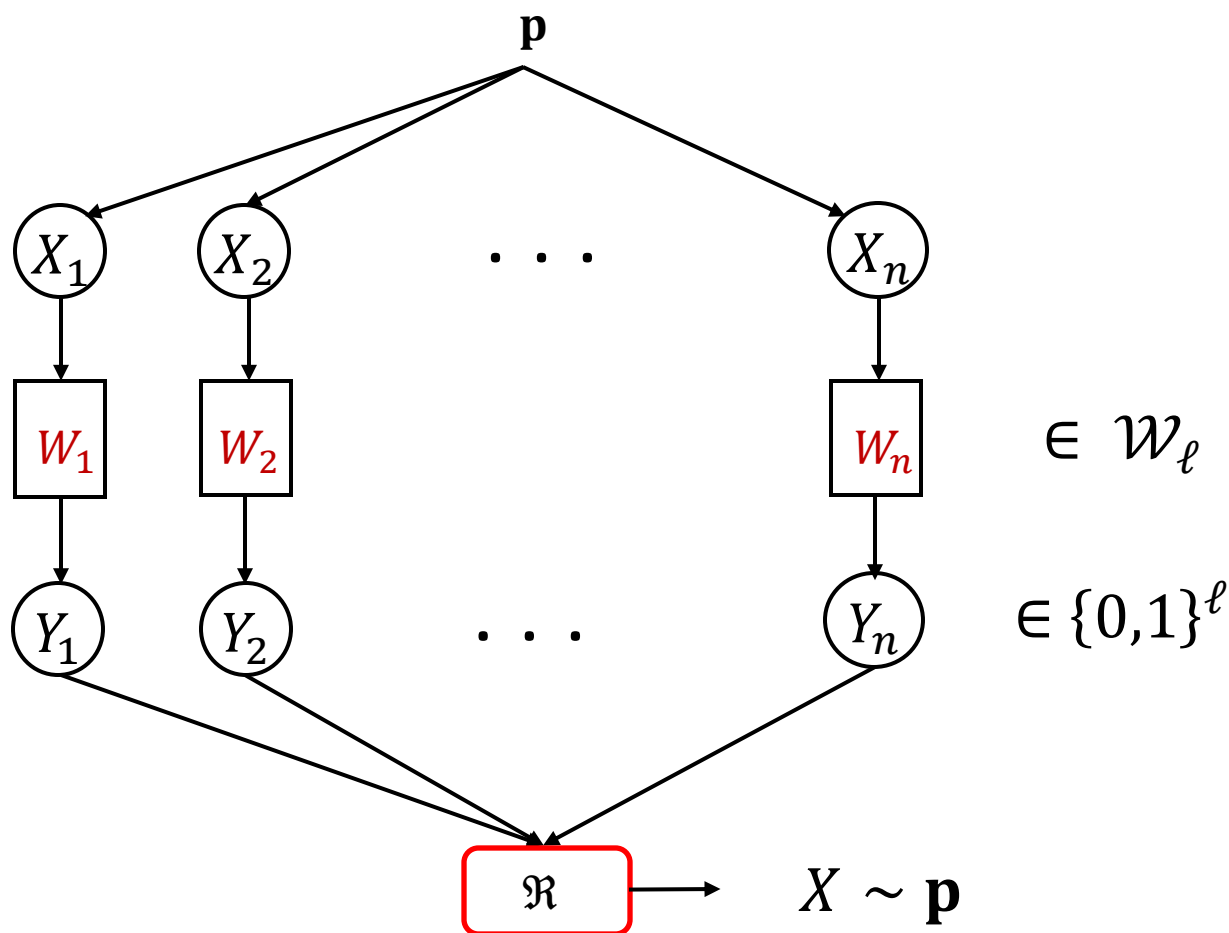
$$\begin{aligned} \mathbb{E}[\ell_1(\hat{\mathbf{p}}, \mathbf{p})]^2 &\leq \mathbb{E}[\ell_1(\hat{\mathbf{p}}, \mathbf{p})^2] && \text{(Jensen)} \\ &\leq d \cdot \mathbb{E}[\ell_2^2(\hat{\mathbf{p}}, \mathbf{p})] && \text{(Cauchy Schwarz)} \\ &\leq \frac{d}{n} \end{aligned}$$

# Under communication constraints



# A simulation puzzle ...

**Goal:** To simulate a sample from messages



# One simulation to solve them all ...

**Theorem.** Suppose **simulation** is possible with  $f(d, \ell)$  samples.

Let  $T$  be some task with **sample complexity**  $T(d, \varepsilon)$ .

Then  $T$  can be solved with  $f(d, \ell) \cdot T(d, \varepsilon)$  samples under  $\mathcal{W}_\ell$ .

What is  $f(d, \log_2 d) = ?$

# One simulation to solve them all ...

**Theorem.** There is a private-coin SMP protocol with

$$f(d, \ell) \approx \max \left\{ \frac{k}{2^\ell}, 1 \right\}.$$

No protocol (even interactive) can do better!

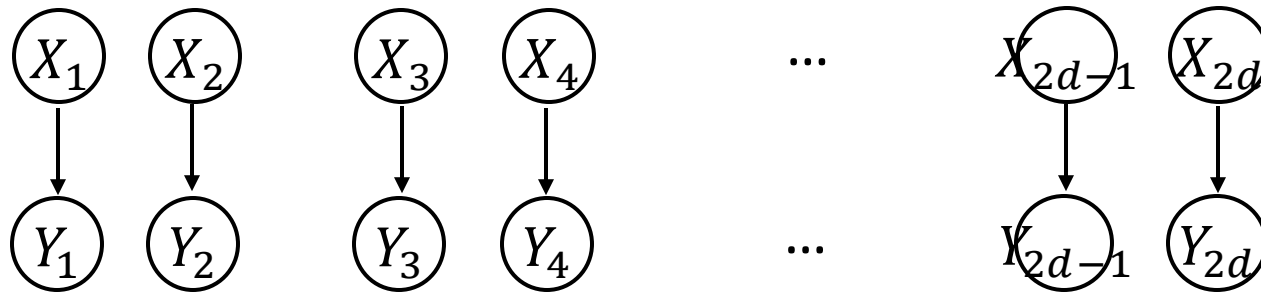
Estimation with  $\Theta \left( \frac{d}{\varepsilon^2} \cdot \frac{d}{2^\ell} \right)$  and testing with  $\Theta \left( \frac{\sqrt{d}}{\varepsilon^2} \cdot \frac{d}{2^\ell} \right)$

# Algorithm for one-bit

Take  $2d$  players:

- First two tell if their input is symbol 1
- Next two tell if their input is symbol 2
- And so on ...

