CMPUT 605: Theoretical Foundations of Reinforcement Learning, Winter 2023
Homework #4

Instructions

Submissions You need to submit a single PDF file, named `p04_<name>.pdf` where `<name>` is your name. The PDF file should include your typed up solutions (we strongly encourage to use pdfLaTeX). Write your name in the title of your PDF file. We provide a LaTeXtemplate that you are encouraged to use. To submit your PDF file you should send the PDF file via private message to Vlad Tkachuk on Slack before the deadline.

Collaboration and sources Work on your own. You can consult the problems with your classmates, use books or web, papers, etc. Also, the write-up must be your own and you must acknowledge all the sources (names of people you worked with, books, webpages etc., including class notes.) Failure to do so will be considered cheating. Identical or similar write-ups will be considered cheating as well. Students are expected to understand and explain all the steps of their proofs.

Scheduling Start early: It takes time to solve the problems, as well as to write down the solutions. Most problems should have a short solution (and you can refer to results we have learned about to shorten your solution). Don't repeat calculations that we did in the class unnecessarily.

Deadline: March 26 at 11:55 pm

Large action set query lower bound

We recall a few definitions and results from Lecture 9. For a featurized MDP (M, ϕ), let

$$\varepsilon^*(M, \Phi) := \sup_{\pi \text{ memoryless}} \inf_{\theta \in \mathbb{R}^d} \|\Phi\theta - q^\pi\|_\infty . \tag{1}$$

Definition 1. An online planner is (δ, ε)-sound if for any finite discounted MDP $M = (\mathcal{S}, \mathcal{A}, P, r, \gamma)$ and feature-map $\varphi : \mathcal{S} \times \mathcal{A} \to \mathbb{R}^d$ such that $\varepsilon^*(M, \Phi) \le \varepsilon$, when interacting with (M, φ), the planner induces a δ-suboptimal policy of M.

The following was proven in the said lecture:

Theorem 1 (Query lower bound: large action sets)**.** *For any $\varepsilon > 0$, $0 < \delta \le 1/2$, positive integer d and for any (δ, ε)-sound online planner $\mathcal{P}$ there exists a featurized-MDP (M, φ) with rewards in $[0, 1]$ with $\varepsilon^*(M, \Phi) \le \varepsilon$ such that when interacting with a simulator of (M, φ), the expected number of queries used by $\mathcal{P}$ is at least $\Omega(f(d, \varepsilon, \delta))$ where*

$$f(d, \varepsilon, \delta) = \exp\left(\frac{1}{32}\left(\frac{\sqrt{d}\varepsilon}{\delta}\right)^2\right) .$$

Question 1. The lecture notes provide a proof sketch for this theorem. Formally prove this theorem, explicitly explain each step of your proof.

Total: **20 points**

Solution. First, recall the JL feature matrix construction:

Proposition 1 (JL feature matrix). *For any $\tau > 0$, integers $d, k > 0$ such that*

$$d \le k \le \exp\left(\frac{d\tau^2}{8}\right), \tag{2}$$

there exists a matrix $\Phi \in \mathbb{R}^{k\times d}$ such that for any $i \in [k]$,

$$\max_{i\in[k]} \inf_{\theta\in\mathbb{R}^d} \|\Phi\theta - e_i\|_\infty \le \tau\,, \tag{3}$$

where e_i is the ith basis vector of standard Euclidean basis of $\mathbb{R}^k$, and in particular if $\varphi_i^\top$ is the ith row of Φ, $\|\Phi\varphi_i - e_i\|_\infty \le \tau$ holds.

Fix the planner $\mathcal{P}$ with the said properties. Let k be a positive integer to be chosen later. We construct a feature map $\varphi : \mathcal{S}\times\mathcal{A} \to \mathbb{R}^d$ and k MDPs $M_1, \ldots, M_k$ that share $\mathcal{S} = \{s, s_{\text{end}}\}$ and $\mathcal{A} = [k]$ as state- and action-spaces, respectively. Here s will be chosen as the initial state where the planners will be tested from and s_{end} will be an absorbing state with zero reward. The MDPs share the same deterministic transition dynamics: All actions in s end up in s_{end} with probability one and all actions taken in s_{end} end up in s_{end} with probability one. The rewards for actions taken in s_{end} are all zero. Finally, we choose the reward of MDP M_i in state s to be

$$r_a^{(i)}(s) = \mathbb{I}(a = i)r^*\,,$$

where the value of $r^* \in [0, 1]$ is left to be chosen later.

