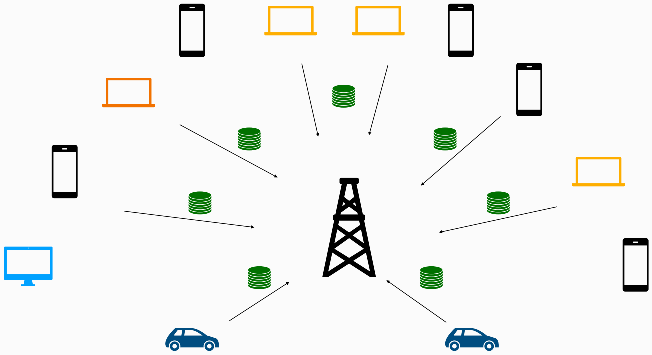
Domain compression: a new primitive

Paper: Domain Compression and its Application to Randomness-Optimal Distributed Goodness-of-Fit

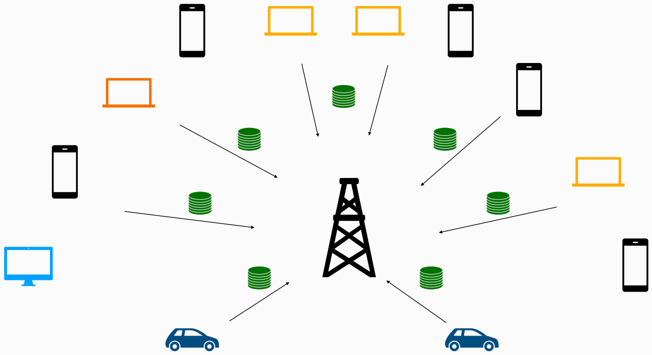
COLT 2020, everywhere on earth

Jayadev Acharya, Cornell University
Clément Canonne, IBM Research
Yanjun Han, Stanford University
Ziteng Sun, Cornell University
Himanshu Tyagi, IISc Bangalore

Distributed Hypothesis Testing

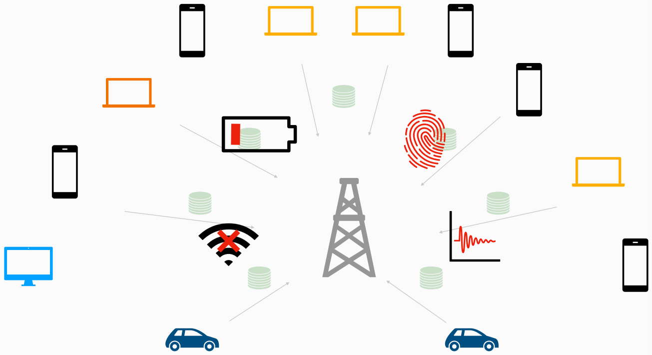


Distributed Hypothesis Testing

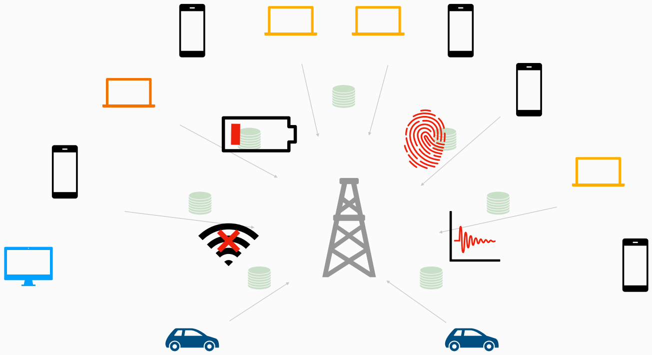


Does the data satisfy a postulated hypothesis/property?

Distributed Hypothesis Testing



Distributed Hypothesis Testing



Only constrained observations are available.

Identity Testing (IT), Goodness of Fit

- $\mathcal{X} = [k] := \{0, 1, 2, ..., k-1\}$, a discrete set of size k.

Identity Testing (IT), Goodness of Fit

- $\mathcal{X} = [k] := \{0, 1, 2, ..., k-1\}$, a discrete set of size k.
- q: a **known** reference distribution.