# Algorithm for one-bit



$$Y_{2t-1} = I\{X_{2t-1} = t\}$$
$$Y_{2t} = I\{X_{2t} = t\}$$

# Algorithm for one-bit

- Output  $t$  if:
  - Player  $2t - 1$  is the **only** odd player sending 1
  - Player  $2t$  sends 0
- If no such  $i$ , output  $\perp$

Conditioned on not outputting  $\perp$ , a sample from  $p$



# Algorithm for one-bit

Player  $2t - 1$  is the **only** odd player sending 1

$$\Pr(Y_{2t-1} = 1, Y_{2t'-1} = 0 \text{ for } t' \neq t) = \mathbf{p}(t) \prod_{t' \neq t} (1 - \mathbf{p}(t'))$$

Player  $2t$  sends 0

$$\Pr(Y_{2t} = 0) = (1 - \mathbf{p}(t))$$

$$\Pr(\text{output } t \mid \text{not } \perp) = \mathbf{p}(t) \cdot \prod_{t' \neq t} (1 - \mathbf{p}(t')) \propto \mathbf{p}(t)$$

# Corollary

Inference Task	Centralized	One-bit <b>private-SMP</b>
Estimation	$\Theta\left(\frac{d}{\varepsilon^2}\right)$	$O\left(\frac{d^2}{\varepsilon^2}\right)$
Testing	$\Theta\left(\frac{\sqrt{d}}{\varepsilon^2}\right)$	$O\left(\frac{d^{3/2}}{\varepsilon^2}\right)$

# Corollary

Inference Task	Centralized	One-bit private-SMP	One-bit public-SMP
Estimation	$\Theta\left(\frac{d}{\varepsilon^2}\right)$	$\Theta\left(\frac{d^2}{\varepsilon^2}\right)$	$\Theta\left(\frac{d^2}{\varepsilon^2}\right)$
Testing: $I_u(k, \varepsilon)$	$\Theta\left(\frac{\sqrt{d}}{\varepsilon^2}\right)$	$\Theta\left(\frac{d^{3/2}}{\varepsilon^2}\right)$	$\Theta\left(\frac{d}{\varepsilon^2}\right)$

Bounds are tight ... simulate and infer is optimal for private-coin SMP

# Related work

Under SMP protocols these bounds are tight for communications  
[HanMukherjeeOzgur19, AcharyaCanonnetyagi'19] and LDP [DuchiJordanWainwright14]

## **Sample complexity with interactivity and general channels?**

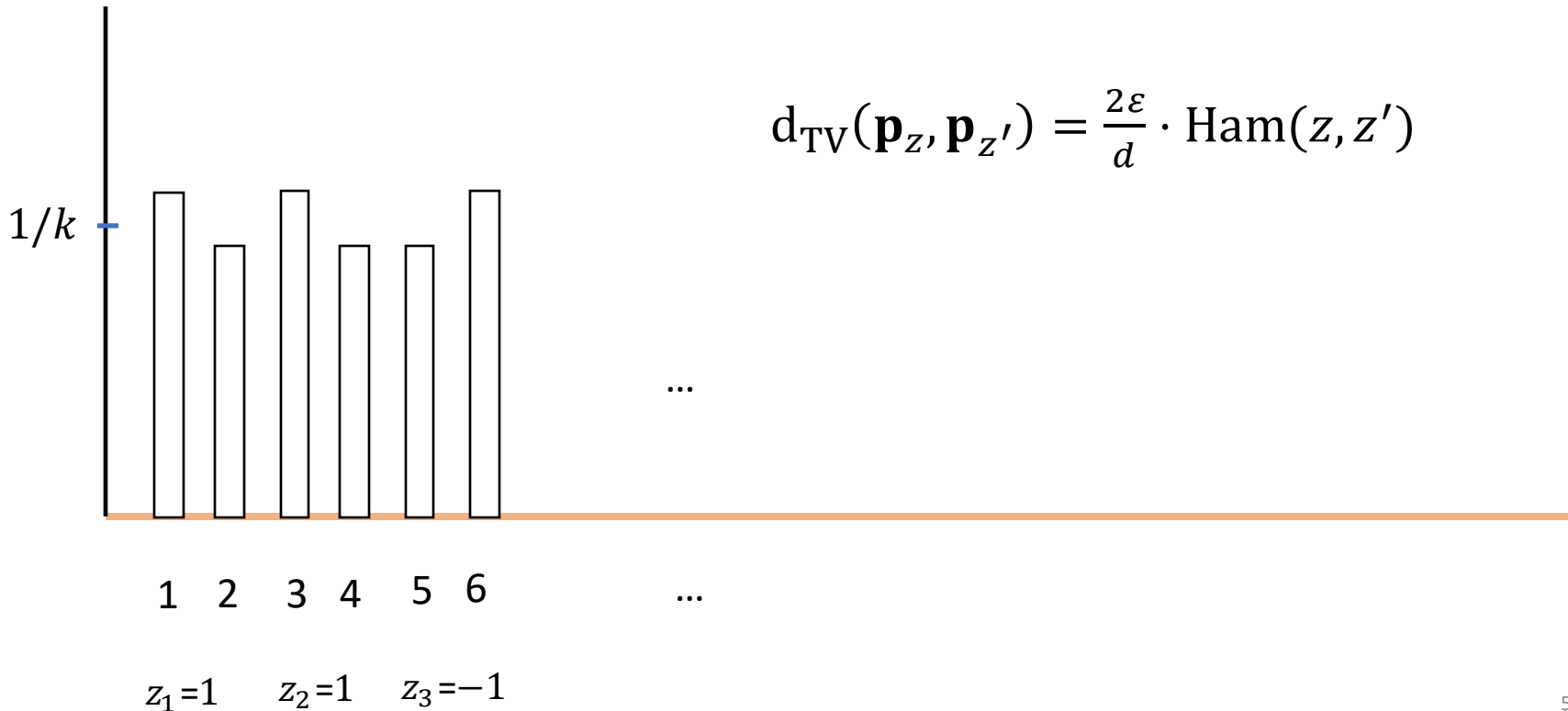
[ACLST20] J. Acharya, C. Canonne, Y. Liu, Z. Sun, H. Tyagi, “Interactive inference under information constraints” *arXiv: 2007.10976 (in submission)*

