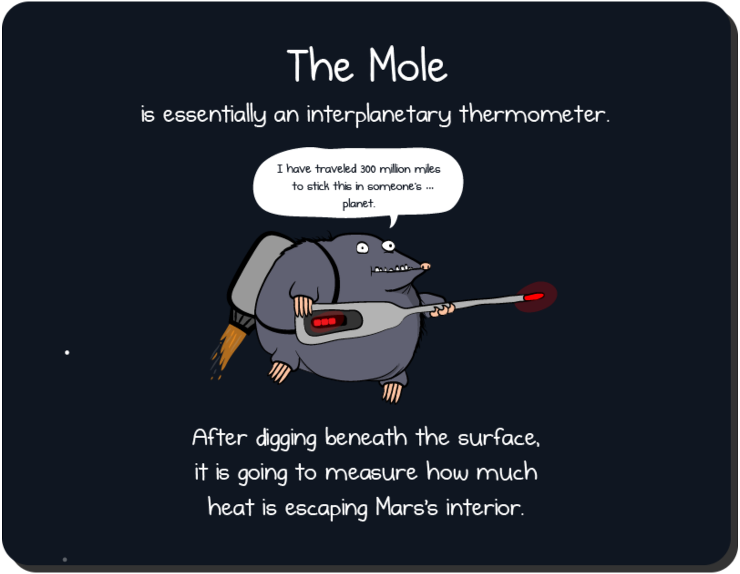
Testing and Learning Distributions Under Local Information Constraints

Clément Canonne (Stanford University)

Based on joint works with **Jayadev Acharya** (Cornell University), **Cody Freitag** (Cornell University), and **Himanshu Tyagi** (IISc Bangalore)

Simons Workshop – November 27, 2018

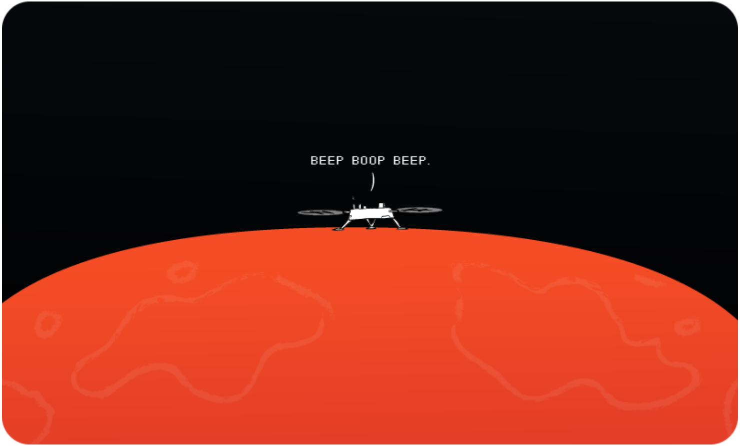

Right now,
as you're reading this,
a spacecraft is traveling toward Mars..


The Mole
is essentially an interplanetary thermometer.
I have traveled 300 million miles to stick this in someone's ... planet.
After digging beneath the surface, it is going to measure how much heat is escaping Mars's interior.

Why...

... am I telling you this?

(for a start, it's pretty great.)


BEEP BOOP BEEP.

In space, no one can hear you stream

In space, no one can hear you stream

- Harsh communication constraints
  - various types of noise
  - energy and battery bottlenecks
  - limited window of communication
  - transmitter size

In space, no one can hear you stream

- Harsh communication constraints
  - various types of noise
  - energy and battery bottlenecks
  - limited window of communication
  - transmitter size
- Cost of deployment

Protocols for the task

Minimize cost, risk of failure, etc. accounting for constraints. How many sensors? How many different spacecrafts? How to send the information?

This work

General question

How to make machine learning "work" with limited resources?

This work

General question

How to make machine learning "work" with limited resources?

This work

Statistical inference under information constraints

Setting: “Simultaneous Communication Protocol” (SMP)

Setting: "Simultaneous Communication Protocol" (SMP)

- an inference task $\mathcal{P}$ over k-ary distributions

Setting: "Simultaneous Communication Protocol" (SMP)

- an inference task $\mathcal{P}$ over k-ary distributions
- an unknown k-ary distribution p

Setting: "Simultaneous Communication Protocol" (SMP)

- an inference task $\mathcal{P}$ over k-ary distributions
- an unknown k-ary distribution p
- one centralized "referee" $\mathcal{R}$ who needs to solve $\mathcal{P}$ on p

Setting: "Simultaneous Communication Protocol" (SMP)

- an inference task $\mathcal{P}$ over k-ary distributions
- an unknown k-ary distribution p
- one centralized "referee" $\mathcal{R}$ who needs to solve $\mathcal{P}$ on p
- n locally-constrained players, each with a channel $W \in \mathcal{W}$

