Interactive Inference under Information Constraints

Clément Canonne (IBM Research)

Joint work with **Jayadev Acharya** (Cornell University), **Yuhan Liu** (Cornell University), **Ziteng Sun** (Cornell University), and **Himanshu Tyagi** (IISc Bangalore)

September 2, 2020

This talk

Interactive Inference under Information Constraints

This talk

Interactive Inference under Information Constraints

Distributed setting

Distributed setting

Data distributed across many users, each user with one observation.
Central server wants to perform specific task on the whole data.

Distributed setting

Noninteractive

Users can only send simultaneously a message to the server, do not get to interact between themselves.

Distributed setting

Noninteractive

Users can only send simultaneously a message to the server, do not get to interact between themselves.

(May or may not share a common random seed ahead of time.)

Distributed setting

(Sequentially) interactive

Users send a message to the server sequentially, and see messages sent by users before them.

Distributed setting

(Sequentially) interactive

Users send a message to the server sequentially, and see messages sent by users before them.

(Can assume they also share a common random seed.)

Distributed setting

Note: there exist other settings which allow for more adaptivity: multi-round sequential protocols, blackboard protocols.

Focus here on the sequentially interactive model, and whether sequentially interactive $\gg$ noninteractive.

This talk

Interactive Inference under Information Constraints

Inference

Inference tasks

Focus on density estimation (learning) and identity testing (one-sample goodness-of-fit) for discrete distributions

Inference tasks

Density estimation

n independent samples from unknown $\mathbf{p}$ over $[k] = \{1, 2, \ldots, k\}$, distance parameter $\varepsilon \in (0, 1]$. Output $\hat{\mathbf{p}}$ such that

$$\ell_1(\mathbf{p}, \hat{\mathbf{p}}) \leq \varepsilon$$

with high probability.

Inference tasks

Identity testing

n independent samples from unknown $\mathbf{p}$ over $[k] = \{1, 2, \ldots, k\}$, reference distribution $\mathbf{q}$, distance parameter $\varepsilon \in (0, 1]$. Distinguish between

$$\mathcal{H}_0 : \mathbf{p} = \mathbf{q} \qquad\qquad \mathcal{H}_1 : \ell_1(\mathbf{p}, \mathbf{q}) > \varepsilon$$

with high probability.

Inference: our goal

(Minimax) sample complexity

Characterize the minimum number of independent samples n to solve the task, as a function of k and ε, over all possible $\mathbf{p}, \mathbf{q}$.

This talk

Interactive Inference under Information Constraints

Information Constraints

Information constraints

Each user cannot simply send or fully observe their datum (sample) to the server.

Information constraints

Each user cannot simply send or fully observe their datum (sample) to the server.

- sensitive data (untrusted server)
- bandwidth constraints
- limited type of measurements
- etc.

Information constraints

Model that type of constraints in a unified fashion by a family $\mathcal{W}$ of allowed channels $W\colon [k] \to \{0,1\}^*$:

- each user chooses some $W \in \mathcal{W}$
- given data x sends message y with probability $W(y \mid x)$

Examples of information constraints

No constraint: $\mathrm{Identity} \in \mathcal{W}$

Bandwidth constraints: messages are at most ℓ bits long.

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

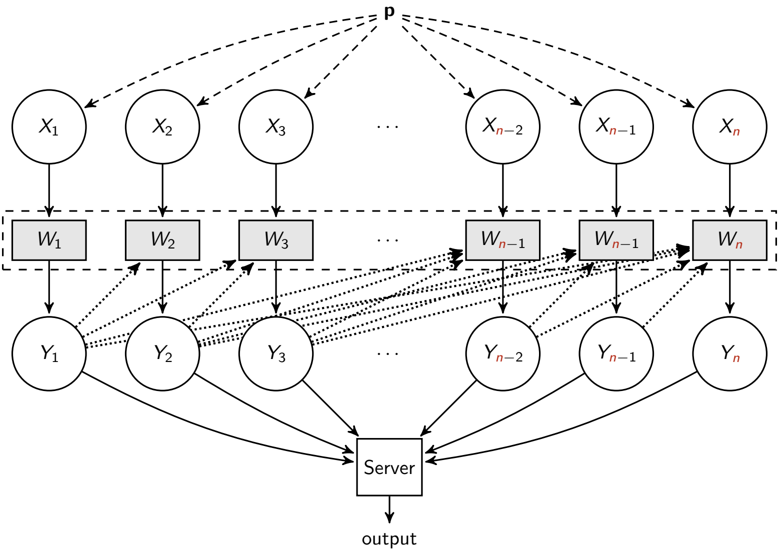
Local privacy constraints: messages must satisfy ϱ-local differential privacy (ϱ-LDP)