A hard instance

# A hard instance

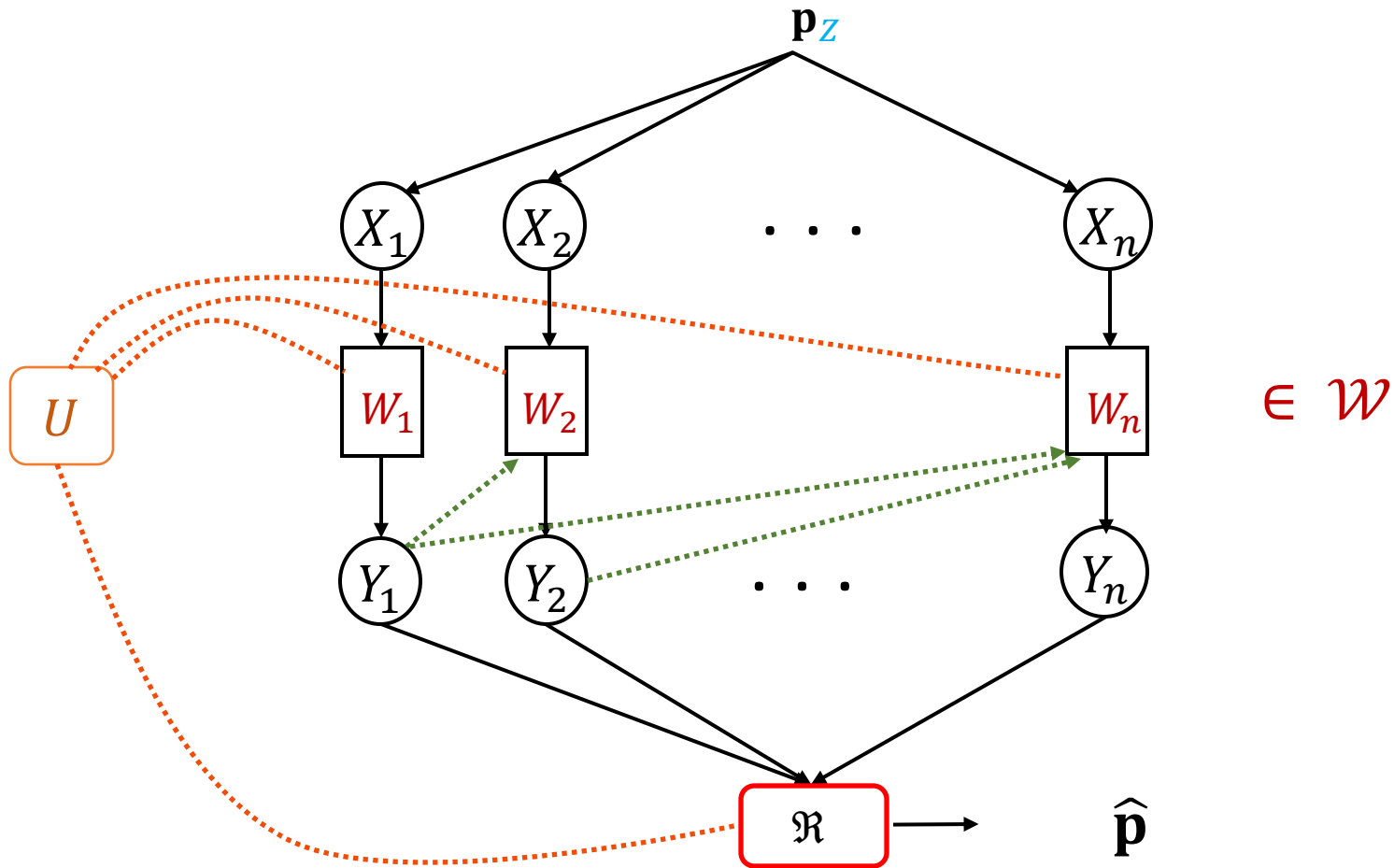
[Paninski'08] Let  $\mathcal{Z} = \{-1, 1\}^{d/2}$ , and  $\mathcal{P}_{\mathcal{Z}} = \{\mathbf{p}_z : z \in \mathcal{Z}\}$ , where

$$\mathbf{p}_z(2i-1) = \frac{1 + z_i \cdot 2\varepsilon}{d}, \quad \mathbf{p}_z(2i) = \frac{1 - z_i \cdot 2\varepsilon}{d}, \quad i = 1, \dots, d/2.$$



# Learning lower bounds

$Z = (Z_1, \dots, Z_{d/2}) \sim_{\text{uar}} \mathcal{Z}$ , ie, each  $Z_i \sim^{\text{iid}} \text{Bern}(0.5)$



# Learning lower bounds –

**Exercise:** Let  $z \in \mathcal{Z}$  and  $\hat{\mathbf{p}}$  satisfies  $d_{\text{TV}}(\hat{\mathbf{p}}, \mathbf{p}_z) < \frac{\varepsilon}{10}$ .

Then,

$$z^* = \arg \min_{z'} d_{\text{TV}}(\hat{\mathbf{p}}, \mathbf{p}_{z'})$$

satisfies

$$\text{Ham}(z, z^*) < \frac{d}{10}.$$



# Assouad's method

If we can estimate  $\mathbf{p}_Z \in_{\text{uar}} \mathcal{P}_Z$ , then we can estimate  $Z$ !

