B9145: Problem Set 2

Due: Oct 30, 11:59pm

Carefully follow submission instructions announced on Canvas.

Question 2.1 (Minimax bounds for estimation (30 points))**:** We derive information theoretic lower bounds for statistical estimation problems, analogous to those for stochastic optimization we saw in class. For a class of distributions $\mathcal{P}$, let $\theta : \mathcal{P} \to \mathbb{R}^d$ be the statistical functional of interest; $\theta(P)$ is often called the "parameter". Let d be a metric on $\Theta := \{\theta(P) : P \in \mathcal{P}\}$, and let $\Phi : \mathbb{R}_+ \to \mathbb{R}_+$ be a non-decreasing function such that $\Phi(0) = 0$. For n observations $X_i \stackrel{\text{iid}}{\sim} P$, we measure performance of an estimator $\widehat{\theta}_n(X_1, \ldots, X_n)$ by

$$\sup_{P \in \mathcal{P}} \mathbb{E}_{X_1^n \sim P} \left[\Phi \left(d(\widehat{\theta}(X_1^n), \theta(P)) \right) \right].$$

The minimax risk for estimation is given by

$$\mathfrak{M}_n(\mathcal{P}, \Phi \circ d) := \inf_{\widehat{\theta}} \sup_{P \in \mathcal{P}} \mathbb{E}_P \left[\Phi \left(d(\widehat{\theta}(X_1^n), \theta(P)) \right) \right],$$

where the infimum is taken over all measurable functions of $X_1, \ldots, X_n$.

For parts (a)-(b), you may give a concise derivation based on results from class.

(a) Derive Le Cam's method: for any fixed $\delta > 0$, and $P_1, P_{-1} \in \mathcal{P}$ such that $d(\theta(P_1), \theta(P_{-1})) \geq 2\delta$,

$$\mathfrak{M}_n(\mathcal{P}, \Phi \circ d) \geq \frac{\Phi(\delta)}{2} \left(1 - \left\| P_1^n - P_{-1}^n \right\|_{\text{TV}} \right).$$

(b) Derive Assouad's method. Let $\mathcal{V} := \{-1, +1\}^d$ be the binary hypercube, and let $\{P_v^n\}_{v \in \mathcal{V}}$ be a collection of distributions on X_1^n. We say $\{P_v^n\}_{v \in \mathcal{V}}$ is δ-separated in the Hamming distance if there exists a mapping $\widehat{v} : \Theta \mapsto \mathcal{V}$ such that

$$\Phi \left(d(\theta, \theta(P_v)) \right) \geq \delta \sum_{j=1}^{d} \mathbf{1} \left\{ \widehat{v}(\theta)_j \neq v_j \right\}.$$

Define $P_{+j}^n := \frac{1}{2^{d-1}} \sum_{v : v_j = 1} P_v^n$ and $P_{-j}^n := \frac{1}{2^{d-1}} \sum_{v : v_j = -1} P_v^n$. Then, we have

$$\mathfrak{M}_n(\mathcal{P}, \Phi \circ d) \geq \frac{\delta}{2} \sum_{j=1}^{d} \left(1 - \left\| P_{+j}^n - P_{-j}^n \right\|_{\text{TV}} \right)$$

whenever $\{P_v^n\}_{v \in \mathcal{V}}$ is δ-separated in Hamming distance.

(c) Consider the normal location model $\mathcal{P}_{\sigma^2} := \{N(\theta, \sigma^2 I) : \theta \in \mathbb{R}^d\}$, where I is the d-by-d dimensional identity matrix, and $\sigma^2 > 0$ is a fixed variance. We're interested in estimating the location parameter θ in the squared Euclidean distance $\|\cdot\|_2^2$. Show the following lower bound

$$\mathfrak{M}_n\left(\mathcal{P}_{\sigma^2}, \|\cdot\|_2^2\right) \geq \frac{d\sigma^2}{16n}. \tag{1}$$

(d) Argue that the lower bound (1) is tight up to numerical constants.

Question 2.2 (Differentially private estimation (50 points))**:** We study estimation under a privacy constraint, when the data collector cannot be trusted with sensitive information. Instead of observing true data $X_i \in \mathcal{X}$, a perturbed version $Z_i \in \mathcal{Z}$ is viewed; given $X = x$, we write $Z \sim Q(\cdot \mid X = x)$, and call Q a "channel". For $\alpha > 0$, we say Z_i is *α-differentially private* if for any measurable subset $A \subset \mathcal{Z}$ and any pair $x, x' \in \mathcal{X}$,

$$\frac{Q(Z \in A \mid X = x)}{Q(Z \in A \mid X = x')} \leq \exp(\alpha). \tag{2}$$

Intuitively, differential privacy asks that x and x' are similarly likely to have generated the observed signal Z. Letting $q(z \mid x) := Q(Z = z \mid X = x)$ be the conditional density of $Z \mid X$, the condition (2) is equivalent to $\frac{q(z|x)}{q(z|x')} \leq e^\alpha$ for all $x, x' \in \mathcal{X}$, and almost surely all $z \in \mathcal{Z}$. In what follows, we assume $\alpha < 1$.

As we will show, differential privacy acts as a contraction on probabilities. For arbitrary probabilities P_1, P_2 on $\mathcal{X}$, let densities p_1 and p_2 be their densities w.r.t. a base measure μ; you may treat this as a continuous density for convenience. Define the *marginal* distributions

$$M_i(Z \in A) := \int_{\mathcal{X}} Q(Z \in A \mid X = x) p_i(x) d\mu(x), \quad i \in \{1, 2\}.$$

We will prove there is a universal (numerical) constant $C < \infty$ such that for any P_1, P_2,

$$D_{\rm kl}\left(M_1 \| M_2\right) + D_{\rm kl}\left(M_2 \| M_1\right) \leq C(e^\alpha - 1)^2 \left\|P_1 - P_2\right\|_{\rm TV}^2. \tag{3}$$

We show this result assuming $\mathcal{Z} = \{1, \ldots, k\}$ for some finite $k \in \mathbb{N}$; this is without loss of generality, but you don't have to justify this.