$$\mathcal{W} = \left\{W \colon [k] \to \{0,1\}^* : \sup_{x,x'} \sup_{y} \frac{W(y \mid x)}{W(y \mid x')} \leq e^{\varrho}\right\}$$

Linear measurements, erasure channels, quantization, "leaky-query"...

Setting: summary

Interactive Inference under Information Constraints



Prior work

Prior and related work

[Adaptive?]	Estimation	Testing
Communication	[BGM+16], [HMÖW18]*, [HÖW18]*, [ACT19a]	[Tsi93], [FMO18], [DGKR19]†
Local privacy	[DJW13a]*, [YB18], [ASZ18], [AS19], [Bas19], [BCÖ20]	[She18], [ACFT19a], [AJM20], [BB20]
General	[BHÖ19]	[ACH+20]
	[ACT18], [ACT19b]	

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

Prior and related work

[Adaptive?]	Estimation	Testing
Communication	[BGM$^+$16], [HMÖW18]*, [HÖW18]*, [ACT19a]	[Tsi93], [FMO18], [DGKR19]†
Local privacy	[DJW13a]*, [YB18], [ASZ18], [AS19], [Bas19], [BCÖ20]	[She18], [ACFT19a], [AJM20], [BB20]
General	[BHÖ19]	[ACH$^+$20]
General	[ACT18], [ACT19b], [ALCST] **(this work)**	[ACT18], [ACT19b], [ALCST] **(this work)**

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

Questions and results

Adaptivity. . .

. . . has a long history in Statistics, and a (much shorter, but very active) history in computer science and machine learning.

Yet

in many cases we don't understand how adaptivity helps in designing a protocol, or choosing a channel.

Interactive Inference under Information Constraints

Information constraints

Can we establish learning and testing lower bounds in a unified way?

Power of sequential interactivity

Does adaptivity help for these tasks? If so, for which types of constraints?

Conceptual message

Can we use the lower bounds to design better protocols (upper bounds)?

Establish learning and testing lower bounds in a unified way

To each channel W, we associate a channel information matrix $H(W)$, and set

$$\|\mathcal{W}\|_{\rm op} = \sup_{W\in\mathcal{W}} \|H(W)\|_{\rm op}, \quad \|\mathcal{W}\|_* = \sup_{W\in\mathcal{W}} \|H(W)\|_*$$

Establish bounds as a function of those spectral quantities:

how much can the users communicate by (adaptively) choosing their channels $\iff$ which directions in $\mathbb{R}^k$ can the channels let the users provide most information about

Establish learning and testing lower bounds in a unified way

	Learning		Testing	
Constraints $\mathcal{W}$	Noninteractive	Interactive	Noninteractive	Interactive
No constraint	$\frac{k}{\varepsilon^2}$		$\frac{\sqrt{k}}{\varepsilon^2}$	
General	$\frac{k}{\varepsilon^2} \cdot \frac{k}{\lVert\mathcal{W}\rVert_*}$		$\frac{\sqrt{k}}{\varepsilon^2} \cdot \frac{\sqrt{k}}{\lVert\mathcal{W}\rVert_F}$	$\frac{\sqrt{k}}{\varepsilon^2} \cdot \frac{\sqrt{k}}{\sqrt{\lVert\mathcal{W}\rVert_* \lVert\mathcal{W}\rVert_{\mathrm{op}}}}$
Bandwidth	$\frac{k}{\varepsilon^2} \cdot \frac{k}{2^\ell}$		$\frac{\sqrt{k}}{\varepsilon^2} \cdot \sqrt{\frac{k}{2^\ell}}$	$\frac{\sqrt{k}}{\varepsilon^2} \cdot \sqrt{\frac{k}{2^\ell}}$
Privacy	$\frac{k}{\varepsilon^2} \cdot \frac{k}{\varrho^2}$		$\frac{\sqrt{k}}{\varepsilon^2} \cdot \frac{\sqrt{k}}{\varrho^2}$	$\frac{\sqrt{k}}{\varepsilon^2} \cdot \frac{\sqrt{k}}{\varrho^2}$
"Leaky-Query"	$\frac{k}{\varepsilon^2} \cdot \sqrt{k}$		$\frac{\sqrt{k}}{\varepsilon^2} \cdot \sqrt{k}$	$\frac{\sqrt{k}}{\varepsilon^2} \cdot \sqrt[4]{k}$