**Theorem.** Pick  $Z \sim_{\text{uar}} \mathcal{Z}$ .

If

$$\mathbb{E}_Z \left[ \mathbb{E}_{\mathbf{p}_Z} [\text{d}_{\text{TV}}(\hat{\mathbf{p}}(Y^n, U), \mathbf{p}_Z)] \right] < \frac{\varepsilon}{10}$$

then there exists an estimator  $\hat{Z}(Y^n, U)$  such that

$$\sum_{1 \leq i \leq d/2} \Pr(\hat{Z}_i = Z_i) > 0.8 \times \frac{d}{2}.$$

- **Note:** We could write this as  $\sum_i I(Z_i \wedge Y^n | U) = \Omega(d)$

# Assouad's method

Exercise. If

$$\sum_{1 \leq i \leq d/2} \Pr(\hat{Z}_i = Z_i) > 0.8 \times \frac{d}{2},$$

then there exists a subset  $S \subseteq \{1, \dots, d/2\}$  with  $|S| > d/6$  s.t. if  $i \in S$ ,

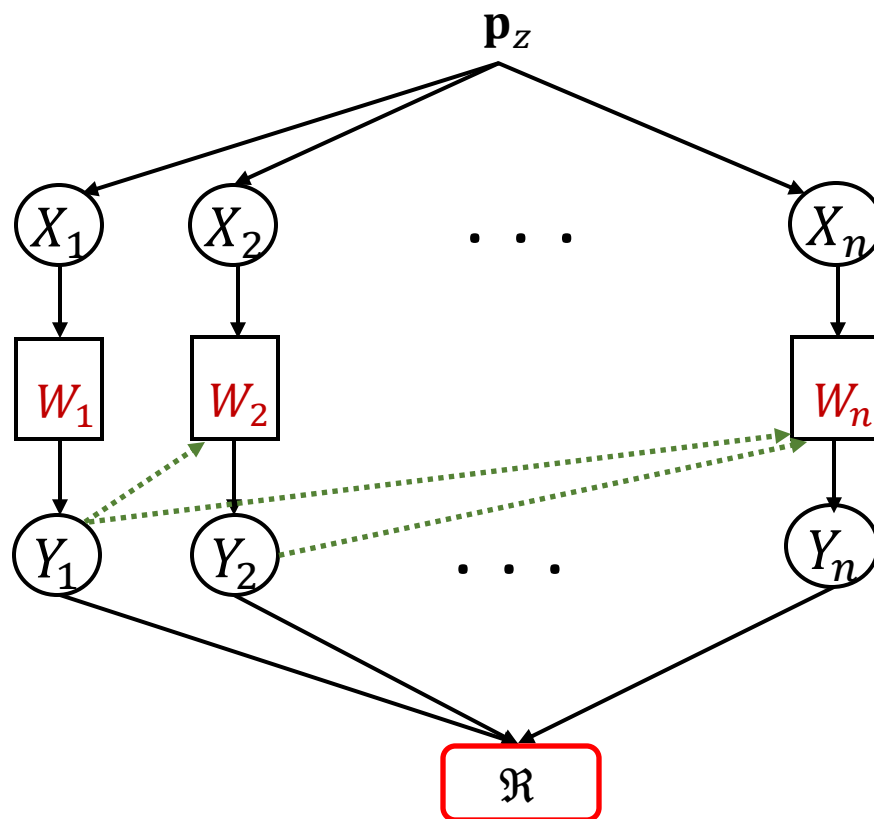
$$\Pr(\hat{Z}_i = Z_i) > 0.7.$$

Now we need a lower bound on  $n$  for this to happen

# Notation

Fix  $i \in [d/2]$ , when can we figure  $Z_i$ ?

$\mathbf{p}_Z^{Y^n}$ : distribution of  $Y^n$  when input distribution  $\mathbf{p}_Z$



# Information bound on one coordinate

average output distribution fixing  $Z_i = \pm 1$ :

When  $Z_i = 1$ :  $\mathbf{p}_{+i}^{Y^n} := \frac{1}{2^{d/2-1}} \sum_{\mathbf{z}: z_i = +1} \mathbf{p}_{\mathbf{z}}^{Y^n}$

When  $Z_i = -1$ :  $\mathbf{p}_{-i}^{Y^n} := \frac{1}{2^{d/2-1}} \sum_{\mathbf{z}: z_i = -1} \mathbf{p}_{\mathbf{z}}^{Y^n}$

If we can guess  $Z_i$  from  $Y^n$

$\Leftrightarrow d_{\text{TV}}(\mathbf{p}_{+i}^{Y^n}, \mathbf{p}_{-i}^{Y^n})$  must be large

$\Rightarrow$  bound distance between  $\mathbf{p}_{+i}^{Y^n}$  and  $\mathbf{p}_{-i}^{Y^n}$

# Total variation and hypothesis testing

$\mathbf{p}_1, \mathbf{p}_2$  be any two distributions over  $\mathcal{Y}$

$j \in \{0,1\}$  be picked at random

Given  $Y \sim \mathbf{p}_j$ , design a  $\hat{j}(Y)$  that is a guess for  $j$

For any  $\hat{j}(Y)$ :

$$\Pr(\hat{j}(Y) = j) \leq \frac{1}{2} (1 + d_{\text{TV}}(\mathbf{p}_1, \mathbf{p}_2))$$

# Information bound on one coordinate

In our case,  $\mathbf{p}_1 = \mathbf{p}_{+i}^{Y^n}$ ,  $\mathbf{p}_2 = \mathbf{p}_{-i}^{Y^n}$ , and

$$\Pr(\hat{Z}_i = Z_i) > 0.7 \Rightarrow d_{\text{TV}}(\mathbf{p}_{+i}^{Y^n}, \mathbf{p}_{-i}^{Y^n}) \geq 0.4$$

Since this holds for at least  $d/6$  coordinates,

$$\sum_i d_{\text{TV}}(\mathbf{p}_{+i}^{Y^n}, \mathbf{p}_{-i}^{Y^n})^2 \geq \frac{d}{6} \times 0.16.$$

# Some ingredients

$$D(\mathbf{p}_1 || \mathbf{p}_2) := \sum_y \mathbf{p}_1(y) \log \frac{\mathbf{p}_1(y)}{\mathbf{p}_2(y)}, \chi^2(\mathbf{p}_1, \mathbf{p}_2) := \sum_y \frac{(\mathbf{p}_1(y) - \mathbf{p}_2(y))^2}{\mathbf{p}_2(y)}$$