Fix $i \in [k]$. Let $\mathbb{P}$ be the probability distribution induced by the interconnection of planner $\mathcal{P}$ and MDP M_i. While this depends on i (since planners use observed rewards to plan and the reward distributions between the MDPs are different), this dependence is suppressed to minimize clutter. Let $\mathbb{E}$ be the corresponding expectation operator.

Let A be the action that is chosen by planner $\mathcal{P}$ when fed with initial state s. Let $\bar{r} = \mathbb{E}[r_A^{(i)}(s)]$. By the MDPs construction, the value of the policy induced by the planner in state s and MDP M_i is $v := \bar{r}$. Note that the optimal value in state s is r^*. By our assumption, $\mathcal{P}$ is (δ, ε)-sound. Hence, provided that we can construct an appropriate feature map so that M_i satisfies $\varepsilon^*(M_i, \Phi) \le \varepsilon$, we must have

$$\bar{r} \ge r^* - \delta\,. \tag{4}$$

Now, choose

$$r^* = 2\delta$$

which makes the right-hand side of (4) $r^*/2$. Note that $r^* \in [0, 1]$ since we assumed that $\delta \le 1/2$. Thus, we have $\bar{r} \ge r^*/2$. By construction, all the rewards are zero except the reward of action i. Hence, $\bar{r} = \mathbb{P}(A = i)r^*$ and thus we get that

$$\mathbb{P}(A = i) \ge 1/2\,.$$

Hence, for i when the planner $\mathcal{P}$ is interconnected with the simulator of M_i, it returns action i with at least probability $1/2$. As a result, we can use the planner to search any binary array of length k for the single nonzero entry in the array. Indeed, assume that we want to use the planner to search in the array $b \in \{0, 1\}^n$. Then, when the planner queries (s, a) with $a \in [k]$, we issue a query to the array to get the value of b_a. Then we feed the planner with s (as the next state) and the reward $b_a r^*$. Clearly, if and only if b is such that $b_i = 1$, this way we simulate the data that would be generated if the planner was interconnected with MDP M_i. Finally, when the planner stops and outputs A, we return A. By our previous argument,

$\mathbb{P}(A = i) \geq 1/2$. Hence, by the high-probability needle lemma, the expected number of queries used by $\mathcal{P}$ on at least one of the k MDPs is at least $\Omega(k)$.

It remains to choose k and the feature-map. For this, we use JL feature matrix. First. note that the action-value functions of the memoryless policies in any of the k MDPs belong to the set

$$\{q_i : \mathcal{S} \times \mathcal{A} \to \mathbb{R} \,:\, q_i(s_{\text{end}}, \cdot) = \mathbf{0}\,, q_i(s,a) = r^* \mathbb{I}(a = i), a \in \mathcal{A}, i \in [k]\}.$$

Hence, we need a feature-map that approximates the functions in this set uniformly well up to an ε accuracy. Fix $\tau > 0$ to be chosen later. Take the JL feature matrix $\Phi \in \mathbb{R}^{k \times d}$ such that Eq. (3) holds for this τ. Let the rows of Φ be $\varphi_1, \ldots, \varphi_k$. We let

$$\phi(s,k) = \varphi_k\,, \quad \phi(s_{\text{end}}, k) = \mathbf{0}\,, \qquad k \in [A]\,.$$

Then, to approximate the function q_i with some $i \in [k]$, we set $\theta = r^* \varphi_i$. Clearly, $q_i(s_{\text{end}}, \cdot) = \phi(s_{\text{end}}, \cdot)^\top \theta = \mathbf{0}$. For s and for $a \in [k]$ we have

$$|\varphi(s,a)^\top \theta - q_i(s,a)| = |r^* \varphi_a^\top \varphi_i - q_i(s,a)| \leq \begin{cases} r^* \tau\,, & \text{if } a \neq i\,; \\ 0\,, & \text{otherwise}\,. \end{cases}$$

To control the error of approximating q_i, it suffices to choose τ so that $r^* \tau \leq \varepsilon$. Recalling $r^* \tau = 2\delta\tau$, we choose τ so that $2\delta\tau = \varepsilon$. Finally, to choose the value of k we plug in the value of τ into Eq. (2) and get

$$k = \left\lfloor \exp\left(\frac{d(\frac{\varepsilon}{2\delta})^2}{8} \right) \right\rfloor .$$

The proof is finished by recalling that the expected number of queries made by $\mathcal{P}$ is at least $\Omega(k)$ on at least one of $M_1, \ldots, M_k$. □

Fixed-horizon fundamental theorem

The same lecture stated the fundamental theorem for fixed-horizon problems, which we copy here for convenience. For the definitions of the quantities used in the theorem, see the lecture notes.