(Bounds for noninteractive inference with a common random seed available to the users.)

Establish learning and testing lower bounds in a unified way

Consider the "local perturbation" around the reference uniform distribution u: for $z \in \{-1, 1\}^{k/2}$,

$$\forall x \in [k], \quad \mathbf{p}_z(x) = \begin{cases} \frac{1-2\varepsilon z_i}{k} & \text{if } x = 2i-1 \\ \frac{1+2\varepsilon z_i}{k} & \text{if } x = 2i \end{cases}$$

For fixed interactive protocol Π, $\mathbf{p}_z$ induces a distribution $\mathbf{p}_z^\Pi$ over messages $Y^n = (Y_1, \ldots, Y_n)$.

Goal

Assoud (learning)

Le Cam (testing)

$$k \lesssim \sum_{i=1}^{k/2} I(Z_i \wedge Y^n) \leq \text{bound}$$

$$1 \lesssim \mathrm{KL}(\mathbb{E}_Z[\mathbf{p}_Z^\Pi] \| u^\Pi) \leq \text{bound}$$

with bound as a function of $n, \mathcal{W}, k, \varepsilon$.

Establish learning and testing lower bounds in a unified way

Theorem (Information Bound)

For every $1 \leq t \leq n$,

$$\frac{1}{k}\sum_{i=1}^{k} I(Z_i \wedge Y^t) \leq \frac{t\varepsilon^2}{k^2} \cdot \|\mathcal{W}\|_*.$$

Theorem (Testing Bound)

$$\mathrm{KL}(\mathbb{E}_Z[\mathbf{p}_Z^{\Pi}] \| u^{\Pi}) \leq \frac{\varepsilon^2}{k}\|\mathcal{W}\|_{\mathrm{op}} \cdot \sum_{t=0}^{n-1}\sum_{i=1}^{k} I(Z_i \wedge Y^t).$$

(actually slightly more refined, term-wise bounds)

Adaptivity can help (but sometimes doesn't)

Immediate corollaries

Interactivity does not help for learning or testing under privacy or communication constraints.

But...

It may help for constraints $\mathcal{W}$ s.t. $\|\mathcal{W}\|_F \ll \sqrt{\|\mathcal{W}\|_* \|\mathcal{W}\|_{\mathrm{op}}}$.

This is possible

We provide a "natural" family a constraint $\mathcal{W}$ showing a maximal separation (factor $k^{1/4}$) for identity testing.

Adaptivity can help (but sometimes doesn't)

Leaky-Query constraints: $\mathcal{W} = \{W_u\}_{u \in \{0,1\}^k}$

$$W_u(y \mid x) = \begin{cases} \eta & \text{if } y = x \\ (1-\eta)u_x & \text{if } y = \mathbf{1}^* \\ (1-\eta)(1-u_x) & \text{if } y = \mathbf{0}^* \end{cases}$$

($\eta \approx 1/\sqrt{k}$: leakage parameter).

Meaning

Leaks the full data point w.p. η, otherwise indicates whether it belongs to the query set $S_u \subseteq [k]$.

Adaptivity helps for these!