Setting: "Simultaneous Communication Protocol" (SMP)

- an inference task $\mathcal{P}$ over k-ary distributions
- an unknown k-ary distribution p
- one centralized "referee" $\mathcal{R}$ who needs to solve $\mathcal{P}$ on p
- n locally-constrained players, each with a channel $W \in \mathcal{W}$
- each player independently gets one sample x from p and sends a message $y = W(x)$ to $\mathcal{R}$

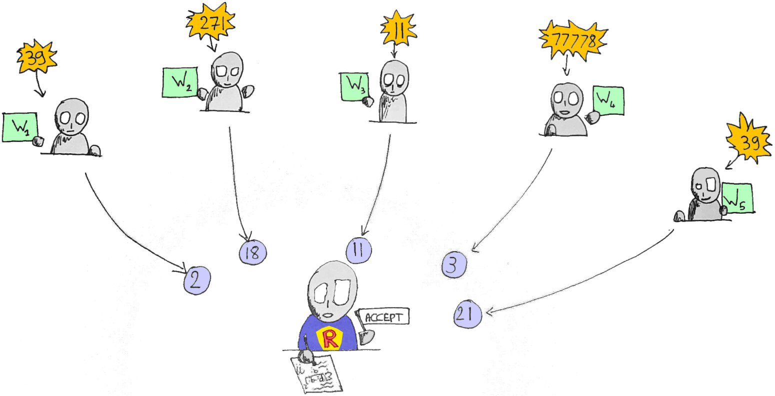
Setting: "Simultaneous Communication Protocol" (SMP)

- an inference task $\mathcal{P}$ over k-ary distributions
- an unknown k-ary distribution p
- one centralized "referee" $\mathcal{R}$ who needs to solve $\mathcal{P}$ on p
- n locally-constrained players, each with a channel $W \in \mathcal{W}$
- each player independently gets one sample x from p and sends a message $y = W(x)$ to $\mathcal{R}$

Question

As a function of k, $\mathcal{W}$, and all relevant parameters of $\mathcal{P}$, what is the number of players n required?

Setting, cont'd



Setting, cont'd

Setting, cont'd

- $\mathcal{W}$ encodes the local constraints: if $\mathrm{Id} \in \mathcal{W}$, trivial

Setting, cont'd

- $\mathcal{W}$ encodes the local constraints: if $\mathrm{Id} \in \mathcal{W}$, trivial
- Inference tasks: density estimation, parameter estimation, functional estimation, hypothesis testing/property testing...

Setting, cont'd

- $\mathcal{W}$ encodes the local constraints: if $\mathrm{Id} \in \mathcal{W}$, trivial
- Inference tasks: density estimation, parameter estimation, functional estimation, hypothesis testing/property testing...
- Different available resources s.t. randomness: public- or private-coin

Enough with the fancy "P"... what are we talking about anyway?

Focused on two specific fundamental* inference tasks:

Distribution Learning

Must output: $\hat{p}$ such that $\ell_1(p, \hat{p}) \leq \varepsilon$

(and be correct on any p with probability at least 2/3)

Enough with the fancy "P"... what are we talking about anyway?

Focused on two specific fundamental* inference tasks:

Distribution Learning

Must output: $\hat{p}$ such that $\ell_1(p, \hat{p}) \leq \varepsilon$

(and be correct on any p with probability at least $2/3$)

Uniformity Testing

Must decide: $p = u_k$ (uniform), or $\ell_1(p, u_k) \geq \varepsilon$?

(and be correct on any p with probability at least $2/3$)

* "If we can make it here, we can make it anywhere." [DK16, Gol16]

Distribution learning and uniformity testing

What is known without local constraints:

Task $\mathcal{P}$	n
Distribution learning	$\frac{k}{\varepsilon^2}$
Uniformity testing	$\frac{\sqrt{k}}{\varepsilon^2}$

Distribution learning and uniformity testing

What is known without local constraints:

Task $\mathcal{P}$	n
Distribution learning	$\frac{k}{\varepsilon^2}$
Uniformity testing	$\frac{\sqrt{k}}{\varepsilon^2}$

What happens with them?

Distribution learning and uniformity testing

What is known without local constraints:

Task $\mathcal{P}$	n
Distribution learning	$\frac{k}{\varepsilon^2}$
Uniformity testing	$\frac{\sqrt{k}}{\varepsilon^2}$

What happens with them? And does public randomness help then?