Theorem 2 (Fixed-horizon fundamental theorem)**.** *We have $v_0^* \equiv \mathbf{0}$ and for any $h \geq 0$, $v_{h+1}^* = T v_h^*$. Furthermore, for any $\pi_0^*, \ldots, \pi_h^*, \ldots$ such that for $i \geq 0$, π_i^* is greedy with respect to v_i^*, for any $h > 0$ it holds that $\pi = (\pi_{h-1}^*, \ldots, \pi_0^*, \ldots)$ (i.e., the policy which in step 0 uses π_{h-1}^*, in step 1 uses π_{h-2}^*, ..., in step $(h-1)$ uses π_0^*, after which it continues arbitrarily) is h-step optimal:*

$$v_h^\pi = v_h^*\,.$$

In the lecture notes we did not give a proof.

Question 2. Prove Theorem 2. **Hint**: Use induction and mimic the previous proofs.

Total: **50 points**

Solution. We follow the advice by mimicking the proofs seen before. We need some definitions. For a policy π and for (s,a), $i \geq 0$ arbitrary, let $\nu_{\mu,i}^\pi(s,a) = \mathbb{P}_\mu^\pi(S_i = s, A_i = a)$. By abusing notation, we also let $\nu_{\mu,i}^\pi(s) = \mathbb{P}_\mu^\pi(S_i = s)$. First, we show the following:
Claim 1: For any $\pi = (\pi_0, \pi_1, \ldots)$ policy, $\mu \in \mathcal{M}_1(\mathcal{S})$, there is a nonstationary memoryless policy $\pi' = (\pi_0', \pi_1', \ldots)$ such that for any $i \geq 0$, $(s,a) \in \mathcal{S} \times \mathcal{A}$, $\nu_{\mu,i}^\pi(s,a) = \nu_{\mu,i}^{\pi'}(s,a)$ (here, $\pi_0', \pi_1', \ldots$ are memoryless policies).

Fix π. The policy π' is defined by $(\pi'_0, \pi'_1, \dots)$ where for $i \geq 0$, π'_i is defined using

$$\pi'_i(a|s) = \frac{\nu^\pi_{\mu,i}(s,a)}{\nu^\pi_{\mu,i}(s)}\,, \qquad (s,a) \in \mathcal{S} \times \mathcal{A}\,.$$

Clearly, π'_i is a memoryless policy. We now claim that for any s, $\nu^\pi_{\mu,i}(s) = \nu^{\pi'}_{\mu,i}(s)$. We prove this by induction on i. For $i = 0$, the claim is clearly true as $\nu^\pi_{\mu,0}(s) = \mu(s) = \nu^{\pi'}_{\mu,0}(s)$. Assume that the claim holds for $i \geq 0$. Then, applying the law of total probability and using definitions of $\mathbb{P}^\pi_\mu$ and $\mathbb{P}^{\pi'}_\mu$,

$$\begin{aligned}
\nu^\pi_{\mu,i+1}(s') &= \sum_{s,a} \nu^\pi_{\mu,i}(s,a)\mathbb{P}^\pi_\mu(S_{i+1} = s'|S_i = s, A_i = a) \\
&= \sum_{s,a} \nu^\pi_{\mu,i}(s,a)P_a(s,s') \\
&= \sum_{s,a} \nu^{\pi'}_{\mu,i}(s,a)P_a(s,s') && \text{(by the I.H.)} \\
&= \sum_{s,a} \nu^{\pi'}_{\mu,i}(s,a)\mathbb{P}^{\pi'}_\mu(S_{i+1} = s'|S_i = s, A_i = a) \\
&= \nu^{\pi'}_{\mu,i+1}(s')\,.
\end{aligned}$$

It also immediately follows that for any $i \geq 0$ and any s, a,

$$\begin{aligned}
\nu^\pi_{\mu,i}(s,a) &= \frac{\nu^\pi_{\mu,i}(s,a)}{\nu^\pi_{\mu,i}(s)}\nu^\pi_{\mu,i}(s) \\
&= \pi'_i(a|s)\nu^{\pi'}_{\mu,i}(s) \\
&\overset{(*)}{=} \frac{\nu^{\pi'}_{\mu,i}(s,a)}{\nu^{\pi'}_{\mu,i}(s)}\nu^{\pi'}_{\mu,i}(s) \\
&= \nu^{\pi'}_{\mu,i}(s,a)\,,
\end{aligned}$$

where the equality marked by $(*)$ follows because π' is a sequence of memoryless policies and the ith policy in π' is exactly π'_i. This finishes the proof of the claim.

