Lower Bounds for Locally Private Estimation via Communication Complexity

John Duchi and Ryan Rogers

Goal

- Lower bounds for estimation under local differential privacy constraints.
- Allow arbitrary interaction and all privacy parameters $\varepsilon \in [0, \infty]$.

Local Privacy Definitions

- A random variable Z is (ε, δ)-DP for $X \in \mathcal{X}$ if conditional on $X = x$, Z has distribution $Q(\cdot \mid x)$ and for any measurable set S and x, x', we have
$$Q(S \mid x) \le e^{\varepsilon} Q(S \mid x') + \delta.$$
- Z is (ε, α)- Renyí DP if for all x and x' we have
$$D_\alpha(Q(\cdot \mid x) \| Q(\cdot \mid x')) \le \varepsilon$$

Fully Interactive Privacy Schemes

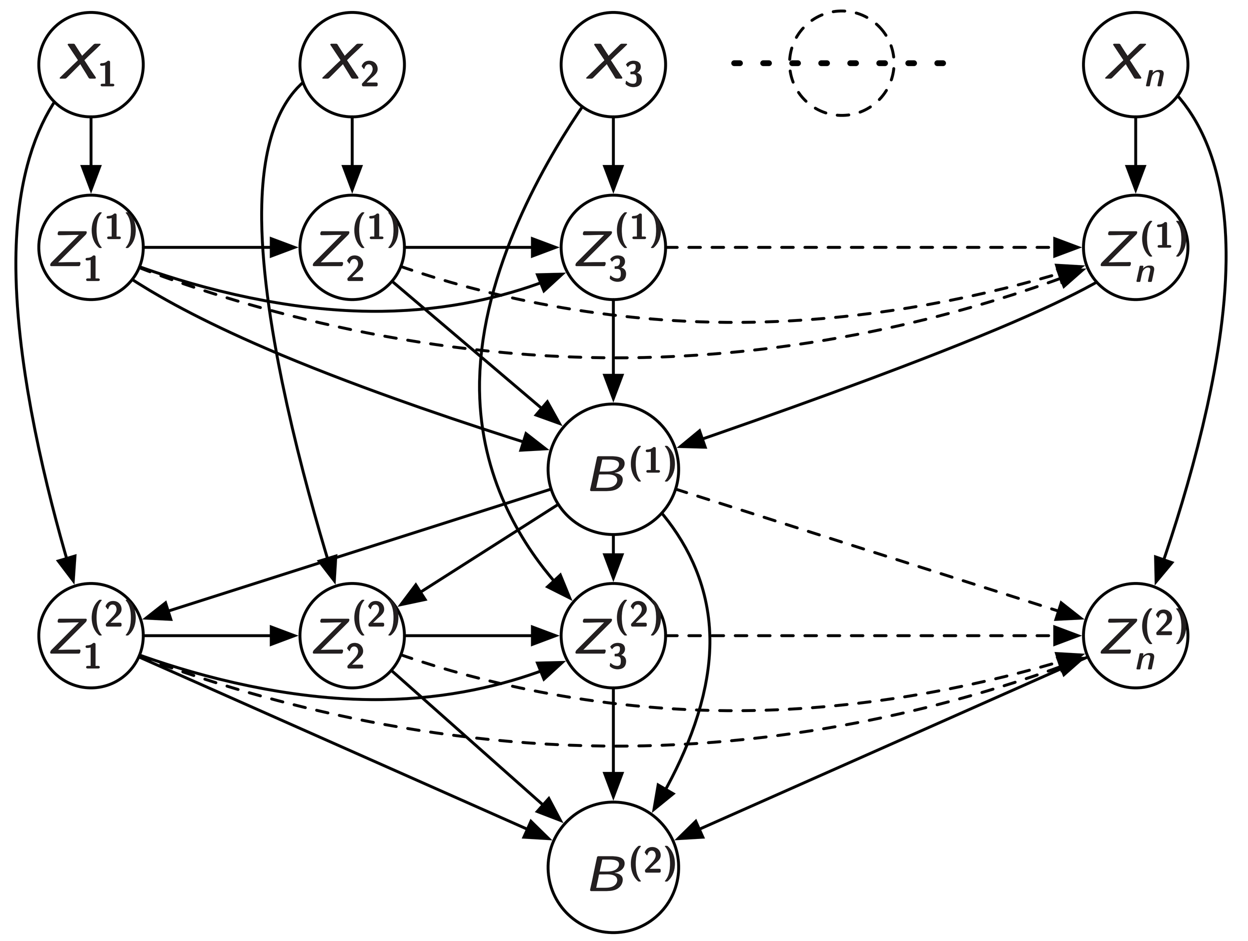
- Let $Z = \{Z_i^{(t)}\}$ be the full communication transcript
- Let the samples $x_{[1:n]}$ and $x_{[1:n]}^{(i)} \in \mathcal{X}^n$ differ in only example i, otherwise being arbitrary. The output Z is *ε-KL-locally private on average* if
$$\frac{1}{n}\sum_{i=1}^{n} D_{\rm kl}\left(Q(Z \in \cdot \mid x_{[1:n]}) \| Q(Z \in \cdot \mid x_{[1:n]}^{(i)})\right) \le \varepsilon_{\rm kl}.$$
- **Assumption 1**: The entire transcript Z is $\varepsilon_{\rm kl}$-KL-locally private on average.
- **Assumption 2**: The entire transcript Z is (ε, δ)-DP for small enough δ.
- **Note:** ε differential privacy implies $\varepsilon_{\rm kl} \le \min\{\varepsilon, \varepsilon^2\}$