Related work

- Learning under communication constraints: [HÖW18] (same model, allows (some) adaptivity), [DGL$^+$17] (different model and focus)
- Testing under communication constraints: [FMO18] (related model, different focus), [AMS18] (different (two-party) model and focus)
- Locally private learning: [DJW13, YB17, ASZ18]
- Locally private testing: [She18]
- Decentralized detection: [Tsi93] (same model, similar-ish focus)

(+ **many** in adjacent areas/models)

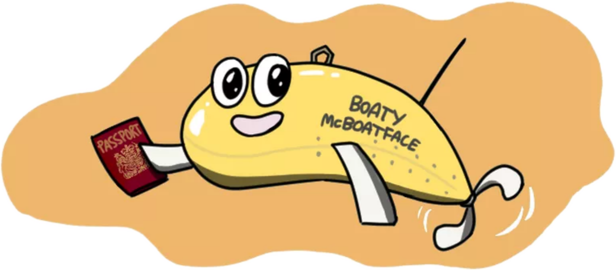
Plan for the talk

1. Communication-Starved Setting
2. Local Differential Privacy
3. General Lower Bound Framework

Part I: Communication–Starved Setting

Boaty McBoatface is starting its first mission today!
It's going to Antarctica to study global warming, not to play.

The world's oceans are changing, you see.
It's freezing down there, but not as cold as it used to be.



Boaty's findings will be sent to scientists with care,
By way of a radio link, but with a certain flair.

McBoatfaces are expensive

What is the most ship-efficient protocol to reliably test whether the distribution of temperatures matches the one on record?

Setting: what is $\mathcal{W}$?

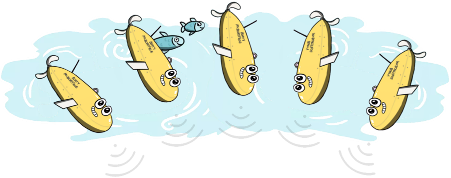
n communication-limited players, each can send ℓ bits to $\mathcal{R}$:

$$\mathcal{W} = \{W\colon [k] \to \{0,1\}^\ell\}$$

Setting: what is $\mathcal{W}$?

n communication-limited players, each can send ℓ bits to $\mathcal{R}$:

$$\mathcal{W} = \{W \colon [k] \to \{0,1\}^{\ell}\}$$



One Approach To Solve It All

One Approach To Solve It All

Key Observation

If the referee can simulate independent samples from p using the messages from the players, then it can do anything as in the centralized setting.

One Approach To Solve It All

Key Observation

If the referee can simulate independent samples from p using the messages from the players, then it can do anything as in the centralized setting.

Begging the question

Can the referee simulate independent samples from p using the messages from the players?

One Approach To Solve It All?

Theorem (No.)

For every $k \geq 1$ *and* $\ell < \log k$*, there exists no SMP with* ℓ *bits of communication per player for distributed simulation over* $[k]$ *with any finite number of players. (Even allowing public-coin and interactive protocols.)*

One Approach To Solve It All?

Theorem (No.)

For every $k \geq 1$ and $\ell < \log k$, there exists no SMP with ℓ bits of communication per player for distributed simulation over $[k]$ with any finite number of players. (Even allowing public-coin and interactive protocols.)

Proof.

By contradiction, [...] pigeonhole principle [...]. □

One Approach To Solve It All?

Theorem (Yes!)

For every $k, \ell \geq 1$, *there exists a private-coin protocol with* ℓ *bits of communication per player for distributed simulation over* $[k]$, *with expected number of players* $O(k/2^\ell \vee 1)$.

One Approach To Solve It All?

Theorem (Yes!)

For every $k, \ell \geq 1$, there exists a private-coin protocol with ℓ bits of communication per player for distributed simulation over $[k]$, with expected number of players $O(k/2^{\ell} \vee 1)$.

Proof.

Case $\ell = 1$.

One Approach To Solve It All?

Theorem (Yes!)

For every $k, \ell \geq 1$, there exists a private-coin protocol with ℓ bits of communication per player for distributed simulation over $[k]$, with expected number of players $O(k/2^\ell \vee 1)$.

Proof.

Case $\ell = 1$. Player $2i - 1$ and $2i$ both send 1 if their sample "hits" i;

One Approach To Solve It All?

Theorem (Yes!)

For every $k, \ell \geq 1$, there exists a private-coin protocol with ℓ bits of communication per player for distributed simulation over $[k]$, with expected number of players $O(k/2^\ell \vee 1)$.

Proof.

Case $\ell = 1$. Player $2i-1$ and $2i$ both send 1 if their sample "hits" i; the referee outputs i if (i) player $2i-1$ is the only odd player sending 1, and player $2i$ sends 0.

One Approach To Solve It All?

Theorem (Yes!)

