COLT2021 Tutorial – Recitation

Jayadev Acharya Clément Canonne Himanshu Tyagi

August 4, 2021

This document contains: (1) a short prelude, to check that some of the bounds obtained or claimed during the lecture part of the tutorial make sense; (2) a two-parter online component, where you will derive upper and lower bounds on some high-dimensional estimation problems under bandwidth or privacy constraints; (3) extra exercises, if you want additional material to go through at home. Detailed solutions will be made available, for we are not total monsters.

Exercise 0.1 (Sanity Checks)**.** *We have seen that, for discrete distribution estimation (in total variation/ℓ_1 distance) over $\{1,2,\ldots,d\}$, we could derive a lower bound of*

$$\Omega\left(\frac{d^2}{2^\ell \varepsilon^2}\right)$$

*on the sample complexity under ℓ-bits information constraints, and $\Omega\left(\frac{d^2}{\rho^2\varepsilon^2}\right), \Omega\left(\frac{d^2}{e^\rho\varepsilon^2}\right)$ under LDP constraints ($\rho \in (0,1]$ and $\rho > 1$, respectively). Comment on those bounds: do they pass basic "sanity checks"? (*Hint: *Recall that,* without *constraints, the tight bound is known to be $\Theta\left(\frac{d}{\varepsilon^2}\right)$.)*

1 Online: Mean Estimation, Upper and Lower Bounds

In what follows, we consider identity-covariance Gaussian distributions with bounded mean:

$$\mathcal{G}_d \coloneqq \{N(\mu, \mathbb{I}_d) : \mu \in \mathbb{R}^d, \|\mu\|_\infty \le 1\}$$

and Bernoulli mean products:

$$\mathcal{B}_d \coloneqq \{\otimes_{i=1}^d \mathrm{Rad}(\frac{1}{2}(\mu_i + 1)) : \mu \in \mathbb{R}^d, \|\mu\|_\infty \le 1\}$$

(product distributions over $\{\pm 1\}^d$ where the ith coordinate has mean μ_i (equals 1 with probability $\frac{\mu_i+1)}{2}$).

We will establish bounds on the sample complexity of mean estimation (under ℓ_2^2 loss) for both ℓ-bit communication constraints, and ρ-LDP. The first part of the recitation focuses on applying the framework discussed in the lectures to derive lower bounds (for the Bernoulli case) against interactive protocols. The second part will use some ideas evoked in the first part of the tutorial, as well as some general "tricks," to obtain matching upper bounds.

As a baseline to keep in mind, estimating the mean of either identity-covariance Gaussian distributions or Bernoulli products to ℓ_2 loss $\varepsilon \in (0,1]$ has sample complexity

$$\Theta\left(\frac{d}{\varepsilon^2}\right)$$

without any constraint.

1.1 Lower bound on Bernoulli Mean Estimation

We will use the general lower bound framework discussed in the tutorial to obtain a sample complexity lower bound for Bernoulli product mean estimation (i.e., the family $\mathcal{B}_d$) in ℓ_2^2 loss, under ℓ-bits communication and ρ-local privacy constraints. Here is what we want to use:

Consider a family of 2^d "hard distributions" $\mathcal{P} \subseteq \mathcal{B}_d$, indexed by $\mathcal{Z} \coloneqq \{\pm 1\}^d$, which satisfies the assumptions. For us, $\mathcal{X} = \{\pm 1\}^d$ (our Bernoulli product distributions are over $\mathcal{X}$), and we will assume for notational simplicity that $\mathcal{Y}$ is discrete.

Assumption 1 (Additive loss)**.** *Fix $\varepsilon > 0$ and $\tau \in (0, 1/2]$. For all $z, z' \in \mathcal{Z}$,*

$$\ell_2(\mu_z, \mu_{z'}) \geq 4\varepsilon \left(\frac{\|z - z'\|_0}{\tau d} \right)^{1/2}$$

Assumption 2 (Densities exist)**.** *For all $z \in \mathcal{Z}$ and $i \in [d]$, there are $\alpha_{z,i} \in \mathbb{R}$ and $\phi_{z,i} \colon \mathcal{X} \to \mathbb{R}$ s.t.*

$$\frac{d\mathbf{p}_{z^{\oplus i}}}{d\mathbf{p}_z} = 1 + \alpha_{z,i}\phi_{z,i}$$

and $|\alpha_{z,i}| \leq \alpha$, where $\alpha \in \mathbb{R}$ is a fixed constant independent of z, i.