From this follows our second claim:

Claim 2: For any $\pi = (\pi_0, \pi_1, \dots)$ policy, $s \in \mathcal{S}$ there is a nonstationary memoryless policy $\pi' = (\pi'_0, \pi'_1, \dots)$ such that for any $i \geq 0$, $v^\pi_i(s) = v^{\pi'}_i(s)$.

Fix π, $s \in \mathcal{S}$. Let π' be as in the previous claim. By abusing notation, let $r(s,a) = r_a(s)$ so that for $\nu \in \mathcal{M}_1(\mathcal{S} \times \mathcal{A})$, $\nu r = \sum_{s,a} \nu(s,a)r(s,a)$ is well-defined. Then, $v^\pi_i(s) = \sum_{t=0}^{i-1} \mathbb{E}^\pi_s[r_{A_t}(S_t)] = \sum_{t=0}^{i-1} \nu^\pi_{\mu,t} r = \sum_{t=0}^{i-1} \nu^{\pi'}_{\mu,t} r = v^{\pi'}_i(s)$, finishing the proof.

Denote by NML the set of nonstationary memoryless policies. Hence, for $i \geq 0$,

$$v^*_i(s) = \sup_\pi v^\pi_i(s) = \sup_{\pi \in \mathrm{NML}} v^\pi_i(s)\,. \tag{5}$$

We prove the statement by induction on i, but we first recall the statement itself. Let us start by recalling the definition of π^*_i: π^*_i is greedy with respect to v^*_i:

$$T_{\pi^*_i} v^*_i = T v^*_i\,.$$

Now, for $i > 0$, let $\tilde{\pi}_i = (\pi^*_{i-1}, \pi^*_{i-2}, \dots, \pi^*_0, \dots)$. The following needs to be proven:

1. $v^*_0 = \mathbf{0}$;
2. $v^*_h = T v^*_{h-1}$ holds for $h \geq 1$;

3. $v_h^{\tilde{\pi}_h} = v_h^*$ holds for $h \geq 1$;

First, let $h = 0$. By definition, for any policy π, $v_0^\pi = \mathbf{0}$. Hence, $v_0^* = \mathbf{0}$. We prove the second two statements together, by induction on h. For the base case let $h = 1$. We need to prove that

$$v_1^* = v_1^{\tilde{\pi}_1} = T\mathbf{0}\,.$$

Take any nonstationary memoryless policy $\pi = (\pi_0, \pi_1, \dots)$. It is easy to see that for any $h \geq 1$,

$$v_h^\pi = T_{\pi_0} \dots T_{\pi_{h-1}} \mathbf{0}\,. \tag{6}$$

In particular, for $h = 1$,

$$v_1^\pi = T_{\pi_0}\mathbf{0}\,.$$

Then,

$$v_1^\pi \leq T\mathbf{0} = T_{\pi_0^*}\mathbf{0} = v_1^{\tilde{\pi}_1} \leq v_1^*\,.$$

Now fix s. Taking the supremum over π, by Eq. (5),

$$v_1^*(s) = \sup_{\pi \in \mathrm{NML}} v_1^\pi(s) \leq (T\mathbf{0})(s) = (T_{\pi_0^*}\mathbf{0})(s) = v_1^{\tilde{\pi}_1}(s) \leq v_1^*(s)\,,$$

hence equality holds everywhere above. Since s is arbitrary,

$$v_1^* = T\mathbf{0} = v_1^{\tilde{\pi}_1}$$

as required.