Identity Testing (IT), Goodness of Fit

- $\mathcal{X} = [k] := \{0, 1, 2, ..., k-1\}$, a discrete set of size k.
- q: a **known** reference distribution.
- Given $X^n = X_1 \ldots X_n$ independent samples from **unknown** p.

Identity Testing (IT), Goodness of Fit

- $\mathcal{X} = [k] := \{0, 1, 2, ..., k-1\}$, a discrete set of size k.
- q: a **known** reference distribution.
- Given $X^n = X_1 \dots X_n$ independent samples from **unknown** p.
- **Is $p = q$?**

Identity Testing (IT), Goodness of Fit

- $\mathcal{X} = [k] := \{0, 1, 2, ..., k-1\}$, a discrete set of size k.
- q: a **known** reference distribution.
- Given $X^n = X_1 \dots X_n$ independent samples from **unknown** p.
- **Is** $p = q$**?**
- Test: $\mathcal{A}\colon [k]^n \to \{0,1\}$, which satisfies the following:

With probability at least $1-\delta$,

$$\mathcal{A}(X^n) = \begin{cases} 1, & \text{if } p = q \\ 0, & \text{if } \|p-q\|_{\mathrm{TV}} > \varepsilon \end{cases}$$

Identity Testing (IT), Goodness of Fit

- $\mathcal{X} = [k] := \{0, 1, 2, ..., k-1\}$, a discrete set of size k.
- q: a **known** reference distribution.
- Given $X^n = X_1 \dots X_n$ independent samples from **unknown** p.
- **Is** $p = q$**?**
- Test: $\mathcal{A}\colon [k]^n \to \{0, 1\}$, which satisfies the following:

With probability at least 2/3,

$$\mathcal{A}(X^n) = \begin{cases} 1, & \text{if } p = q \\ 0, & \text{if } \|p - q\|_{\mathrm{TV}} > \varepsilon \end{cases}$$

Identity Testing (IT), Goodness of Fit

- $\mathcal{X} = [k] := \{0, 1, 2, ..., k-1\}$, a discrete set of size k.
- q: a **known** reference distribution.
- Given $X^n = X_1 \ldots X_n$ independent samples from **unknown** p.
- **Is** $p = q$**?**
- Test: $\mathcal{A} \colon [k]^n \to \{0, 1\}$, which satisfies the following:

With probability at least 2/3,

$$\mathcal{A}(X^n) = \begin{cases} 1, & \text{if } p = q \\ 0, & \text{if } \|p - q\|_{\mathrm{TV}} > \varepsilon \end{cases}$$

Sample complexity: Smallest n for which such a test exists.

Identity Testing (IT), Goodness of Fit

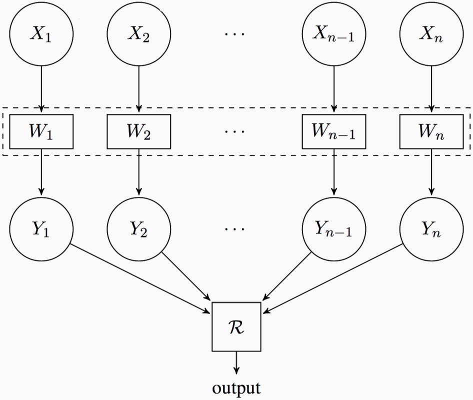
- $\mathcal{X} = [k] := \{0, 1, 2, ..., k-1\}$, a discrete set of size k.
- q: a **known** reference distribution.
- Given $X^n = X_1 \ldots X_n$ independent samples from **unknown** p.
- **Is** $p = q$**?**
- Test: $\mathcal{A} \colon [k]^n \to \{0, 1\}$, which satisfies the following:

> With probability at least 2/3,
> $$\mathcal{A}(X^n) = \begin{cases} 1, & \text{if } p = q \\ 0, & \text{if } \|p - q\|_{\mathrm{TV}} > \varepsilon \end{cases}$$

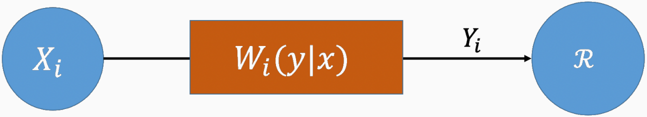
$$\Theta\left(\sqrt{k}/\varepsilon^2\right).$$

Simultaneous Message Passing (SMP) Protocol

Observations $Y_i = W_i(X_i) \in \mathcal{Y}$. $W_i \in \mathcal{W}$.