Assumption 3 (Bounded Ratios)**.** *There exists $\lambda \in [1, \infty]$ s.t.*

$$\sup_{z \in \mathcal{Z}} \sup_{y \in \mathcal{Y}} \sup_{W \in \mathcal{W}} \frac{\mathbb{E}_{X \sim \mathbf{p}_{z^{\oplus i}}}[W(y \mid X)]}{\mathbb{E}_{X \sim \mathbf{p}_z}[W(y \mid X)]} \leq \lambda$$

Assumption 4 (Orthonormality)**.** *For all $z \in \mathcal{Z}$ and $1 \leq i, j \leq d$,*

$$\mathbb{E}_{X \sim \mathbf{p}_z}[\phi_{z,i}(X)\phi_{z,j}(X)] = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise.} \end{cases}$$

Assumption 5 (Subgaussianity)**.** *There is $\sigma \geq 0$ s.t., for all $z \in \mathcal{Z}$ and $1 \leq i, j \leq d$,*

$$\phi_z(X) \coloneqq (\phi_{z,1}(X), \ldots, \phi_{z,d}(X)) \in \mathbb{R}^d$$

is σ^2-subgaussian for $X \sim \mathbf{p}_z$, with independent coordinates.

Then we have the following:

Theorem 1. *Suppose Assumptions 1 to 3 hold, and let ε, τ as in Assumption 1. Suppose Π is an interactive protocol using $\mathcal{W}$ with n users, with expected ℓ_2^2 loss at most ε^2 over $\{\mathbf{p}_Z\}_Z$, when $Z \sim \mathrm{Rad}(\tau)^{\otimes d}$. Then, (1) if Assumption 4 holds,*

$$n \cdot \frac{\alpha^2}{d} \min\left(\lambda, \frac{1}{\tau}\right) \max_{z \in \mathcal{Z}} \max_{W \in \mathcal{W}} \sum_{y \in \mathcal{Y}} \frac{\mathrm{Var}_{X \sim \mathbf{p}_z}[W(y \mid X)]}{\mathbb{E}_{X \sim \mathbf{p}_z}[W(y \mid X)]} = \Omega(1)$$

and (2) if Assumption 5 holds,

$$n \cdot \frac{\alpha^2\sigma^2}{d} \min\left(\lambda, \frac{1}{\tau}\right) \max_{z \in \mathcal{Z}} \max_{W \in \mathcal{W}} H(\mathbf{p}_z^W) = \Omega(1)\,,$$

where $H(\mathbf{p}_z^W)$ is the Shannon entropy of the distribution $\mathbf{p}_z^W$ induced by $\mathbf{p}_z$ and W on $\mathcal{Y}$ (and thus is at most $\log_2 |\mathcal{Y}|$).

Construction. Now, what should we choose as $\mathcal{P}$? *Discussion:* A natural choice is to have the discrepancy between means "spread out" uniformly over all coordinates (all coordinates of the mean to estimate are equally important) and to make each coordinate as unbiased as possible (since biased v. unbiased coin is the "hardest" in one dimension). This motivates setting, for some "suitable" $\gamma > 0$,

$$\mu_z \coloneqq \gamma z \in [-1,1]^d, \qquad \mathbf{p}_z \coloneqq \otimes_{i=1}^d \mathrm{Rad}(\frac{1}{2}(\mu_{z,i}+1)) \tag{1}$$

for every $z \in \{\pm 1\}^d$ (and $\mathcal{P} \coloneqq \{\mathbf{p}_z\}_{z\in\{\pm1\}^d}$). Moreover, we'll consider the uniform prior of $\mathcal{Z}$, i.e., $Z \sim \mathrm{Rad}(1/2)^{\otimes d}$. *Discussion:* We don't have any sparsity constraint on μ, and all coordinates have a symmetric role, so uniform prior makes sense: each μ_i is chosen to be $\pm\gamma$ independently.