For every $k, \ell \geq 1$, there exists a private-coin protocol with ℓ bits of communication per player for distributed simulation over $[k]$, with expected number of players $O(k/2^\ell \vee 1)$.

Proof.

Case $\ell = 1$. Player $2i-1$ and $2i$ both send 1 if their sample "hits" i; the referee outputs i if (i) player $2i-1$ is the only odd player sending 1, and player $2i$ sends 0. Then, conditioned on $\mathcal{R}$ not outputting $\perp$, i is outputted with probability p_i.

One Approach To Solve It All?

Theorem (Yes!)

For every $k, \ell \geq 1$, there exists a private-coin protocol with ℓ bits of communication per player for distributed simulation over $[k]$, with expected number of players $O(k/2^\ell \vee 1)$.

Proof.

Case $\ell = 1$. Player $2i-1$ and $2i$ both send 1 if their sample "hits" i; the referee outputs i if (i) player $2i-1$ is the only odd player sending 1, and player $2i$ sends 0. Then, conditioned on $\mathcal{R}$ not outputting $\perp$, i is outputted with probability p_i. And the probability to output $\perp$ is

$$1 - \prod_{i=1}^{k}(1-p_i) \leq 1 - \text{blah}(\|p\|_2)$$

One Approach To Solve It All?

Theorem (Yes!)

For every $k, \ell \geq 1$, there exists a private-coin protocol with ℓ bits of communication per player for distributed simulation over $[k]$, with expected number of players $O(k/2^\ell \vee 1)$.

Proof.

Case $\ell = 1$. Player $2i-1$ and $2i$ both send 1 if their sample "hits" i; the referee outputs i if (i) player $2i-1$ is the only odd player sending 1, and player $2i$ sends 0. Then, conditioned on $\mathcal{R}$ not outputting $\perp$, i is outputted with probability p_i. And the probability to output $\perp$ is

$$1 - \prod_{i=1}^{k}(1-p_i) \leq 1 - \text{blah}(\|p\|_2)$$

(and some complications to bound this away from 1). □

One Approach To Solve It All!

Corollary (Informal)

For any inference task $\mathcal{P}$ over k-ary distributions with sample complexity s in the centralized model, there is a private-coin protocol for $\mathcal{P}$, with ℓ bits of communication per player, and $n = O(s \cdot k/2^{\ell})$ players.

Illustration ©Dami Lee

One Approach To Solve It All!

Corollary (Distribution Learning)

For every $k, \ell \leq \log_2 k$, there is a private-coin protocol for learning k-ary distributions with ℓ bits per player, and $n = O(\frac{k^2}{2^\ell \varepsilon^2})$ players.

Corollary (Uniformity Testing)

For every $k, \ell \leq \log_2 k$, there is a private-coin protocol for testing uniformity over $[k]$ with ℓ bits per player, and $n = O(\frac{k^{3/2}}{2^\ell \varepsilon^2})$ players.

One Approach To Really, Really Solve It All?

One Approach To Really, Really Solve It All?

Natural Question

Is this "simulate-and-infer" approach optimal?

One Approach To Really, Really Solve It All?

Natural Question

Is this "simulate-and-infer" approach optimal?

Answer

Not if one allows public coins!

Distributed Uniformity Testing with Public Coins

Theorem (Upper Bound)

For every $k, \ell \leq \log_2 k$, there is a public-coin protocol for testing uniformity over $[k]$ with ℓ bits per player, and $n = O\left(\frac{k}{2^{\ell/2}\varepsilon^2}\right)$ players.

MCH (Minimally Contracting Hashing)

Theorem (χ^2 contraction)

Choose u.a.r. a balanced partition Π of $[k]$ in L parts, and let p_Π be the distribution induced by p on Π. Then

$$\Pr_{\Pi}[\ell_1(p_\Pi, u_L) \geq \Omega(\sqrt{L/k})\ell_1(p, u_k)] \geq \Omega(1)\,.$$

MCH (Minimally Contracting Hashing)

Theorem (χ^2 contraction)

Choose u.a.r. a balanced partition Π of $[k]$ in L parts, and let p_Π be the distribution induced by p on Π. Then

$$\Pr_\Pi[\ell_1(p_\Pi, u_L) \geq \Omega(\sqrt{L/k})\ell_1(p, u_k)] \geq \Omega(1)\,.$$

Proof.

Not hard (but technical). Dealing with dependencies when computing second and fourth moments + Paley–Zygmund. □

MCH (Minimally Contracting Hashing)

Theorem (χ^2 contraction)

Choose u.a.r. a balanced partition Π of $[k]$ in L parts, and let p_Π be the distribution induced by p on Π. Then

$$\Pr_{\Pi}[\ell_1(p_\Pi, u_L) \geq \Omega(\sqrt{L/k})\ell_1(p, u_k)] \geq \Omega(1)\,.$$

Proof.