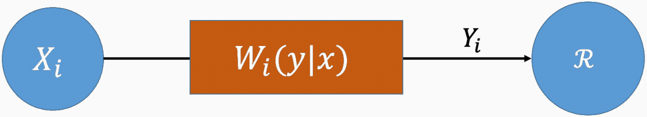


Local Information Constraints



- **Communication.** Only ℓ-bits from each user.

$$|\mathcal{Y}| \leq 2^{\ell}.$$



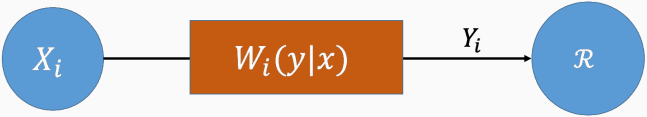
- **Communication.** Only ℓ-bits from each user.

$$|\mathcal{Y}| \leq 2^{\ell}.$$

- **Privacy.** W_i's satisfy ρ-local differentially privacy (LDP).

$$\sup_{y \in \mathcal{Y}} \sup_{x, x' \in \mathcal{X}} \frac{W_i(y|x)}{W_i(y|x')} \leq e^{\rho}.$$

Local Information Constraints



- **Communication.** Only ℓ-bits from each user.

$$|\mathcal{Y}| \leq 2^{\ell}.$$

- **Privacy.** W_i's satisfy ρ-local differentially privacy (LDP).

$$\sup_{y \in \mathcal{Y}} \sup_{x, x' \in \mathcal{X}} \frac{W_i(y|x)}{W_i(y|x')} \leq e^{\rho}.$$

- **Restricted Measurement.** E.g. Linear measurements, noisy measurements.

Related Works

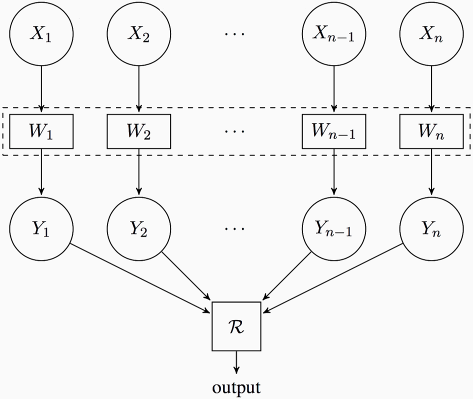
Identity Testing:
Paninski '08, Valiant-Valiant '17, ADK '15, Goldreich '16, DGPP '18

Communication-limited Inference:
ZDJW '13, GMN '14, AMS '18, FMO '18, HMW '18, BarnesH '19

LDP-constrainted Inference:
EPR '13, DJW '13, YB '17, ASZ '18, Sheffet '17, AJM '19, CSU '19

And many more.

How are Channels Selected

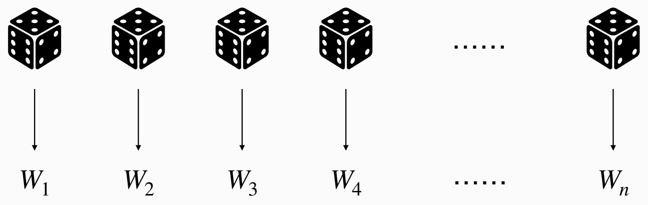


How are Channels Selected

Private-coin protocols:

$U_1, U_2, ..., U_n$: independent random seeds at each user

$W_i = g_i(U_i) \in \mathcal{W}$.

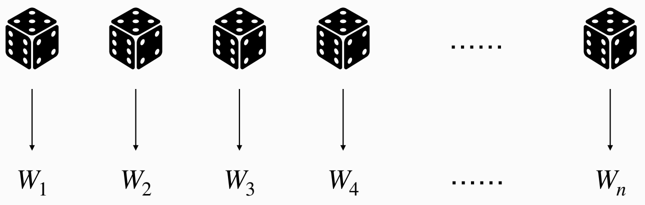


How are Channels Selected

Private-coin protocols:

$U_1, U_2, ..., U_n$: independent random seeds at each user

$W_i = g_i(U_i) \in \mathcal{W}$.