Lower bound proof hints at what to do

Testing: set $\mathbf{q} = \mathbb{E}_Z[\mathbf{p}_Z^\Pi]$. Lower bound framework gives

$$\mathrm{KL}(\mathbf{q}^\Pi \| u^\Pi) \leq \frac{\varepsilon^2}{k} \boxed{\|\mathcal{W}\|_{\mathrm{op}}} \cdot \sum_{t=0}^{n-1} \sum_{i=1}^{k} I(Z_i \wedge Y^t).$$

How: using first chain rule, then χ^2 divergence

$$\begin{aligned}
&\mathbb{E}_{\mathbf{q}^{Y^t}}\left[\mathrm{KL}(\mathbf{q}^{Y_{t+1}|Y^t} \| u^{Y_{t+1}|Y^t})\right] \\
&\leq \mathbb{E}_{\mathbf{q}^{Y^t}}\left[\chi^2(\mathbf{q}^{Y_{t+1}|Y^t}, u^{Y_{t+1}|Y^t})\right] \\
&= k \cdot \mathbb{E}_{\mathbf{q}^{Y^t}}\left[\sum_y \frac{\left(\sum_x W^{Y^t}(y \mid x)(\mathbf{q}_{X_{t+1}|Y^t}(x) - \frac{1}{k})\right)^2}{\sum_x W^{Y^t}(y \mid x)}\right] \\
&= \frac{\varepsilon^2}{k} \mathbb{E}_{\mathbf{q}^{Y^t}}\left[\mathbb{E}\left[Z \mid Y^t\right]^T H(W^{Y^t}) \mathbb{E}\left[Z \mid Y^t\right]\right]. \qquad (\star)
\end{aligned}$$

Lower bound proof hints at what to do

Key step: bounding this bilinear form

$$\mathbb{E}\left[Z \mid Y^t\right]^T H(W^{Y^t})\mathbb{E}\left[Z \mid Y^t\right] \leq \boxed{\|H(W^{Y^t})\|_{\mathrm{op}}} \cdot \|\mathbb{E}\left[Z \mid Y^t\right]\|_2^2,$$

Not tight in general, but suggests protocol: user t picks channel $W \in \mathcal{W}$ s.t. $H(W)$ maximizes bilinear form. Depends on the structure of the spectrum of the $H(W)$s and the set of achievable eigenvectors!

$\rightsquigarrow$ Heuristic, but leads to optimal protocol for leaky-query channels.

Conclusion

This talk

Interactive Inference under Information Constraints

- Plug-and-play bounds for density estimation and identity testing of discrete distributions
- Corollary: tight bounds for privacy and communication constraints
- Separation between noninteractive and interactive protocols for composite hypothesis testing under natural class of constraints
- New versatile lower bound technique: further extensions to high-dimensional estimation, optimization. . .

Thank you

[arXiv:2007.10976]

Jayadev Acharya, Clément L. Canonne, Cody Freitag, and Himanshu Tyagi.
Inference under information constraints III: Local privacy constraints, 2019.
In preparation. Preprint available at arXiv:abs/1808.02174; Full version of [ACFT19b].

Jayadev Acharya, Clément L. Canonne, Cody Freitag, and Himanshu Tyagi.
Test without Trust: Optimal Locally Private Distribution Testing.
Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics, AISTATS 2019, 2019.

Jayadev Acharya, Clément L Canonne, Yanjun Han, Ziteng Sun, and Himanshu Tyagi.
Domain compression and its application to randomness-optimal distributed goodness-of-fit.
In Jacob Abernethy and Shivani Agarwal, editors, *Proceedings of Thirty Third Conference on Learning Theory*, volume 125 of *Proceedings of Machine Learning Research*, pages 3–40. PMLR, 09–12 Jul 2020.

Jayadev Acharya, Clément L. Canonne, and Himanshu Tyagi.
Inference under information constraints I: lower bounds from chi-square contraction.
CoRR, abs/1812.11476, 2018.
In submission. Full version of [ACT19c].

Jayadev Acharya, Clément L. Canonne, and Himanshu Tyagi.
Communication-constrained inference and the role of shared randomness.
In Kamalika Chaudhuri and Ruslan Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 30–39, Long Beach, California, USA, jun 2019. PMLR.

Jayadev Acharya, Clément L. Canonne, and Himanshu Tyagi.
Inference under information constraints II: Communication constraints and shared randomness, 2019.
In preparation. Preprint available at arXiv:abs/1804.06952.

Jayadev Acharya, Clément L. Canonne, and Himanshu Tyagi.
Inference under information constraints: Lower bounds from chi-square contraction.