Not hard (but technical). Dealing with dependencies when computing second and fourth moments + Paley–Zygmund. □

(This is tight).

MCH (Minimally Contracting Hashing)

Apply with $L := 2^{\ell}$, choosing a common random Π using public coins.
Test p_{Π} with $\varepsilon' := \sqrt{L/k}\varepsilon$:

$$\frac{\sqrt{L}}{\varepsilon'^2} = \frac{k}{2^{\ell/2}\varepsilon^2} .$$

MCH (Minimally Contracting Hashing)

Apply with $L := 2^{\ell}$, choosing a common random Π using public coins.
Test p_Π with $\varepsilon' := \sqrt{L/k}\varepsilon$:

$$\frac{\sqrt{L}}{\varepsilon'^2} = \frac{k}{2^{\ell/2}\varepsilon^2}\,.$$

Repeat in parallel to amplify probability. $\square$

MCH (Minimally Contracting Hashing)

Interpretation

Use public randomness to randomly map the domain to a smaller one, which provides the best tradeoff domain reduction/distance shrinkage to test, w.r.t. χ^2 distance, given the communication constraints.

MCH (Minimally Contracting Hashing)

- Simple.
- χ^2 contraction theorem: very general.
- Randomness: $O(k\ell)$ bits (Improve using 4-wise independence)
- We'll see it again: with Batman.

Distribution learning and uniformity testing

With local communication constraints (upper bounds):

Task $\mathcal{P}$	n (private-coin)	n (public-coin)
Distribution learning	$\frac{k}{\varepsilon^2} \cdot \frac{k}{2^\ell}$	$\frac{k}{\varepsilon^2} \cdot \frac{k}{2^\ell}$
Uniformity testing	$\frac{\sqrt{k}}{\varepsilon^2} \cdot \frac{k}{2^\ell}$	$\frac{\sqrt{k}}{\varepsilon^2} \cdot \sqrt{\frac{k}{2^\ell}}$

Part II: Local Differential Privacy

Local Differential Privacy (LDP)

“No one can know.”

Setting: what is W?

n privacy-conscious players, each can send a ϱ-private message to the $\mathcal{R}$:

$$\mathcal{W} = \{W\colon [k] \to \{0,1\}^* : W \ \varrho\text{-LDP}\}$$

i.e., for all $x, x' \in [k]$, $y \in \{0,1\}^*$,

$$\frac{W(y \mid x)}{W(y \mid x')} \leq e^{\varrho}$$

Private-coin upper bounds

Two protocols: Rappor-based, Hadamard-Response-based.

Uniformity testing

Private-coin upper bounds

Two protocols: RAPPOR-based, HADAMARD-RESPONSE-based.

Public-coin upper bound

RAPTOR: uses the χ^2-contraction theorem, for $\ell = 1$.

MCH (Minimally Contracting Hashing)

Interpretation

Use public randomness to randomly map the domain to a smaller one, which provides the best tradeoff domain reduction/distance shrinkage to test, w.r.t. χ^2 distance, given the privacy constraints.

Distribution learning and uniformity testing

With local privacy constraints (upper bounds):

Task $\mathcal{P}$	n (private-coin)	n (public-coin)
Distribution learning	$\frac{k}{\varepsilon^2} \cdot \frac{k}{\varrho^2}$	$\frac{k}{\varepsilon^2} \cdot \frac{k}{\varrho^2}$
Uniformity testing	$\frac{\sqrt{k}}{\varepsilon^2} \cdot \frac{k}{\varrho^2}$	$\frac{\sqrt{k}}{\varepsilon^2} \cdot \frac{\sqrt{k}}{\varrho^2}$

Part III: Lower Bounds via χ^2 contraction

The Lower Bounds

Theorem (Upper Bounds are Lower Bounds)

Every upper bound mentioned in this talk is optimal.

The Lower Bounds

Theorem (Upper Bounds are Lower Bounds)

Every upper bound mentioned in this talk is optimal.

Corollary

Sharing (randomness) helps a lot for testing, not at all for learning.

The Lower Bound (I)

By Le Cam's two-point method, consider a distribution $\mathcal{Z}$ over "hard instances":

$$\forall 1 \leq i \leq k/2, \qquad p(2i-1), p(2i) = \left(\frac{1 \pm \varepsilon}{k}, \frac{1 \mp \varepsilon}{k} \right)$$

uniformly and independently at random. (Paninski's construction [Pan08]).