(a) Recall the definition of the total variation distance $\|P_1 - P_2\|_{\rm TV} = \sup_{A \subset \mathcal{X}} \{P_1(A) - P_2(A)\}$. Show $\|P_1 - P_2\|_{\rm TV} = \frac{1}{2} \int |p_1(x) - p_2(x)| d\mu(x)$.

(b) Define $m_j(z) := \int q(z \mid x) p_j(x) d\mu(x)$, prove that for a universal constant $c < \infty$,

$$|m_1(z) - m_2(z)| \leq c(e^\alpha - 1) \inf_{x \in \mathcal{X}} q(z \mid x) \cdot \left\|P_1 - P_2\right\|_{\rm TV}.$$

(c) Show the result (3) when $\mathcal{Z} = \{1, \ldots, k\}$ for some finite $k \in \mathbb{N}$.

Hint Use the following simple inequality: for any $a, b > 0$, we have $\left|\log \frac{a}{b}\right| \leq \frac{|a-b|}{\min\{a,b\}}$. To see this, use $\log(1 + x) \leq x$ to note

$$\log \frac{a}{b} = \log\left(1 + \frac{a}{b} - 1\right) \leq \frac{a - b}{b} \quad \text{and} \quad \log \frac{b}{a} \leq \frac{b - a}{a}.$$

We now use the inequality (3) to prove minimax lower bounds for differentially private estimation. Consider a survey data on individuals $i = 1, \ldots, n$, where we ask each individual about illicit drug use: $X_i = 1$ if person i uses illicit drugs, 0 otherwise ($\mathcal{X} = \{0, 1\}$). Define $\theta(P) = P(X = 1) = \mathbb{E}_P[X]$. To protect privacy, we perturb each answer X_i in a α-differentially private manner, and use Z_i's as our data.

To make sure everyone feels suitably private, assume $\alpha < 1/2$; in this case, $(e^\alpha - 1)^2 \le 2\alpha^2$. Let $\mathcal{Q}_\alpha$ be the family of all α-differentially private channels, and let $\mathcal{P}$ denote the Bernoulli distributions with parameter $\theta(P) = P(X_i = 1) \in [0, 1]$. We consider the minimax risk for private estimation of the proportion $\theta(P)$

$$\mathfrak{M}_n(\theta(\mathcal{P}), |\cdot|, \alpha) := \inf_{Q \in \mathcal{Q}_\alpha} \inf_{\widehat{\theta}} \sup_{P \in \mathcal{P}} \mathbb{E}\left[|\widehat{\theta}(Z_1, \ldots, Z_n) - \theta(P)|\right],$$

where the infimum is over (differentially private) channels Q and estimators $\widehat{\theta}$, and the expectation is taken with respect to both the X_i (according to P) and the Z_i (according to $Q(\cdot \mid X_i)$).

(d) Use Le Cam's method to argue that whenever P_1, P_2 satisfy $|\theta(P_1) - \theta(P_2)| \ge \delta$,

$$\mathfrak{M}_n(\theta(\mathcal{P}), |\cdot|, \alpha) := \inf_{Q \in \mathcal{Q}_\alpha} \inf_{\widehat{\theta}} \sup_{P \in \mathcal{P}} \mathbb{E}\left[|\widehat{\theta}(Z_1, \ldots, Z_n) - \theta(P)|\right] \ge \frac{\delta}{2} \inf_{Q \in \mathcal{Q}_\alpha} [1 - \|M_1^n - M_2^n\|_{\mathrm{TV}}].$$

Then, use inequality (3) to show that for some universal constant $c' > 0$.

$$\mathfrak{M}_n(\theta(\mathcal{P}), |\cdot|, \alpha) \ge \frac{c'}{\sqrt{n\alpha^2}}.$$

(e) Give a rate-optimal estimator for this problem. i.e., define a α-differentially private channel Q and an estimator $\widehat{\theta}$ such that $\mathbb{E}[|\widehat{\theta}(Z_1^n) - \theta|] \le C'/\sqrt{n\alpha^2}$, where $C' > 0$ is a universal constant.

Hint Consider perturbing the data with probability $1 - q_\alpha$, where $q_\alpha = e^\alpha/(1 + e^\alpha)$. Note that $(2q_\alpha - 1)^{-2} = \left(\frac{e^\alpha+1}{e^\alpha-1}\right)^2 \approx 4/\alpha^2$ for $\alpha \approx 0$.

Question 2.3 (Adversarial robustness for linear logistic regression (10 points))**:** Consider a binary classification problem with label $y \in \{-1, +1\}$ and features $x \in \mathbb{R}^d$. We study the logistic regression loss $\ell(\theta; x, y) = -\log \sigma(y\theta^\top x)$, where $\sigma(a) = \frac{1}{1+\exp(-a)}$. Derive an alternative form for the adversarial loss:

$$\max_{\bar{x} \in \mathbb{R}^d : \|\bar{x} - x\|_\infty \le \epsilon} \ell(\theta; \bar{x}, y) = -\log \sigma\left(y\theta^\top x - \epsilon\|\theta\|_1\right).$$

Give an interpretation of this result.

Question 2.4 (ICLR 2020 Vision talk (10 points))**:** Watch Ruha Benjamin's ICLR talk on "Reimagining the default settings of technology and society" via the url `https://iclr.cc/virtual_2020/speaker_3.html`. In 2-3 sentences, discuss how this may relate to your research, or other professional activities.