Exercise 1.1. *From Eq.* (1)*, write the expression for* $\mathbf{p}_z(x)$*, for arbitrary* $z \in \{\pm1\}^d$ *and* $x \in \{\pm1\}^d$*.*

Exercise 1.2. *Given our choice of "dense" prior (uniform prior over* $\mathcal{Z}$*), which of the 5 assumptions do we not need to check?*

Exercise 1.3. *Given our construction of* $\mathbf{p}_z$ *(Eq.* (1)*) and* τ*, how should we set* γ*? (*Hint: *Assumption 1.)*

Exercise 1.4. *Show that if* $\alpha_{z,i}, \phi_{z,i}$ *satisfy Assumption 2, then* $\mathbb{E}_{X\sim\mathbf{p}_z}[\phi_{z,i}(X)] = 0$*.*

Exercise 1.5. *Find* $\alpha_{z,i}$*,* $\phi_{z,i}$*,* α *to satisfy Assumptions 2 and 4. (*Hint: *Start by finding the expression for the product* $\alpha_{z,i}\phi_{z,i}$*. Then compute* $\mathbb{E}_{\mathbf{p}_z}\left[(\alpha_{z,i}\phi_{z,i})^2\right]$ *and normalise to get* $\alpha_{z,i}$*, since* $\mathbb{E}_{\mathbf{p}_z}\left[\phi_{z,i}^2\right]$ *must equal 1 for Assumption 4. Finally, use the previous exercise to show Assumption 4 holds.)*

We so far have shown Assumptions 1, 2 and 4 hold (and argued we didn't need Assumption 3). We already can apply part of Theorem 1!

Exercise 1.6. *Apply Theorem 1 to get the LDP lower bound for mean estimation to* ℓ_2 *loss* ε*: for* $\rho \in (0,1]$*, we must have*

$$n = \Omega\left(\frac{d^2}{\varepsilon^2\rho^2}\right) \tag{2}$$

and some *communication-constrained lower bound: for* $\ell \geq 1$*, we must have*

$$n = \Omega\left(\frac{d^2}{\varepsilon^2 2^\ell}\right). \tag{3}$$

*(*Hint: *Use the statement from Exercise 2.4 for LDP, seen in the tutorial; for communication constraints, use* $\mathrm{Var}[W] \leq \mathbb{E}[W]$*.)*

The bound for communication constraints in Eq. (3) can be significantly improved, however. To do so, let's show Assumption 5 holds. First, some useful facts.

Recall that a mean-zero r.v. $U \in \mathbb{R}^d$ is σ^2-subgaussian if, for every unit vector $u \in \mathbb{R}^d$, the univariate r.v. $\langle u, U\rangle$ is σ^2-subgaussian: $\mathbb{E}\left[e^{t\langle u,U\rangle}\right] \leq e^{\sigma^2t^2/2}$ for all $t \in \mathbb{R}$. We also have the following:

Lemma 1 (Hoeffding's Lemma)**.** *Let* X *be a real-valued r.v. s.t.* $a \leq X \leq b$ *almost surely. Then*

$$\mathbb{E}\left[e^{t(X-\mathbb{E}[X])}\right] \leq e^{\frac{t^2(b-a)^2}{8}}$$

for all $t \in \mathbb{R}$*.*

Exercise 1.7. *Show that Assumption 5 holds for* $\sigma^2 = \frac{1+\gamma}{1-\gamma}$*. Use then the second part of Theorem 1 the communication-constrained lower bound*

$$n = \Omega\left(\frac{d^2}{\varepsilon^2\ell}\right) \tag{4}$$

for $\ell \geq 1$*. (*Hint: *Use Hoeffding's lemma.)*

In the next part, we will show that Eqs. (2) and (4) are optimal, and are actually achieved by *noninteractive* (even more, private-coin!) protocols.