If $\mathcal{W}$ is convex,

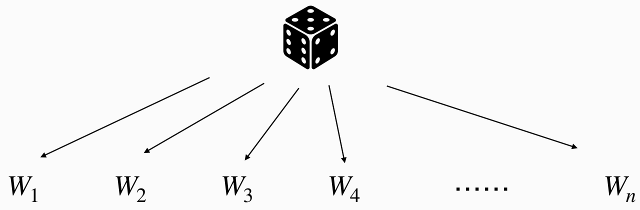
$$\bar{W}_i = \mathbb{E}_{U_i}[g_i(U_i)].$$

How are Channels Selected

Public-coin protocols:

U: shared random seeds available to all players and the referee.

$W_i = g_i(U) \in \mathcal{W}$.



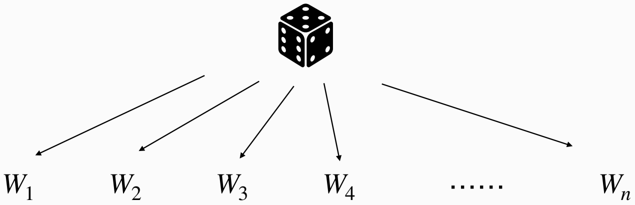
Previous Works

Acharya, Canonne, and Tyagi, 2019:

	Public-Coin Protocols	Private-Coin Protocols
No Constraint	$\Theta\left(\frac{\sqrt{k}}{\varepsilon^2}\right)$	
ℓ-bit	$\Theta\left(\frac{\sqrt{k}}{\varepsilon^2}\sqrt{\frac{k}{2^\ell}}\right)$	$\Theta\left(\frac{\sqrt{k}}{\varepsilon^2}\frac{k}{2^\ell}\right)$
ρ-LDP	$\Theta\left(\frac{\sqrt{k}}{\varepsilon^2}\frac{\sqrt{k}}{\rho^2}\right)$	$\Theta\left(\frac{\sqrt{k}}{\varepsilon^2}\frac{k}{\rho^2}\right)$

Limited Shared-Randomness

What if we can only throw the dice s times ($\Theta(s)$ bits of shared-randomness)?



Our Contribution

	Public-Coin Protocols	Private-Coin Protocols	s-bit shared randomness
No Constraint	$\Theta\left(\frac{\sqrt{k}}{\varepsilon^2}\right)$		
ℓ-bit	$\Theta\left(\frac{\sqrt{k}}{\varepsilon^2}\sqrt{\frac{k}{2^\ell}}\right)$	$\Theta\left(\frac{\sqrt{k}}{\varepsilon^2}\frac{k}{2^\ell}\right)$	$\Theta\left(\frac{\sqrt{k}}{\varepsilon^2}\frac{k}{\min\{2^{s/2},\sqrt{k}\}\,2^\ell}\right)$
ρ-LDP	$\Theta\left(\frac{\sqrt{k}}{\varepsilon^2}\frac{\sqrt{k}}{\rho^2}\right)$	$\Theta\left(\frac{\sqrt{k}}{\varepsilon^2}\frac{k}{\rho^2}\right)$	$\Theta\left(\frac{\sqrt{k}}{\varepsilon^2}\frac{k}{\min\{2^{s/2},\sqrt{k}\}\,\rho^2}\right)$

Our Contribution

	Public-Coin Protocols	**Private**-Coin Protocols	***s*-bit shared randomness**
No Constraint	$\Theta\left(\frac{\sqrt{k}}{\varepsilon^2}\right)$		
ℓ-bit	$\Theta\left(\frac{\sqrt{k}}{\varepsilon^2}\sqrt{\frac{k}{2^\ell}}\right)$	$\Theta\left(\frac{\sqrt{k}}{\varepsilon^2}\frac{k}{2^\ell}\right)$	$\Theta\left(\frac{\sqrt{k}}{\varepsilon^2}\frac{k}{\min\{2^{s/2},\sqrt{k}\}\,2^\ell}\right)$
ρ-LDP	$\Theta\left(\frac{\sqrt{k}}{\varepsilon^2}\frac{\sqrt{k}}{\rho^2}\right)$	$\Theta\left(\frac{\sqrt{k}}{\varepsilon^2}\frac{k}{\rho^2}\right)$	$\Theta\left(\frac{\sqrt{k}}{\varepsilon^2}\frac{k}{\min\{2^{s/2},\sqrt{k}\}\,\rho^2}\right)$

One bit of shared-randomness is worth 0.5 bit of communication!