Pinsker's inequality, convexity of logarithms:

$$2 \cdot d_{\text{TV}}(\mathbf{p}_1, \mathbf{p}_2)^2 \leq D(\mathbf{p}_1 || \mathbf{p}_2) \leq \chi^2(\mathbf{p}_1, \mathbf{p}_2)$$

Chain rule of KL divergence: If  $\mathbf{p}_1$  and  $\mathbf{p}_2$  are over  $\mathcal{Y}_1 \times \mathcal{Y}_2$ :

$$\begin{aligned} & D(\mathbf{p}_1(Y_1, Y_2) || \mathbf{p}_2(Y_1, Y_2)) \\ &= D(\mathbf{p}_1(Y_1) || \mathbf{p}_2(Y_1)) + \mathbb{E}_{Y_1} [D(\mathbf{p}_1(Y_2 | Y_1) || \mathbf{p}_2(Y_2 | Y_1))] \end{aligned}$$

## KL $\leq$ chi-squared (DIY)

Since  $\log(1 + x) \leq x$  (why?)

$$\begin{aligned} D(\mathbf{p}||\mathbf{q}) &:= \sum_x \mathbf{p}(x) \log \left( 1 + \frac{\mathbf{p}(x) - \mathbf{q}(x)}{\mathbf{q}(x)} \right) \\ &\leq \sum_x \mathbf{p}(x) \frac{(\mathbf{p}(x) - \mathbf{q}(x))}{\mathbf{q}(x)} = \chi^2(\mathbf{p}, \mathbf{q}) \end{aligned}$$

**Exercise:** Prove the chain rule of KL.



# Why go to KL?

By Pinsker's inequality,

$$4 \cdot d_{\text{TV}}(\mathbf{p}_{+i}^{Y^n}, \mathbf{p}_{-i}^{Y^n})^2 \leq \left( D(\mathbf{p}_{+i}^{Y^n} \parallel \mathbf{p}_{-i}^{Y^n}) + D(\mathbf{p}_{-i}^{Y^n} \parallel \mathbf{p}_{+i}^{Y^n}) \right)$$

Summing over  $i$ ,

$$\begin{aligned} & \sum_i \left( D(\mathbf{p}_{+i}^{Y^n} \parallel \mathbf{p}_{-i}^{Y^n}) + D(\mathbf{p}_{-i}^{Y^n} \parallel \mathbf{p}_{+i}^{Y^n}) \right) \\ & \geq \sum_i 4 \cdot d_{\text{TV}}(\mathbf{p}_{+i}^{Y^n}, \mathbf{p}_{-i}^{Y^n})^2 \geq 4 \cdot \frac{d}{6} \times 0.16 \geq \frac{d}{10} \end{aligned}$$

$\mathbf{p}_{+i}^{Y^n}$  are **mixture distributions**!

Handling mixtures is painful, leads to **issues** to extend SMP lower bounds to interactive setting

# Convexity to the rescue

**Exercise:** KL divergence is convex.

For any distributions  $\mathbf{p}_1, \mathbf{p}_2$  and  $\mathbf{q}_1, \mathbf{q}_2$  and  $\lambda \in [0,1]$ ,

$$\begin{aligned} &D(\lambda\mathbf{p}_1 + (1 - \lambda)\mathbf{q}_1 || \lambda\mathbf{p}_2 + (1 - \lambda)\mathbf{q}_2) \\ &\leq \lambda \cdot D(\mathbf{p}_1 || \mathbf{p}_2) + (1 - \lambda) \cdot D(\mathbf{q}_1 || \mathbf{q}_2) \end{aligned}$$

Prove using concavity of logarithms

# Convexity to handle mixtures

$z \in \{-1, 1\}^{k/2}$ ,  $z^{\oplus i}$  obtained by flipping the  $i$ th coordinate of  $z$

**Theorem.**

$$\frac{1}{2} \left( D(\mathbf{p}_{+i}^{Y^n} || \mathbf{p}_{-i}^{Y^n}) + D(\mathbf{p}_{-i}^{Y^n} || \mathbf{p}_{+i}^{Y^n}) \right) \leq \mathbb{E}_Z [D(\mathbf{p}_Z^{Y^n} || \mathbf{p}_{Z^{\oplus i}}^{Y^n})]$$

**Proof.** Convexity of divergence to the definitions of  $\mathbf{p}_{+i}^{Y^n}$  and  $\mathbf{p}_{-i}^{Y^n}$  ■

Information about  $Z_i$  bounded by average divergence in message distribution upon **changing only**  $Z_i$  when all others are fixed!

# Convexity to handle mixtures

Summing over  $i$

$$\frac{d}{20} \leq \mathbb{E}_Z \left[ \sum_i D(\mathbf{p}_Z^{Y^n} || \mathbf{p}_{Z \oplus i}^{Y^n}) \right]$$

What do we have here ... Fix a  $Z$  and then change one coordinate at a time ...