1.2 Upper Bound on Gaussian and Bernoulli Mean Estimation

We now will complement the lower bounds (against interactive protocols) from the previous section by a *noninteractive, private-coin* protocol achieving the same sample complexity. The idea will be to first get the communication-constrained upper bound using 2 techniques:

- Reducing Gaussian to Bernoulli product mean estimation *(simpler! Everything is bits!)*
- Using SIMULATE-AND-INFER for Bernoulli products

and, finally, to use the communication-constrained algorithm for $\ell = 1$, along with a simple LDP protocol (Randomized Response) to get the LDP upper bound.

Exercise 1.8 (Reduction between Bernoulli Product and Gaussian)**.** *Suppose one has a mean estimation protocol for $\mathcal{B}_d$ under ℓ_2^2 loss with sample complexity n. Show that this implies a mean estimation protocol for $\mathcal{G}_d$ under ℓ_2^2 loss with sample complexity $O(n)$. Is the converse true? (*Hint: Erf *is well-behaved on* $[-1,1]$*.)*

Exercise 1.9 (Upper Bound for Bernoulli Product)**.** *Use the exercise above to obtain a noninteractive, private-coin protocol for Gaussian mean estimation (under ℓ_2^2 loss) under ℓ-bit communication constraints with sample complexity $O\left(\frac{d^2}{\varepsilon^2 \min(\ell,d)}\right)$. (*Hint: *Simulate-and-Infer.)*

Exercise 1.10 (From communication to local privacy)**.** *Deduce, for $\rho \in (0,1]$, the existence of a noninteractive, private-coin protocol for Gaussian mean estimation (under ℓ_2^2 loss) under ρ-LDP (no communication constraints) with sample complexity $O\left(\frac{d^2}{\varepsilon^2\rho^2}\right)$. (*Hint: *Case $\ell = 1$, and Randomized Response.)*

Exercise 1.11 (The boundedness assumption)**.** *Generalize the above exercises to*

$$\mathcal{G}_d(R) \coloneqq \{N(\mu, \mathbb{I}_d) : \mu \in \mathbb{R}^d, \|\mu\|_\infty \le R\}$$

where $R > 0$ is a parameter given as input to the protocol. Can we set $R = \infty$?

2 Extra (offline): More exercises

The exercises marked with an asterisk do not have a solution provided. Feel free to email us if you're stuck!

2.1 Lower bounds: more sanity checks, and practise

Exercise 2.1. *Recall that we were able to derive a $\Omega\left(\frac{d^2}{\varepsilon^2\ell}\right)$ lower bound for Gaussian mean estimation under ℓ-bits communication constraints. Why did we only get a 2^ℓ for discrete distribution estimation? (*Hint: *Check the subgaussianity+independence assumption. Also, the upper bound, for a "Well, duh!" answer.)*

Exercise 2.2 $(*)$**.** *Suppose you are in a setting where each user sends their data via an erasure (oblivious) communication channel, which just "swallows" the message with fixed probability $\eta \in [0,1]$. Model this as a family $\mathcal{W}_\eta$ of local constraints, and derive a sample complexity lower bound for discrete distribution estimation (in total variation distance). (*Hint: $\mathcal{W}_\eta \subseteq \{W : [d] \to [d] \cup \{\perp\}\}$*.)*
Complement this with a matching upper bound.

2.2 Simulate-and-Infer

Exercise 2.3 (Simulate-and-Infer)**.** *Suppose you want to prove a (noninteractive) lower bound on the sample complexity of some estimation problem $\mathcal{P}$ under ℓ-bit communication constraints, of the form*

$$n = \Omega\left(\frac{f(\mathcal{P})}{2^\ell}\right).$$

Show that it suffices to prove it for $\ell = 1$. Does it extend to interactive protocols?