The Lower Bound (I)

By Le Cam's two-point method, consider a distribution $\mathcal{Z}$ over "hard instances":

$$\forall 1 \leq i \leq k/2, \qquad p(2i-1), p(2i) = \left(\frac{1 \pm \varepsilon}{k}, \frac{1 \mp \varepsilon}{k}\right)$$

uniformly and independently at random. (Paninski's construction [Pan08]).

...then look at them through the channels: $W^n \circ p^n$.

But...

... needs to upper bound the TV distance between (i) distribution of n messages sent to the referee when $p = u_k$, and (ii) distribution of n messages under average hard instance.

The Lower Bound (I)

By Le Cam's two-point method, consider a distribution $\mathcal{Z}$ over "hard instances":

$$\forall 1 \leq i \leq k/2, \qquad p(2i-1), p(2i) = \left(\frac{1 \pm \varepsilon}{k}, \frac{1 \mp \varepsilon}{k}\right)$$

uniformly and independently at random. (Paninski's construction [Pan08]).

...then look at them through the channels: $W^n \circ p^n$.

But...

... needs to upper bound the TV distance between (i) distribution of n messages sent to the referee when $p = u_k$, and (ii) distribution of n messages under average hard instance. The latter is not a product distribution...

The Lower Bound (I)

Want to bound TV distance between transcripts – “right” proxy is χ^2:

The Lower Bound (I)

Want to bound TV distance between transcripts – "right" proxy is χ^2:

$$\ell_1(\mathbb{E}_{Z\sim\mathcal{Z}}[Y_n^Z], Y_n^u)^2 \leq \chi^2(\mathbb{E}_{Z\sim\mathcal{Z}}[Y_n^Z], Y_n^u) \overset{\text{(goal)}}{\ll} 1$$

where Y_n^u is the distribution of the n messages under the uniform distribution, and Y_n^Z the distribution of the n messages under p_Z.

The Lower Bound (I)

To a channel W, we associate a p.s.d. matrix $H(W) \in \mathbb{R}^{k/2 \times k/2}$:

$$H(W)_{i_1,i_2} := \sum_y \frac{(W(y \mid 2i_1 - 1) - W(y \mid 2i_1))(W(y \mid 2i_2 - 1) - W(y \mid 2i_2))}{\sum_{i \in [k]} W(y \mid i)} .$$

The Lower Bound (I)

To a channel W, we associate a p.s.d. matrix $H(W) \in \mathbb{R}^{k/2 \times k/2}$:

$$H(W)_{i_1,i_2} := \sum_{y} \frac{(W(y \mid 2i_1 - 1) - W(y \mid 2i_1))(W(y \mid 2i_2 - 1) - W(y \mid 2i_2))}{\sum_{i \in [k]} W(y \mid i)} .$$

We characterize the contraction in chi-square distances in terms of the Frobenius and trace norms of this matrix: $\|H(W)\|_F$ and $\|H(W)\|_*$.

The Lower Bound (I)

Reminiscent

... of the SQ learning bounds via χ^2 [FGR$^+$13, SVW16, Fel17], esp. in view of the relation of SQ learning to local privacy [KLN$^+$11]. However, different quantities at play here (trace/Frobenius vs. spectral norms), leading to tighter bounds.

The Lower Bound (I)

Reminiscent

… of the SQ learning bounds via χ^2 [FGR$^+$13, SVW16, Fel17], esp. in view of the relation of SQ learning to local privacy [KLN$^+$11]. However, different quantities at play here (trace/Frobenius vs. spectral norms), leading to tighter bounds.

Works

… for public-coin protocols. But for private-coin (higher) lower bound, we need a more specifically designed perturbation $\mathcal{Z}$ to get optimal bound.

The Lower Bound (II)

Generalization: design a perturbation distribution $\mathcal{Z}$ over $[-1,1]^{k/2}$:

$$\forall 1 \leq i \leq k/2, \qquad p(2i-1), p(2i) = \left(\frac{1+\varepsilon Z}{k}, \frac{1-\varepsilon Z}{k}\right)$$

such that $Z \sim \mathcal{Z}$ has $\|Z\|_1 \geq 1/100$ w.h.p. (Generalizes Paninski's construction).

The Lower Bound (II)

Generalization: design a perturbation distribution $\mathcal{Z}$ over $[-1,1]^{k/2}$:

$$\forall 1 \leq i \leq k/2, \qquad p(2i-1), p(2i) = \left(\frac{1+\varepsilon Z}{k}, \frac{1-\varepsilon Z}{k}\right)$$

such that $Z \sim \mathcal{Z}$ has $\|Z\|_1 \geq 1/100$ w.h.p. (Generalizes Paninski's construction).