Now let $h > 1$ and assume that the statement has been proven up to $h - 1$. Again, fix an arbitrary nonstationary memoryless policy π. Then, by Eq. (6),

$$\begin{aligned}
v_h^\pi &= T_{\pi_0} T_{\pi_1} \dots T_{\pi_{h-1}} \mathbf{0} \\
&\leq T_{\pi_0} v_{h-1}^* && (T_{\pi_0} \text{ monotone}, T_{\pi_1} \dots T_{\pi_{h-1}}\mathbf{0} = v_{h-1}^{(\pi_1, \pi_2, \dots)} \leq v_{h-1}^*) \\
&\leq T v_{h-1}^* && (T_{\pi_0} \leq T) \\
&= T_{\pi_{h-1}^*} v_{h-1}^* && (\text{def. of } \pi_{h-1}^*) \\
&= T_{\pi_{h-1}^*} v_{h-1}^{\tilde{\pi}_{h-1}} && (\text{induction hypothesis}) \\
&= T_{\pi_{h-1}^*} \dots T_{\pi_0^*} \mathbf{0} && (\text{def. of } \tilde{\pi}_{h-1}, \text{ Eq. (6)}) \\
&= v_h^{\tilde{\pi}_h} && (\text{def. of } \pi_h, \text{ Eq. (6)}) \\
&\leq v_h^*\,. && (\text{def. of } v_h^*)
\end{aligned}$$

As before, fixing a state, taking the supremum over π, by Eq. (5),

$$v_h^* \leq T v_{h-1}^* = v_h^{\tilde{\pi}_h} \leq v_h^*\,,$$

and hence we have equality everywhere. This finishes the inductive step and thus the proof. □

Statisticians also have limits

Let $\mathcal{X}$ be a subset of a Euclidean space equipped with the usual Borel σ-algebra, $\mathcal{P} \subset \mathcal{M}_1(\mathcal{X})$ a set of probability distributions over $\mathcal{X}$. Let $f : \mathcal{P} \to \mathbb{R}$ be a fixed function. We consider statistical estimation problems where a random "data" $X \in \mathcal{X}$ is observed from an unknown $P \in \mathcal{P}$ and the job of the statistician is to produce an estimate of $f(P)$.

That is, the statistician needs to design an estimator; for simplicity we assume that the estimators are not randomizing (an extension to randomizing estimators is trivial). A non-randomizing estimator maps the data to a real; thus, any such estimator is a map $g : \mathcal{X} \to \mathbb{R}$. We assume that g is measurable so that we can talk about the probability of errors.

In particular, for $\delta \in [0,1]$ and $\varepsilon > 0$, we say that g is (δ, ε)**-sound** for the problem specified by $(\mathcal{P}, f)$ if for any $P \in \mathcal{P}$,

$$P(|g(X) - f(P)| > \varepsilon) \leq \delta\,. \tag{7}$$

Here, $X : \mathcal{X} \to \mathcal{X}$ is treated as the identity map, as usual: $X(x) = x$, $x \in \mathcal{X}$. Thus, the above probability is the probability assigned by P to the set

$$\{x \in \mathcal{X} \,:\, |g(x) - f(P)| > \varepsilon\}$$

and condition (7) has the equivalent form that for any $P \in \mathcal{P}$,

$$P(\{x \in \mathcal{X} \,:\, |g(x) - f(P)| > \varepsilon\}) \leq \delta\,.$$

It is just shorter and more elegant to write Eq. (7), hence, we will stick to this usual form.

For two probability measures, P, Q, over the same measurable space $(\Omega, \mathcal{F})$, we define their **relative entropy** by

$$D(P,Q) = \begin{cases} \int \log \frac{dP}{dQ}(\omega)\, dP(\omega)\,, & \text{if } P \ll Q \\ +\infty\,, & \text{otherwise}\,. \end{cases}$$

The relative entropy is also known as the Kullback-Leibler divergence between P and Q (see Chapter 14 in the bandit book for an explanation of its origin and some examples).

The following result, which is Theorem 14.12 in that book, will be useful:

Theorem 3 (Bretagnolle–Huber inequality)**.** *Let P and Q be probability measures on the same measurable space $(\Omega, \mathcal{F})$, and let $A \in \mathcal{F}$ be an arbitrary event. Then,*

$$P(A) + Q(A^c) \geq \frac{1}{2} \exp\left(-D(P,Q)\right)\,, \tag{8}$$

where $A^c = \Omega \setminus A$ is the complement of A.

Question 3. Show that if there is an (δ, ε)-sound estimator for $(\mathcal{P}, f)$ then

$$\log\left(\frac{1}{4\delta}\right) \leq \inf\{D(P_0, P_1) \,:\, P_0, P_1 \in \mathcal{P} \text{ s.t. } |f(P_0) - f(P_1)| > 2\varepsilon\}\,.$$

In words, distributions whose f-values are separated by 2ε cannot be too close to each other if a (δ, ε)-sound estimator exist. This should be quite intuitive.