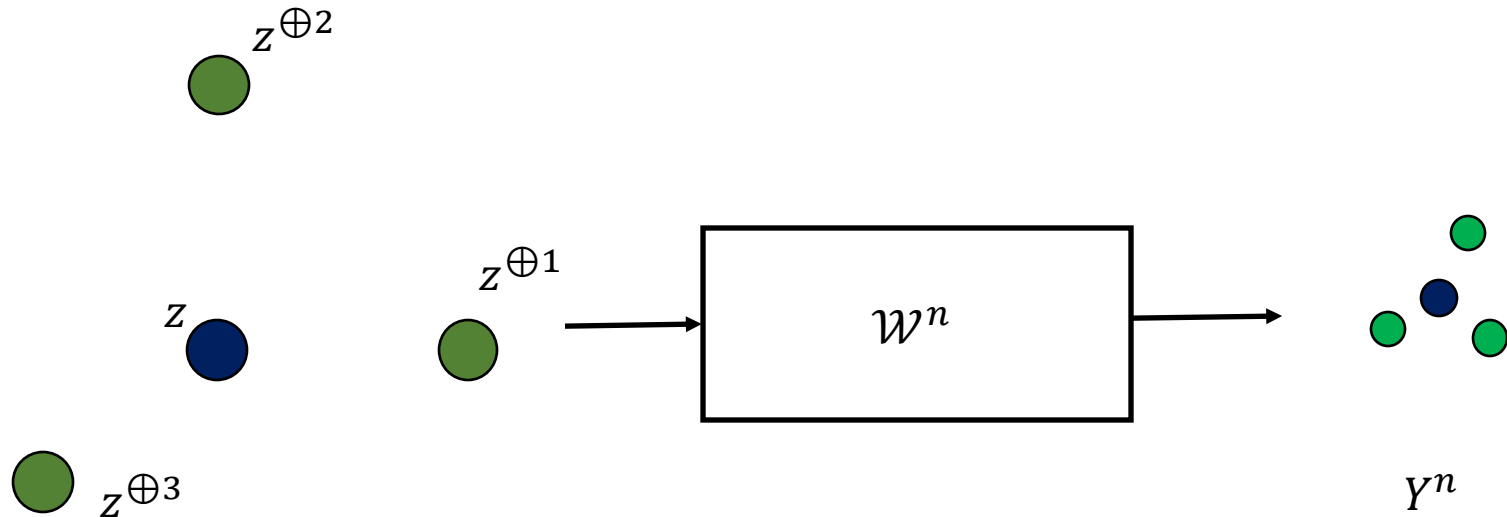
# Focus on one $z$

By expectation<max, and linearity of expectations,

$$\frac{d}{20} \leq \max_z \left[ \sum_i D(p_z^{Y^n} || p_{z \oplus i}^{Y^n}) \right]$$

\*\* the following is the original bound in terms of MI:

$$\sum_i I(Z_i \wedge Y^n) \leq \frac{1}{2} \cdot \max_z \left[ \sum_i D(p_z^{Y^n} || p_{z \oplus i}^{Y^n}) \right]$$

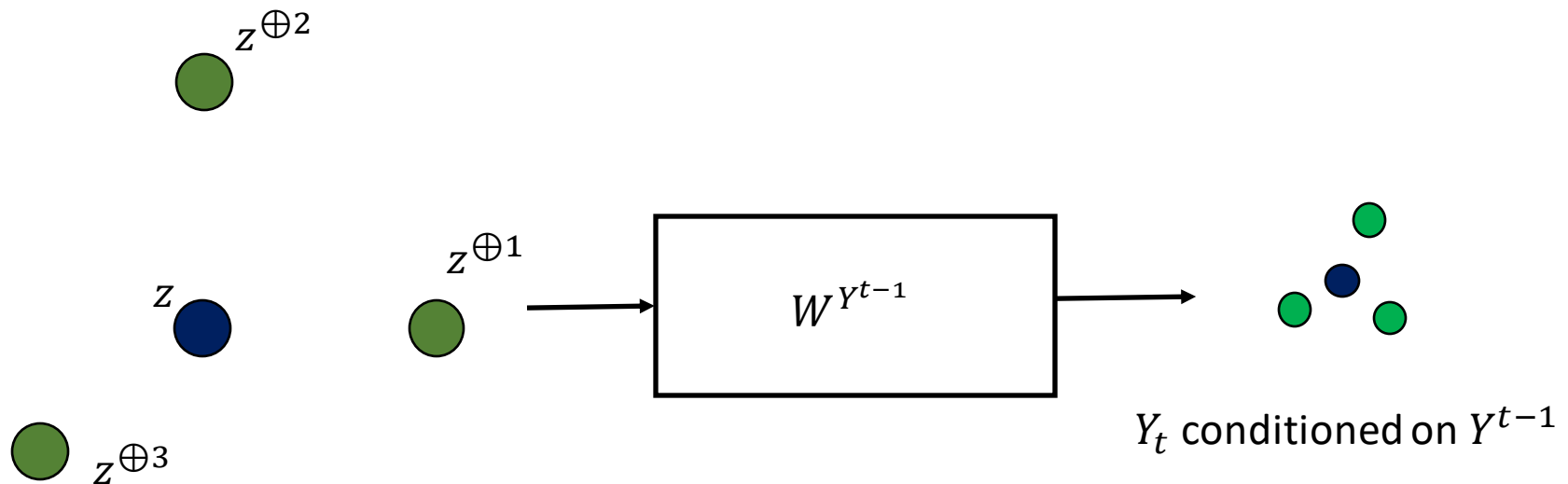


# Bounding $\sum_i D(\mathbf{p}_z^{Y^n} \parallel \mathbf{p}_{z \oplus i}^{Y^n})$

By the chain rule of divergence

$$\sum_i D(\mathbf{p}_z^{Y^n} \parallel \mathbf{p}_{z \oplus i}^{Y^n}) = \sum_t \mathbb{E}_{\mathbf{p}_z^{Y^{t-1}}} \left[ \sum_i D(\mathbf{p}_z^{Y_t|Y^{t-1}} \parallel \mathbf{p}_{z \oplus i}^{Y_t|Y^{t-1}}) \right].$$

- $\mathbf{p}_z^{Y_t|Y^{t-1}}$ : Distribution of  $Y_t$  with input  $\mathbf{p}_z$  conditioned on  $Y^{t-1}$
- Channel at player  $t$  a function only of  $Y^{t-1}$ , denoted  $W^{Y^{t-1}}$