Idea

For private-coin lower bound against a given W, can choose $\mathcal{Z}$ to "focus" on the elements which W does not "look at" too much.

The Lower Bound (II)

Slightly less informal

For private-coin lower bound against a given W, can choose $\mathcal{Z}$ to "focus" on the subspaces orthogonal to $H(W)$'s largest ("most informative") eigenvalues.

The Lower Bound (II)

Slightly less informal

For private-coin lower bound against a given W, can choose $\mathcal{Z}$ to "focus" on the subspaces orthogonal to $H(W)$'s largest ("most informative") eigenvalues.

$\rightsquigarrow$ Leads to $\max\min$-type bounds instead of $\min\max$.

The Lower Bound (II)

Upshot

Lower bounds for learning, testing, with public- or private-coins: all depend on the corresponding χ^2-contraction factors:

$$\max_{W \in \mathcal{W}} \|H(W)\|_F \quad \text{and} \quad \max_{W \in \mathcal{W}} \|H(W)\|_*$$

The Lower Bound (II)

Upshot

Lower bounds for learning, testing, with public- or private-coins: all depend on the corresponding χ^2-contraction factors:

$$\max_{W \in \mathcal{W}} \|H(W)\|_F \quad \text{and} \quad \max_{W \in \mathcal{W}} \|H(W)\|_*$$

Bounding those gives the lower bounds.

The Lower Bound (II)

Upshot

Lower bounds for learning, testing, with public- or private-coins: all depend on the corresponding χ^2-contraction factors:

$$\max_{W \in \mathcal{W}} \|H(W)\|_F \quad \text{and} \quad \max_{W \in \mathcal{W}} \|H(W)\|_*$$

Bounding those gives the lower bounds.

Fact

Bounding those quantities in the communication-starved and the ϱ-LDP cases takes 5 lines.

The Lower Bound (III)

	Learning	Testing	
	Public/Private-Coin	Public-Coin	Private-Coin
General $\mathcal{W}$	$\frac{k}{\varepsilon^2} \cdot \frac{k}{\max_{W \in \mathcal{W}} \lVert H(W) \rVert_*}$	$\frac{\sqrt{k}}{\varepsilon^2} \cdot \frac{\sqrt{k}}{\max_{W \in \mathcal{W}} \lVert H(W) \rVert_F}$	$\frac{\sqrt{k}}{\varepsilon^2} \cdot \frac{k}{\max_{W \in \mathcal{W}} \lVert H(W) \rVert_*}$
Centralized	$\frac{k}{\varepsilon^2}$	$\frac{\sqrt{k}}{\varepsilon^2}$	
ℓ bits	$\frac{k}{2^\ell \varepsilon^2}$	$\frac{\sqrt{k}}{\varepsilon^2} \cdot \sqrt{\frac{k}{2^\ell}}$	$\frac{\sqrt{k}}{\varepsilon^2} \cdot \frac{k}{2^\ell}$
ϱ-LDP	$\frac{k^2}{\varrho^2 \varepsilon^2}$	$\frac{\sqrt{k}}{\varepsilon^2} \cdot \frac{\sqrt{k}}{\varrho^2}$	$\frac{\sqrt{k}}{\varepsilon^2} \cdot \frac{k}{\varrho^2}$

Part IV: Recap and Conclusion

How do things change under information constraints?

Pairwise distances contract: specifically, the "right" measure here is the χ^2 divergence,

$$\chi^2(p, q) = \mathbb{E}_p\left[\left(\frac{q(X)}{p(X)} - 1\right)^2\right]$$

We give a quantitative characterization of this contraction (lower bounds) and protocols achieving it.

Unified View

Gotham needed them.

Locally minimum chi-square contraction principle

Design schemes that minimize the local chi-square contraction.

Conclusion

- General framework for inference problems with local constraints over discrete distributions

Conclusion

- General framework for inference problems with local constraints over discrete distributions
- Captures the communication-starved and the locally private regimes

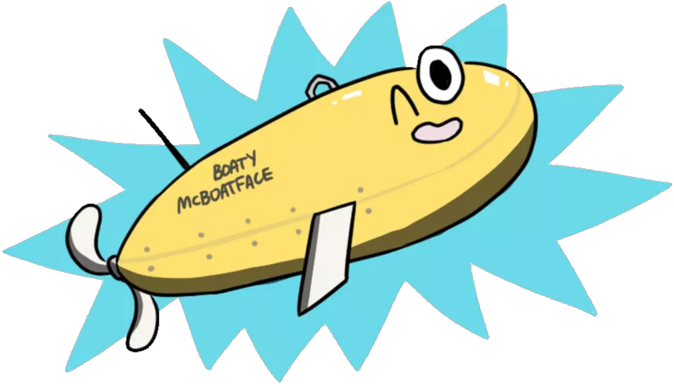
Conclusion

- General framework for inference problems with local constraints over discrete distributions
- Captures the communication-starved and the locally private regimes
- First work on distributed testing; optimal protocols for public-coin and private-coin uniformity testing in all settings considered

Conclusion

- General framework for inference problems with local constraints over discrete distributions
- Captures the communication-starved and the locally private regimes
- First work on distributed testing; optimal protocols for public-coin and private-coin uniformity testing in all settings considered
- Many questions and directions to explore: several samples, continuous case, general parametric settings (high-dimensional statistics)...