Overview of Our Approach

Use shared randomness to *embed* the statistical problem *into a smaller domain*

Overview of Our Approach

Use shared randomness to *embed* the statistical problem
into a smaller domain

High-level description

1. Domain compression: find a set $\mathcal{F}$ of mappings $f : [k] \to [L]$ of size 2^s such that for all distributions p, q supported on $[k]$,

$$\Pr_{f \sim \mathrm{Unif}(\mathcal{F})}\left(d(p^f, q^f) \geq \theta \cdot d(p, q)\right) \geq 1 - \delta$$

holds for small L, large θ, small δ, and suitable d;

Overview of Our Approach

Use shared randomness to *embed* the statistical problem *into a smaller domain*

High-level description

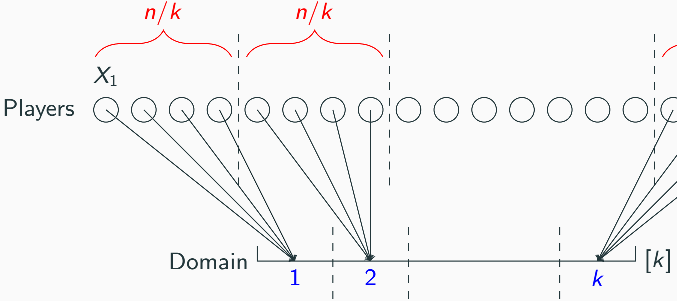
1. Domain compression: find a set $\mathcal{F}$ of mappings $f : [k] \to [L]$ of size 2^s such that for all distributions p, q supported on $[k]$,

$$\Pr_{f \sim \mathrm{Unif}(\mathcal{F})}\left(d(p^f, q^f) \geq \theta \cdot d(p, q)\right) \geq 1 - \delta$$

   holds for small L, large θ, small δ, and suitable d;
2. Reduction to small domain: players use the s-bit shared randomness to apply the same mapping $f \in \mathcal{F}$ to their data, and use the private-randomness scheme for the small domain.

Estimation with No Shared Randomness

Suppose $s = 0$ and $\ell = 1$:

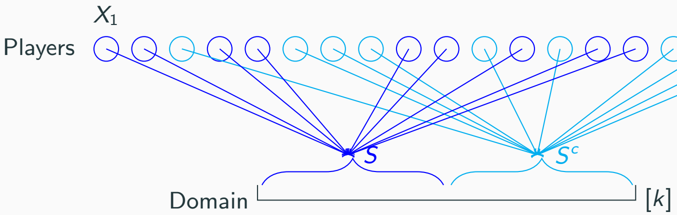


Reduced to uniformity testing with $n' = n/k$, therefore

$$n' = O\left(\frac{\sqrt{k}}{\varepsilon^2}\right) \implies n = O\left(\frac{\sqrt{k}}{\varepsilon^2} \cdot k\right).$$

Estimation with Unlimited Shared Randomness

Suppose $s = \infty$ and $\ell = 1$:



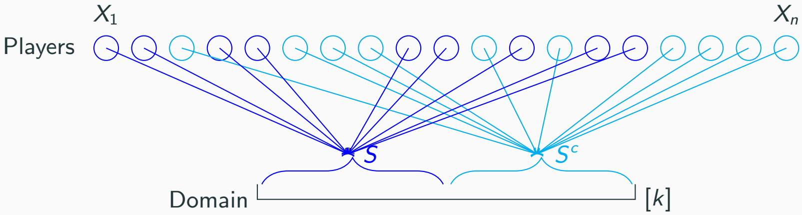
Theorem (ACT'19)

Let $S \subseteq [k]$ be a uniformly random subset of size $k/2$, and p^S be the restriction of p on (S, S^c). For any p, q supported on $[k]$,

$$\Pr_S \left(\|p^S - q^S\|_{\mathrm{TV}} \geq \frac{0.1}{\sqrt{k}} \|p - q\|_{\mathrm{TV}} \right) \geq 0.01.$$

Estimation with Unlimited Shared Randomness

Suppose $s = \infty$ and $\ell = 1$:



Reduced to uniformity testing with $(k', \varepsilon') = (2, \frac{\varepsilon}{10\sqrt{k}})$, giving

$$n = O\left(\frac{\sqrt{k'}}{(\varepsilon')^2}\right) = O\left(\frac{\sqrt{k}}{\varepsilon^2} \cdot \sqrt{k}\right).$$

Derandomization

Selecting a random subset is not randomness efficient.