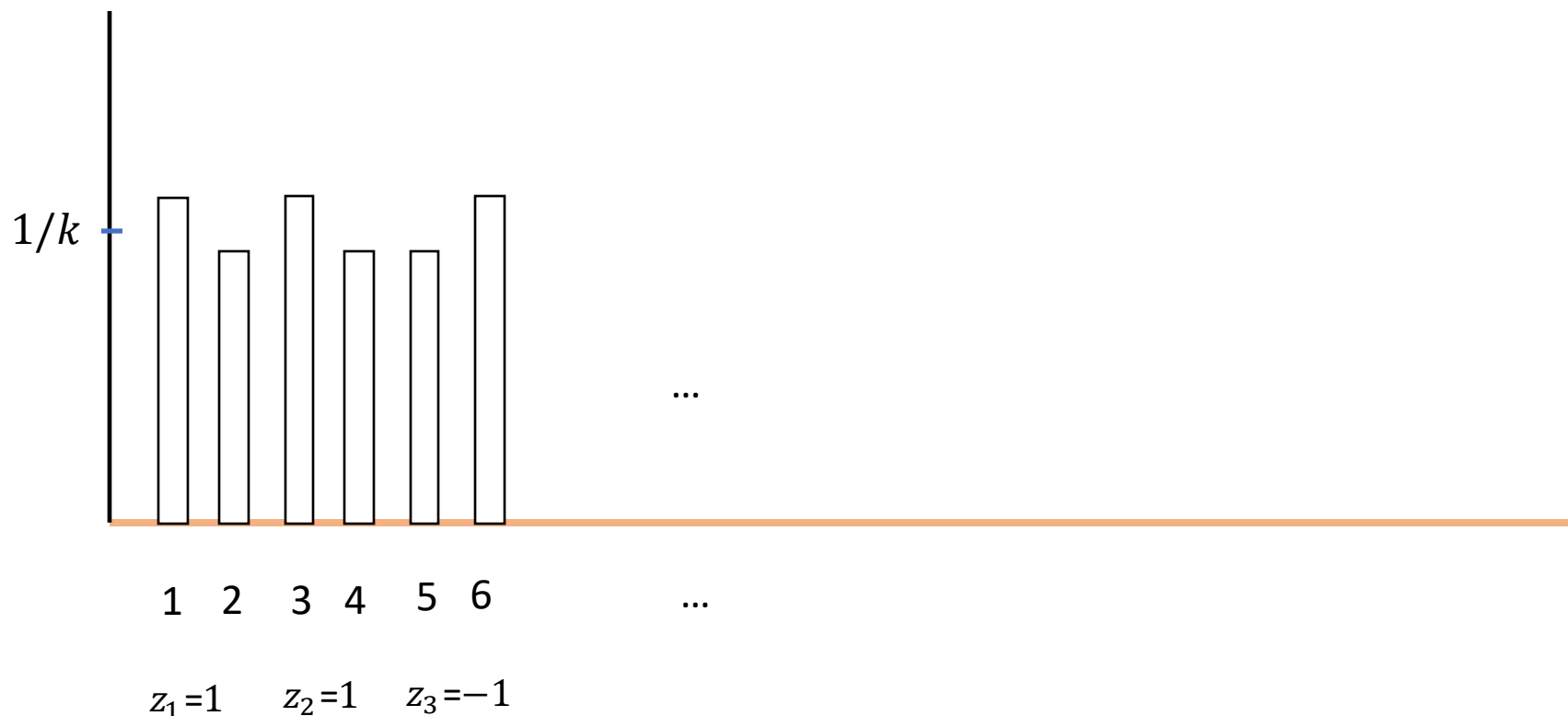


# Recall

For  $z \in \{-1, 1\}^{k/2}$ ,

$$\mathbf{p}_z(2i-1) = \frac{1 + z_i \varepsilon}{k}, \quad \mathbf{p}_z(2i) = \frac{1 - z_i \varepsilon}{k}, \quad i = 1, \dots, k/2.$$

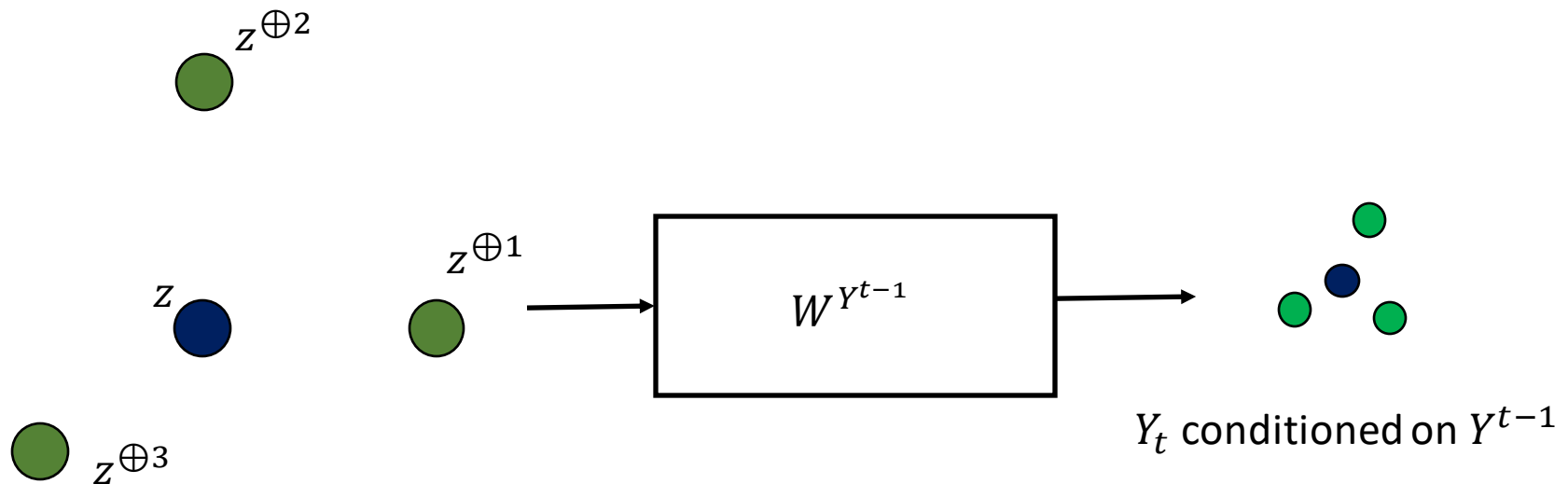
$\mathbf{p}_z$  and  $\mathbf{p}_{z \oplus i}$  differ **only on**  $2i-1$  and  $2i$



# Bounding $\sum_i D \left( \mathbf{p}_z^{Y_t|Y^{t-1}} || \mathbf{p}_{z \oplus i}^{Y_t|Y^{t-1}} \right)$

- Fix  $Y^{t-1}$

$$\mathbf{p}_z^{Y_t|Y^{t-1}}(y) = \mathbf{p}_{z \oplus i}^{Y_t|Y^{t-1}}(y) + \frac{2\varepsilon z_i}{k} \left( W^{Y^{t-1}}(y|2i-1) - W^{Y^{t-1}}(y|2i) \right)$$