In Alina Beygelzimer and Daniel Hsu, editors, *Proceedings of the Thirty-Second Conference on Learning Theory*, volume 99 of *Proceedings of Machine Learning Research*, pages 3–17, Phoenix, USA, jun 2019. PMLR.

Kareem Amin, Matthew Joseph, and Jieming Mao.
Pan-private uniformity testing.
In Jacob Abernethy and Shivani Agarwal, editors, *Proceedings of Thirty Third Conference on Learning Theory*, volume 125 of *Proceedings of Machine Learning Research*, pages 183–218. PMLR, 09–12 Jul 2020.

Jayadev Acharya and Ziteng Sun.
Communication complexity in locally private distribution estimation and heavy hitters.
In Kamalika Chaudhuri and Ruslan Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 51–60, Long Beach, California, USA, jun 2019. PMLR.

Jayadev Acharya, Ziteng Sun, and Huanyu Zhang.
Hadamard response: Estimating distributions privately, efficiently, and with little communication.
In *Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics (AISTATS'19)*, volume abs/1802.04705, 2018.

Raef Bassily.
Linear queries estimation with local differential privacy.
In *AISTATS*, volume 89 of *Proceedings of Machine Learning Research*, pages 721–729. PMLR, 2019.

Thomas B. Berrett and Cristina Butucea.
Locally private non-asymptotic testing of discrete distributions is faster using interactive mechanisms.
CoRR, abs/2005.12601, 2020.

Leighton P. Barnes, Wei-Ning Chen, and Ayfer Özgür.
Fisher information under local differential privacy.

CoRR, abs/2005.10783, 2020.

Mark Braverman, Ankit Garg, Tengyu Ma, Huy L. Nguyen, and David P. Woodruff.
Communication lower bounds for statistical estimation problems via a distributed data processing inequality.
In *Symposium on Theory of Computing Conference, STOC'16*, pages 1011–1020. ACM, 2016.

Leighton P. Barnes, Yanjun Han, and Ayfer Özgür.
Fisher information for distributed estimation under a blackboard communication protocol.
In *ISIT*, pages 2704–2708. IEEE, 2019.

Ilias Diakonikolas, Themis Gouleakis, Daniel M. Kane, and Sankeerth Rao.
Communication and memory efficient testing of discrete distributions.
In Alina Beygelzimer and Daniel Hsu, editors, *Proceedings of the Thirty-Second Conference on Learning Theory*, volume 99 of *Proceedings of Machine Learning Research*, pages 1070–1106, Phoenix, USA, jun 2019. PMLR.

John C. Duchi, Michael I. Jordan, and Martin J. Wainwright.
Local privacy and statistical minimax rates.
CoRR, abs/1302.3203, 2013.
Latest version, v4 (2014). Full version of [DJW13b].

John C. Duchi, Michael I. Jordan, and Martin J. Wainwright.
Local privacy and statistical minimax rates.
In *54th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2013*, pages 429–438. IEEE Computer Society, 2013.

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

Yanjun Han, Pritam Mukherjee, Ayfer Özgür, and Tsachy Weissman.

Distributed statistical estimation of high-dimensional and nonparametric distributions with communication constraints, feb 2018.
Talk given at ITA 2018.

Yanjun Han, Ayfer Özgür, and Tsachy Weissman.
Geometric lower bounds for distributed parameter estimation under communication constraints.
In *Proceedings of the 31st Conference on Learning Theory, COLT 2018*, volume 75 of *Proceedings of Machine Learning Research*, pages 3163–3188. PMLR, 2018.

Or Sheffet.
Locally private hypothesis testing.
In Jennifer Dy and Andreas Krause, editors, *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pages 4612–4621, Stockholmsmässan, Stockholm Sweden, 10–15 Jul 2018. PMLR.

John N. Tsitsiklis.
Decentralized detection.
In H. V. Poor and J. B. Thomas, editors, *Advances in Statistical Signal Processing*, volume 2, pages 297–344. JAI Press, 1993.

Min Ye and Alexander Barg.
Optimal schemes for discrete distribution estimation under locally differential privacy.
IEEE Trans. Inform. Theory, 64(8):5662–5676, 2018.