Minimax Risk

- Let $\mathcal{P}$ be a collection of distributions on $\mathcal{X}$ and $\theta(P) \in \Theta \subset \mathbb{R}^d$ be a parameter of interest for $P \in \mathcal{P}$.
- Given a sample $X_1, \cdots, X_n \sim P$ and any interactive private channel Q, we get the set of privatize observations
$$Z = \{Z_1^{(1)}, Z_2^{(1)}, \cdots, Z_n^{(1)}, Z_1^{(2)}, \cdots, Z_n^{(2)}, \cdots, Z_n^{(T)}\}.$$
- $L(\hat{\theta}, \theta(P))$ is the loss for estimator $\hat{\theta}$ based on the privatized observations and the true parameter $\theta(P)$.
- The *channel minimax risk* for family $\mathcal{P}$, parameter θ, and loss L is
$$\mathfrak{M}_n(\theta(\mathcal{P}), L, Q) = \inf_{\hat{\theta}} \sup_{P \in \mathcal{P}} \mathbb{E}_{P,Q}\left[L\left(\hat{\theta}(Z), \theta(P)\right)\right].$$

Main Results

Take Home: Effective sample size reduction from n to $n \cdot \min\{\varepsilon, \varepsilon^2, d\}/d$



- **Bernoulli Estimation**: Let $\mathcal{P}_d$ be the collection of Bernoulli distributions on $\{0,1\}^d$ and $L(\theta, \theta') = \sum_{j=1}^d \ell(\theta_j - \theta_j')$ for ℓ symmetric then
$$\mathfrak{M}_n(\theta(\mathcal{P}_d), L, Q) \gtrsim d \cdot \ell\left(\sqrt{\frac{d}{n\varepsilon_{\rm kl}}}\right)$$
- **Logistic Risk**: Let $\mathcal{P}_d$ be the collection of logistic distributions with $\ell(\theta; x, y) = \log\left(1 + e^{-y\langle x, \theta\rangle}\right)$, $R_P(\theta) = \mathbb{E}_P[\ell(\theta; (X, Y)]$, and excess risk
$$L(\theta, \theta(P)) = R_P(\theta) - R_P(\theta(P)).$$
Then
$$\mathfrak{M}_n(\theta(\mathcal{P}_d), L, Q) \gtrsim \frac{d}{n} \cdot \frac{d}{\varepsilon_{\rm kl}}$$
- **Gaussian Estimation** (only for Assumption 1): Let $\mathcal{P}_d$ be the collection of Gaussians $N(\theta, \sigma^2 I_d)$ and $\theta \in [-1, 1]^d$ and $\sigma > 0$ is known, then
$$\mathfrak{M}_n(\theta(\mathcal{P}_d), \|\cdot\|_2^2, Q) \gtrsim d \cdot \min\left\{1, \max\left\{\frac{d}{\varepsilon_{\rm kl}} \cdot \frac{\sigma^2}{n}, \frac{\sigma^2}{n}\right\}\right\}$$
- **k-sparse Gaussian Estimation** (only for Assumption 1): Let $\mathcal{P}_d$ be the collection of k-sparse Gaussians $N(\theta, \sigma^2 I_d)$ and $\theta \in [-1, 1]^d$ and $\sigma > 0$ is known, then
$$\mathfrak{M}_n(\theta(\mathcal{P}_d), \|\cdot\|_2^2, Q) \gtrsim k \cdot \min\left\{1, \max\left\{\frac{d}{\varepsilon_{\rm kl}} \cdot \frac{\sigma^2}{n}, \frac{\sigma^2 \log(d/k)}{n}\right\}\right\}$$

Achievability and Analysis

- The lower bounds are achievable. See results in [?]
- Minimax lower bounds build off work in communication limits in estimation [ZDJW13, GMN14, BGM+16].
- Bounds follow from mutual information calculations and communication structure

Extensions

- Also consider *Compositional* locally private schemes [JMNR19], where each randomizer is locally private while ensuring the sum of privacy parameters is bounded.
- Results apply when d-dimensional parameters that are "independent" of each other; when correlations exist between coordinates the lower bounds do not apply.
- Interesting question: can leveraging correlation improve locally private estimation?

Acknowledgements

We thank Vitaly Feldman, Aleksandar Nikolov, Aaron Roth, Adam Smith, and Salil Vadhan for clarifying discussions and feedback on earlier versions of this work, which (among other things) led us to the general KL-locally private on average definition. We also thank the Simons Institute for hosting our visit as part of the *Data Privacy: Foundations and Applications* semester.

References

[BDF+18] Bhowmick, Duchi, Freudiger, Kapoor, and Rogers. Protection Against Reconstruction and Its Applications in Private Federated Learning. *arXiv e-prints*, page arXiv:1812.00984, Dec 2018.

[BGM+16] 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. 2016.

[GMN14] Ankit Garg, Tengyu Ma, and Huy L. Nguyen. On communication cost of distributed statistical estimation and dimensionality. 2014.

[JMNR19] Matthew Joseph, Jieming Mao, Seth Neel, and Aaron Roth. The role of interactivity in local differential privacy. *arXiv:1904.03564 [cs.LG]*, 2019.

[ZDJW13] Yuchen Zhang, John C. Duchi, Michael I. Jordan, and Martin J. Wainwright. Information-theoretic lower bounds for distributed estimation with communication constraints. 2013.