# Bounding $\sum_i D \left( \mathbf{p}_Z^{Y_t|Y^{t-1}} \parallel \mathbf{p}_{Z \oplus i}^{Y_t|Y^{t-1}} \right)$

Since  $\text{KL} \leq \chi^2$ , plugging the expression above

$$\begin{aligned} \sum_i D \left( \mathbf{p}_Z^{Y_t|Y^{t-1}} \parallel \mathbf{p}_{Z \oplus i}^{Y_t|Y^{t-1}} \right) &\leq \sum_i \sum_y \frac{\left( \mathbf{p}_Z^{Y_t}(y) - \mathbf{p}_{Z \oplus i}^{Y_t}(y) \right)^2}{\mathbf{p}_{Z \oplus i}^{Y_t}(y)} \\ &\leq \frac{8\varepsilon^2}{k} \cdot \sum_i \sum_y \frac{\left( W(y|2i-1) - W(y|2i) \right)^2}{\sum_x W(y|x)} \end{aligned}$$

# An average information bound

**Theorem.**

$$\sum_i I(\mathbf{Z}_i \wedge Y^n) \leq n \cdot \frac{8\varepsilon^2}{d} \cdot \sup_{W \in \mathcal{W}} \sum_i \sum_y \frac{(W(y|2i-1) - W(y|2i))^2}{\sum_x W(y|x)}$$

Recall

$$\mathbf{p}_{\mathbf{Z}}(2i-1) = \frac{1 + \mathbf{Z}_i \varepsilon}{d}, \quad \mathbf{p}_{\mathbf{Z}}(2i) = \frac{1 - \mathbf{Z}_i \varepsilon}{d}$$

$|W(y|2i-1) - W(y|2i)|$  large  $\Leftrightarrow$  seeing  $y$  tells about input  
 $\Leftrightarrow$  tells about  $\mathbf{Z}_i$

# An average information bound

**Theorem.** [ACLT20] Under any **interactive protocol**,

$$\sum_i I(Z_i \wedge Y^n) \leq n \cdot \frac{8\varepsilon^2}{k} \cdot \sup_{W \in \mathcal{W}} \sum_i \sum_y \frac{(W(y|2i-1) - W(y|2i))^2}{\sum_x W(y|x)}$$

**Theorem.** If there exists an estimator then

$$\frac{d}{20} \leq n \cdot \frac{8\varepsilon^2}{k} \cdot \sup_{W \in \mathcal{W}} \sum_i \sum_y \frac{(W(y|2i-1) - W(y|2i))^2}{\sum_x W(y|x)}$$

# Applications

For any  $W \in \mathcal{W}_\ell$

$$\sum_i \sum_y \frac{(W(y|2i-1) - W(y|2i))^2}{\sum_x W(y|x)} \leq 2^\ell$$



For any  $W \in \mathcal{W}_\varrho$ ,  $\varrho \leq 1$

$$\sum_i \sum_y \frac{(W(y|2i-1) - W(y|2i))^2}{\sum_x W(y|x)} = o(\varrho^2)$$

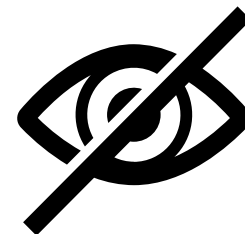


# Interactive lower bound for estimation

$$\frac{d}{20} \leq n \cdot \frac{8\varepsilon^2}{d} \cdot 2^\ell$$
$$n = \Omega\left(\frac{d^2}{2^\ell \varepsilon^2}\right)$$



$$\frac{d}{20} \leq n \cdot \frac{8\varepsilon^2}{d} \cdot \varrho^2$$
$$n = \Omega\left(\frac{d^2}{\varepsilon^2 \varrho^2}\right)$$



# Plug-n-play bounds

$H(W)$  is a  $\frac{d}{2} \times \frac{d}{2}$  PSD matrix:

$$(H(W))_{ij} := \sum_{y \in Y} \frac{(W(y|2i-1) - W(y|2i))(W(y|2j-1) - W(y|2j))}{\sum_j W(y|j)}$$

$$\sum_i \sum_y \frac{(W(y|2i-1) - W(y|2i))^2}{\sum_x W(y|x)} = \|H(W)\|_*$$

# Plug-n-play bounds

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

Testing:

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

Estimation

Classic	Sequentially Interactive
$\Omega\left(\frac{d}{\varepsilon^2}\right)$	$\Omega\left(\frac{d^2}{\varepsilon^2 \  \mathcal{W} \ _*}\right)$



Next 45 minutes:

**Reinforcement Learning** by Himanshu Tyagi ...

# References (click to go)

▼ 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; Zuteng 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  $E_\gamma$ -Divergence.** Shahab Asodeh; 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](#) ▼

# References

∨ 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](#) ∨

# References

- [ACT'18] J. Acharya, C. Canonne, H. Tyagi, “Distributed Simulation and Distributed Inference”, arxiv
- [ACT'18] J. Acharya, C. Canonne, H. Tyagi, “Inference under Information Constraints I: Lower bounds from chi-squared contractions”, arxiv
- [ACFT'18] J. Acharya, C. Canonne, C. Freitag, H. Tyagi, “Test without Trust: Optimal Locally Private Distribution Testing”, arxiv
- [DGL+'17] I. Diakonikolas, E. Grigorescu, J. Li, A. Natarajan, K. Onak, L. Schmidt “Communication-Efficient Distributed Learning of Discrete Distributions” NIPS
- [EPR'11] Úlfar Erlingsson, Vasyl Pihur, Aleksandra Korolova, “RAPPOR: Randomized Aggregatable Privacy-Preserving Ordinal Response”
- [HOW'18] Y. Han, A. Ozgur, T. Weissman, “Geometric Lower Bounds for Distributed Parameter Estimation under Communication Constraints”, COLT
- [Paninski'08] Liam Paninski, “A Coincidence-Based Test for Uniformity Given Very Sparsely Sampled Discrete Data”, IEEE Transactions on Information Theory

## Some references and previous work





## Some references and previous work



(not in that order)

# Some references and previous work

Too many for a single slide, or two. Starts, more or less, with Tsitsiklis'89, picks up again in the mid-2000's with a slightly different focus: local privacy, various types of communication constraints, ML-related motivations...

For a detailed bibliography:

[www.cs.columbia.edu/~ccanonne/tutorial-focs2020/bibliography.html](http://www.cs.columbia.edu/~ccanonne/tutorial-focs2020/bibliography.html)



THE END >>>>