$\Theta(k)$ bits of shared-randomness.

Derandomization

Theorem (ACHST'20)

There exist $m = O(k)$ and subsets $S_1, \ldots, S_m \subseteq [k]$ of size $k/2$ such that for any p, q supported on $[k]$,

$$\Pr_{S \sim \mathrm{Unif}\{S_1,\ldots,S_m\}} \left(\|p^S - q^S\|_{\mathrm{TV}} \geq \frac{0.1}{\sqrt{k}} \|p - q\|_{\mathrm{TV}} \right) \geq 0.01.$$

Derandomization

Theorem (ACHST'20)

There exist $m = O(k)$ and subsets $S_1, \ldots, S_m \subseteq [k]$ of size $k/2$ such that for any p, q supported on $[k]$,

$$\Pr_{S \sim \mathsf{Unif}\{S_1,\ldots,S_m\}} \left(\|p^S - q^S\|_{\mathrm{TV}} \geq \frac{0.1}{\sqrt{k}} \|p - q\|_{\mathrm{TV}} \right) \geq 0.01.$$

Remarks

- subsets $S_1, \ldots, S_m$ chosen before p and q;
- $m = \Omega(k)$ also necessary;

Derandomization

Theorem (ACHST'20)

There exist $m = O(k)$ and subsets $S_1, \ldots, S_m \subseteq [k]$ of size $k/2$ such that for any p, q supported on $[k]$,

$$\Pr_{S \sim \mathrm{Unif}\{S_1,\ldots,S_m\}} \left(\|p^S - q^S\|_{\mathrm{TV}} \geq \frac{0.1}{\sqrt{k}} \|p - q\|_{\mathrm{TV}} \right) \geq 0.01.$$

Remarks

- subsets $S_1, \ldots, S_m$ chosen before p and q;
- $m = \Omega(k)$ also necessary;

$\Theta(\log k)$ bits suffice to achieve public-coin performance.

Domain Compression: A Key Primitive

Domain Compression Theorem (ACHST'20)

There exists constants c, δ_0, $\forall \theta \in [\sqrt{c/k}, \sqrt{c/2}]$ and $L \geq \theta^2 k/c$, there exists a set $\mathcal{F}$ of mappings $f : [k] \to [L]$ of size $O(\frac{1}{\theta^2})$ such that for all distributions p, q supported on $[k]$,

$$\Pr_{f \sim \mathsf{Unif}(\mathcal{F})}(\|p^f - q^f\|_{\mathrm{TV}} \geq \theta \cdot \|p - q\|_{\mathrm{TV}}) \geq 1 - \delta_0.$$

Domain Compression: A Key Primitive

Domain Compression Theorem (ACHST'20)

There exists constants c, δ_0, $\forall \theta \in [\sqrt{c/k}, \sqrt{c/2}]$ and $L \geq \theta^2 k/c$, there exists a set $\mathcal{F}$ of mappings $f : [k] \to [L]$ of size $O(\frac{1}{\theta^2})$ such that for all distributions p, q supported on $[k]$,

$$\Pr_{f \sim \mathrm{Unif}(\mathcal{F})}(\|p^f - q^f\|_{\mathrm{TV}} \geq \theta \cdot \|p - q\|_{\mathrm{TV}}) \geq 1 - \delta_0.$$

Parameter choices:

- Size $|\mathcal{F}| = O(\frac{1}{\theta^2})$. Select $\theta = O(\frac{1}{\sqrt{2^s}})$.
- New domain size $L = O(\theta^2 k) = O(k/2^s)$.
- $\varepsilon' > \frac{\varepsilon}{\sqrt{2^s}}$.