Total: **20 points**

Solution. Let $\Omega = \mathcal{X}$ and $\mathcal{F}$ be the corresponding Borel σ-algebra. Let g be a (δ, ε)-sound estimator for $(\mathcal{P}, f)$. Pick $P_0, P_1 \in \mathcal{P}$ such that

$$|f(P_0) - f(P_1)| > 2\varepsilon \tag{9}$$

Let $A = \{|g(X) - f(P_0)| > \varepsilon\}$. From Eq. (8),

$$D(P_0, P_1) \geq \log\left(\frac{1}{2(P_0(A) + P_1(A^c))}\right)\,.$$

Hence, it suffices to show that $P_0(A) + P_1(A^c) \le 2\delta$.

By definition,

$$\delta \ge P_0(A)\,.$$

Also by definition and by Eq. (9),

$$\delta \ge P_1(|g(X) - f(P_1)| > \varepsilon) \ge P_1(|g(X) - f(P_0)| \le \varepsilon) = P_1(A^c)\,.$$

In particular, the second inequality holds because for any x such that $|g(x) - f(P_0)| \le \varepsilon$, by Eq. (9), $|g(x) - f(P_1)| > \varepsilon$. Putting things together, $P_0(A) + P_1(A^c) \le 2\delta$ and thus $D(P_0, P_1) \ge \log(1/(4\delta))$. Taking the infimum over P_0 and P_1 gives the result. □

In what follows, we will deal with Bernoulli random variables. The relative entropy between Bernoulli distributions has special properties which we will find useful. The next problem asks you to prove some of these properties.

Let $\mathrm{Ber}(p)$ denote the Bernoulli distribution with parameter $p \in [0,1]$. As it is well known (and not hard to see from the definition),

$$D(\mathrm{Ber}(p), \mathrm{Ber}(q)) = d(p,q)$$

where $d(p,q)$ is the so-called **binary relative entropy function**, which is defined as

$$d(p,q) = p\log(p/q) + (1-p)\log((1-p)/(1-q))\,.$$

Question 4. Show that for $p, q \in (0,1)$, defining p^* to be p or q depending on which is further away from $1/2$,

$$d(p,q) \le \frac{(p-q)^2}{2p^*(1-p^*)}\,. \tag{10}$$

Hint: Notice that $d(p,q) = D_R((p,1-p),(q,1-q))$, where D_R is Bregman divergence with respect to our old friend, the unnormalized negentropy R over $[0,\infty)^2$. Then use Theorem 26.12 from the bandit book.

Total: **20 points**

Solution. Let R be the unnormalized negentropy over $[0,\infty)^2$. Then, by Theorem 26.12, for any $x, y \in (0,\infty)^2$,

$$D_R(x,y) = \frac{1}{2}\|x-y\|^2_{\nabla R(z)}$$

for some z on the line segment connecting x to y. We have $R(z) = z_1\log(z_1) + z_2\log(z_2) - z_1 - z_2$. Hence, $\nabla R(z) = [\log(z_1), \log(z_2)]^\top$ and $\nabla R(z) = \mathrm{diag}(1/z_1, 1/z_2)$, both defined for $z \in (0,\infty)^2$. Thus,

$$D_R(x,y) = \frac{(x_1-y_1)^2}{2z_1} + \frac{(x_2-y_2)^2}{2z_2}\,.$$

Now choosing $x = (p, 1-p)$, $y = (q, 1-q)$, we see that $x, y \in (0,\infty)^2$ if $p, q \in (0,1)$. In this case, with some $\alpha \in [0,1]$, $z = \alpha x + (1-\alpha)y = (\alpha p + (1-\alpha)q, \alpha(1-p) + (1-\alpha)(1-q))^\top = (\alpha p + (1-\alpha)q, 1 - (\alpha p + (1-\alpha)q))^\top$. Hence, $z_2 = 1 - z_1$ and

$$d(p,q) = \frac{(p-q)^2}{2z_1} + \frac{(p-q)^2}{2(1-z_1)} = \frac{(p-q)^2}{2z_1(1-z_1)}\,.$$