Thank you



A. Andoni, T. Malkin, and N. Shekel Nosatzki.
Two Party Distribution Testing: Communication and Security.
ArXiv e-prints, November 2018.

Jayadev Acharya, Ziteng Sun, and Huanyu Zhang.
Communication efficient, sample optimal, linear time locally private discrete distribution estimation.
CoRR, abs/1802.04705, 2018.

Ilias Diakonikolas, Elena Grigorescu, Jerry Li, Abhiram Natarajan, Krzysztof Onak, and Ludwig Schmidt.
Communication-efficient distributed learning of discrete distributions.
In *Proceedings of NIPS*, pages 6394–6404, 2017.

John C. Duchi, Michael I. Jordan, and Martin J. Wainwright.
Local privacy and statistical minimax rates.
In *Proceedings of FOCS*, pages 429–438. IEEE Computer Society, 2013.

Ilias Diakonikolas and Daniel M. Kane.
A new approach for testing properties of discrete distributions.
In *Proceedings of FOCS*. IEEE Computer Society, 2016.

Vitaly Feldman.
A general characterization of the statistical query complexity.
In Satyen Kale and Ohad Shamir, editors, *Proceedings of the 30th Conference on Learning Theory, COLT 2017, Amsterdam, The Netherlands, 7-10 July 2017*, volume 65 of *Proceedings of Machine Learning Research*, pages 785–830. PMLR, 2017.

Vitaly Feldman, Elena Grigorescu, Lev Reyzin, Santosh Vempala, and Ying Xiao.
Statistical algorithms and a lower bound for detecting planted cliques.
In Dan Boneh, Tim Roughgarden, and Joan Feigenbaum, editors, *Symposium on Theory of Computing Conference, STOC'13, Palo Alto, CA, USA, June 1-4, 2013*, pages 655–664. ACM, 2013.

Orr Fischer, Uri Meir, and Rotem Oshman.
Distributed uniformity testing.
In *Proceedings of the 2018 ACM Symposium on Principles of Distributed Computing*, PODC '18, pages 455–464, New York, NY, USA, 2018. ACM.

Oded Goldreich.
The uniform distribution is complete with respect to testing identity to a fixed distribution.
Electronic Colloquium on Computational Complexity (ECCC), 23:15, 2016.

Yanjun Han, Pritam Mukherjee, Ayfer Özgür, and Tsachy Weissman.
Distributed statistical estimation of high-dimensional and nonparametric distributions with communication constraints, February 2018.
Talk given at ITA 2018.

Yanjun Han, Ayfer Özgür, and Tsachy Weissman.
Geometric lower bounds for distributed parameter estimation under communication constraints.
volume 75 of *Proceedings of Machine Learning Research*, pages 3163–3188. PMLR, 2018.

Shiva Prasad Kasiviswanathan, Homin K. Lee, Kobbi Nissim, Sofya Raskhodnikova, and Adam D. Smith.
What can we learn privately?
SIAM J. Comput., 40(3):793–826, 2011.

Liam Paninski.
A coincidence-based test for uniformity given very sparsely sampled discrete data.
IEEE Transactions on Information Theory, 54(10):4750–4755, 2008.

Or Sheffet.
Locally private hypothesis testing.
In *Proceedings of the 35th International Conference on Machine Learning (ICML)*, volume 80, pages 4612–4621. PMLR, 10–15 Jul 2018.

Jacob Steinhardt, Gregory Valiant, and Stefan Wager.

Memory, communication, and statistical queries.
In Vitaly Feldman, Alexander Rakhlin, and Ohad Shamir, editors, *29th Annual Conference on Learning Theory*, volume 49 of *Proceedings of Machine Learning Research*, pages 1490–1516, Columbia University, New York, New York, USA, 23–26 Jun 2016. PMLR.

John N Tsitsiklis.
Decentralized detection.
In *In Advances in Statistical Signal Processing*, volume 2, 1993.

Min Ye and Alexander Barg.
Optimal schemes for discrete distribution estimation under locally differential privacy.
CoRR, abs/1702.00610, 2017.