Domain Compression: A Key Primitive

Domain Compression Theorem (ACHST'20)

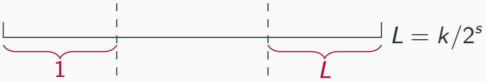
There exists constants c, δ_0, $\forall \theta \in [\sqrt{c/k}, \sqrt{c/2}]$ and $L \geq \theta^2 k/c$, there exists a set $\mathcal{F}$ of mappings $f : [k] \to [L]$ of size $O(\frac{1}{\theta^2})$ such that for all distributions p, q supported on $[k]$,

$$\Pr_{f \sim \mathsf{Unif}(\mathcal{F})}(\|p^f - q^f\|_{\mathrm{TV}} \geq \theta \cdot \|p - q\|_{\mathrm{TV}}) \geq 1 - \delta_0.$$

Remarks

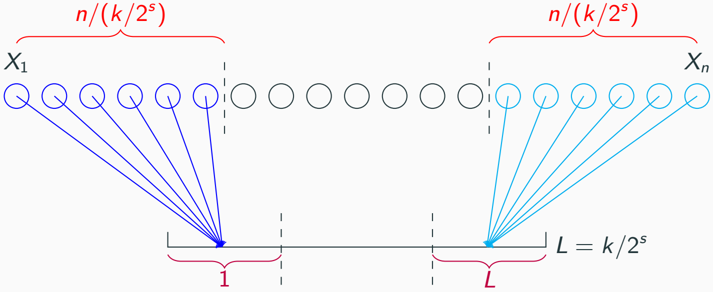
- Each mapping is an almost equal partition of the domain.
- Similar results hold for ℓ_2 in addition to TV.

Estimation with Some Shared Randomness



Reduced to uniformity testing with $(k', \varepsilon') = (\frac{k}{2^s}, \frac{\varepsilon}{\sqrt{2^s}})$.

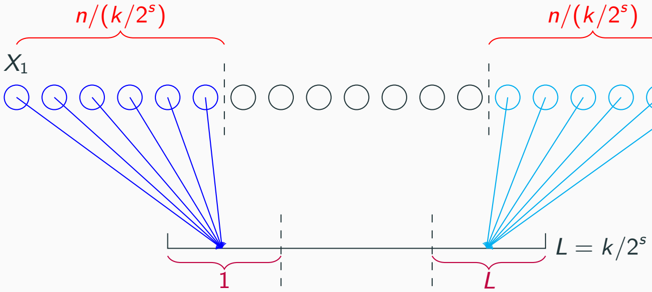
Estimation with Some Shared Randomness



Recall one bit protocol. $n' = \frac{n}{k/2^s}$, therefore

$$n' = O\left(\frac{\sqrt{k'}}{(\varepsilon')^2}\right) \implies n = O\left(\frac{\sqrt{k}}{\varepsilon^2} \cdot \sqrt{k} \cdot \sqrt{\frac{k}{2^s} \vee 1}\right).$$

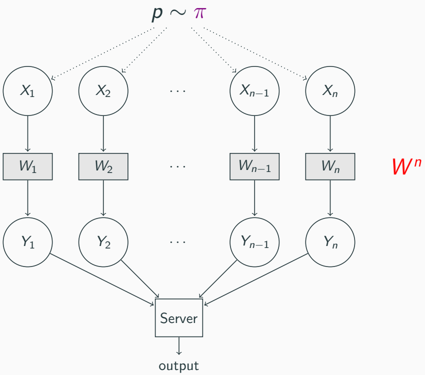
Estimation with Some Shared Randomness



A small catch:

- boosting using repetition requires more shared randomness. ☹
- solution: deterministic amplification. ☺
- see full paper for details.

Lower Bound Idea



Learner: choose communication channel $W^n = (W_1, \cdots, W_n)$ to **perform** constrained inference.

Adversary: choose prior π on the underlying distribution p to **confuse** the learner.

Role of shared randomness:

- without shared randomness: W^n is a product channel;
- with shared randomness: W^n is a mixture of product channels.

Semimaxmin Information

- Let $I(W^n \to \pi)$ be a suitable notion of "information" provided by a given channel W^n to a given prior π.