2.3 Connection to local privacy

Exercise 2.4. *Fix any $\rho > 0$, and suppose that $W\colon \mathcal{X} \to \mathcal{Y}$ is a ρ-locally private channel (for simplicity of notation, with discrete output space). Show that, for any probability distribution $\mathbf{p}$ on $\mathcal{X}$, we have*

$$\sum_{y\in\mathcal{Y}} \frac{\mathrm{Var}_{X\sim\mathbf{p}}[W(y \mid X)]}{\mathbb{E}_{X\sim\mathbf{p}}[W(y \mid X)]} \le \min\big((e^\rho - 1)^2, e^\rho\big)\,.$$

(In particular, for $\rho \in [0,1]$ the RHS is $O(\rho^2)$.)

2.4 Generalizing to per-user different LDP or comnunication constraints

This exercise asks you to generalize the arguments from Section 1 (both upper and lower bounds) to the setting where each user has a possibly different local privacy parameter $\rho_i \in (0,1]$ (or a possibly different bandwidth constraint $\ell_i \ge 1$).

Exercise 2.5 $(*)$**.** *Suppose user i (for $1 \le i \ne n$) has a local privacy parameter $\rho_i \in (0,1]$, and let $\bar\rho^2 := \frac{1}{n}\sum_{i=1}^n \rho_i^2$. Generalize the proofs of Section 1 for ℓ_2 mean estimation for Bernoulli products (or identity-covariance Gaussians) to obtain (i) a noninteractive, private-coin protocol with sample complexity*

$$O\left(\frac{d^2}{\varepsilon^2\bar\rho^2}\right)$$

and (ii) a matching sample complexity lower bound of $\Omega\left(\frac{d^2}{\varepsilon^2\bar\rho^2}\right)$ against interactive protocols. Prove the analogous statement if each user has a communication constraint ℓ_i, letting $\bar\ell := \frac{1}{n}\sum_{i=1}^n \ell_i$: i.e., establish a bound of $\Theta\left(\frac{d^2}{\varepsilon^2\min(\bar\ell,d)}\right)$.

*(*Hint: *For the lower bound, recall (cf. tutorial, or check the paper!) that in Theorem 1 the "$n\sup_z \sup_{W\in\mathcal{W}}$" can be replaced by "$\sum_{i=1}^n \sup_z \sup_{W\in\mathcal{W}_i}$."*)

2.5 From parameter estimation to estimating Z

Exercise 2.6 $(*)$**.** *Let $\mathcal{P}_\Theta := \{\mathbf{p}_\theta\}_{\theta\in\Theta}$ be a family of probability distributions over $\mathcal{X}$ parameterized by $\Theta \subseteq \mathbb{R}^k$, and $\{\mathbf{p}_z\}_{z\in\{\pm1\}^d} \subseteq \mathcal{P}_\Theta$. For simplicity, for each $z \in \{\pm1\}^d$ we denote by θ_z the parameter $\theta \in \Theta$ corresponding to $\mathbf{p}_z$. Suppose that there exists $\gamma > 0$ such that*

$$\|\theta_z - \theta_{z'}\|_2^2 \ge \gamma\cdot\|z - z'\|_0, \qquad \forall z, z' \in \{\pm1\}^d\,.$$

Discussion: *Compare this assumption to Assumption 1. Show that, given an estimator $\hat\theta\colon \mathcal{X}^n \to \Theta$ for $\mathcal{P}_\Theta$ such that*

$$\sup_{\theta\in\Theta} \mathbb{E}_{\mathbf{p}_\theta}\left[\|\theta - \hat\theta\|_2^2\right] \le \varepsilon^2$$

one can obtain an estimator $\hat Z\colon \mathcal{X}^n \to \{\pm1\}^d$ such that

$$\sup_{z\in\{\pm1\}^d} \mathbb{E}_{\mathbf{p}_z}\left[\|z - \hat Z\|_0\right] \le \frac{4\varepsilon^2}{\gamma}$$

What does this imply if $\{\mathbf{p}_z\}_{z\in\{\pm1\}^d}$ is defined by $\theta_z = \frac{10\varepsilon}{\sqrt{d}}z$? $\theta_z = \varepsilon z$?