Semimaxmin Information

- Let $I(W^n \to \pi)$ be a suitable notion of "information" provided by a given channel W^n to a given prior π.
- Semimaxmin information:

$$\underline{\bar{I}} = \qquad I(W^n \to \pi)\,.$$

Semimaxmin Information

- Let $I(W^n \to \pi)$ be a suitable notion of "information" provided by a given channel W^n to a given prior π.
- Semimaxmin information:

$$\underline{\bar{I}} = \max_{\mathcal{W}:|\mathcal{W}|=2^s} \qquad I(W^n \to \pi) \,.$$

Semimaxmin Information

- Let $I(W^n \to \pi)$ be a suitable notion of "information" provided by a given channel W^n to a given prior π.
- Semimaxmin information:

$$\underline{\bar{I}} = \max_{\mathcal{W}:|\mathcal{W}|=2^s} \min_{\pi} \qquad I(W^n \to \pi) .$$

Semimaxmin Information

- Let $I(W^n \to \pi)$ be a suitable notion of "information" provided by a given channel W^n to a given prior π.
- Semimaxmin information:

$$\underline{\bar{I}} = \max_{\mathcal{W}:|\mathcal{W}|=2^s} \min_{\pi} \mathbb{E}_{W^n \sim \mathsf{Unif}(\mathcal{W})}[I(W^n \to \pi)].$$

Semimaxmin Information

- Let $I(W^n \to \pi)$ be a suitable notion of "information" provided by a given channel W^n to a given prior π.
- Semimaxmin information:

$$\overline{\underline{I}} = \max_{\mathcal{W}:|\mathcal{W}|=2^s} \min_{\pi} \mathbb{E}_{W^n \sim \mathrm{Unif}(\mathcal{W})}[I(W^n \to \pi)].$$

- $s = 0$ gives the maxmin information for private randomness:

$$\overline{\underline{I}} \geqslant \underline{I} = \max_{W^n} \min_{\pi} I(W^n \to \pi).$$

Semimaxmin Information

- Let $I(W^n \to \pi)$ be a suitable notion of "information" provided by a given channel W^n to a given prior π.
- Semimaxmin information:

$$\bar{\underline{I}} = \max_{\mathcal{W}:|\mathcal{W}|=2^s} \min_{\pi} \mathbb{E}_{W^n \sim \mathrm{Unif}(\mathcal{W})}[I(W^n \to \pi)].$$

- $s = 0$ gives the maxmin information for private randomness:

$$\bar{\underline{I}} \geqslant \underline{I} = \max_{W^n} \min_{\pi} I(W^n \to \pi).$$

- $s = \infty$ gives the minmax information for public randomness:

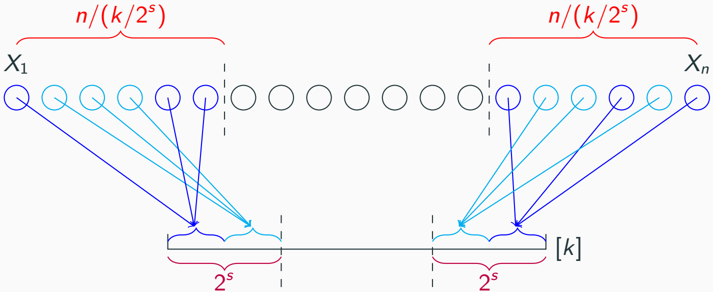
$$\bar{\underline{I}} \leqslant \bar{I} = \min_{\pi} \max_{W^n} I(W^n \to \pi).$$

Conclusion

- randomness-optimal domain compression;
- tight tradeoffs on shared randomness.

Thank You!
arXiv: 1907.08743

Estimation with Some Shared Randomness



Reduced to uniformity testing with $(k', \varepsilon') = (\frac{2k}{2^s}, \frac{\varepsilon}{10\sqrt{2^s}})$ and $n' = \frac{n}{k/2^s}$, therefore

$$n' = O\left(\frac{\sqrt{k'}}{(\varepsilon')^2}\right) \implies n = O\left(\frac{\sqrt{k}}{\varepsilon^2} \cdot \sqrt{k} \cdot \sqrt{\frac{k}{2^s} \vee 1}\right).$$

Estimation with Some Shared Randomness