Now, $z_1(1-z_1) \ge p^*(1-p^*)$ (the function $z \mapsto z(1-z)$ has a maximum at $z = 1/2$ and is decreasing on "either side" of the line $z = 1/2$). Putting things together, we get

$$d(p,q) = \frac{(p-q)^2}{2z_1(1-z_1)} \le \frac{(p-q)^2}{2p^*(1-p^*)}\,.$$

□

Now, for $n > 0$ let $\mathrm{Ber}^{\otimes n}(p)$ denote the n-fold product of $\mathrm{Ber}(p)$ with itself, so that if $X \sim \mathrm{Ber}^{\otimes n}(p)$ then $X = (X_1, \dots, X_n)$ where $X_i \sim \mathrm{Ber}(p)$ and $(X_1, \dots, X_n)$ is an independent sequence.

Take $\mathcal{X} = \{0,1\}^n$ and $\mathcal{P}_n = \{\mathrm{Ber}^{\otimes n}(p) : p \in [0,1]\}$. Let $f : \mathcal{P}_n \to [0,1]$ be defined by $f(\mathrm{Ber}^{\otimes n}(p)) = p$. The problem specified by $(\mathcal{P}_n, f)$ is the problem of estimating the parameter of a Bernoulli distribution given n independent observations from the said, unknown distribution.

Question 5. Show that for the Bernoulli estimation problem described above, for $\delta \in [0,1]$ and $0 \le \varepsilon^2 < 1/32$ fixed, there is no (δ, ε)-sound estimator of the common mean, unless $n \ge \frac{\log(1/(4\delta))}{16\varepsilon^2}$.
Hint: Use that $D(P^{\otimes n}, Q^{\otimes n}) = nD(P,Q)$ and the statements from the previous two problems.

Total: **20 points**

Solution. Assume that there is a (δ, ε)-sound estimator for the said problem. By Question 3,

$$\log\left(\frac{1}{4\delta}\right) \le \inf\{D(P_0, P_1) : P_0, P_1 \in \mathcal{P}_n \text{ s.t. } |f(P_0) - f(P_1)| > 2\varepsilon\}\,.$$

Let $P_i = \mathrm{Ber}^{\otimes n}(p_i)$, $i \in \{0,1\}$. By the hint and Question 4, for $p_0 = 1/2$, $p_1 = 1/2 + \varepsilon'$, $p^* = p_1$ and $p^*(1-p^*) = 1/4 - (\varepsilon')^2$, hence

$$D(P_0, P_1) = nd(p_0, p_1) \le n\frac{(p_0 - p_1)^2}{2p^*(1-p^*)} \le \frac{2n(\varepsilon')^2}{1 - 4(\varepsilon')^2}\,.$$

Using that $f(P_i) = p_i$, we get

$$\log\left(\frac{1}{4\delta}\right) \le 2n \inf\left\{\frac{(\varepsilon')^2}{1-4(\varepsilon')^2} : \varepsilon' > 2\varepsilon\right\} = \frac{2n(2\varepsilon)^2}{1-4(2\varepsilon)^2}\,,$$

where the equality follows because $x \mapsto x^2/(1-4x^2)$ is increasing on $[0, 1/2)$. Reordering, we get that

$$n \ge \frac{\log\left(\frac{1}{4\delta}\right)}{8\varepsilon^2}(1 - 16\varepsilon^2) \ge \frac{\log\left(\frac{1}{4\delta}\right)}{16\varepsilon^2}\,,$$

where the last inequality follows from $\varepsilon^2 \le 1/32$, finishing the proof. □

Now consider the problem when the definition of f is changed to

$$f_\gamma(\mathrm{Ber}^{\otimes n}(p)) = \frac{1}{1-\gamma p}\,, \tag{11}$$

where $0 < \gamma < 1$.

Question 6. Show that for the Bernoulli estimation problem described above with $f = f_\gamma$ as in Eq. (11), with some constants $\gamma_0 > 0$ and $c_0, c_1 > 0$, for $\delta \in [0,1]$, $\varepsilon \le c_0/(1-\gamma)$, $\gamma \ge \gamma_0$, the necessary condition for the existence of (δ, ε)-sound estimator for $(\mathcal{P}_n, f_\gamma)$ is that $n \ge c_1 \frac{\log(1/(4\delta))}{(1-\gamma)^3\varepsilon^2}$.
Hint: Use the same strategy as in the solution of the previous exercise.

Total: **40 points**

Solution. Similarly to the previous calculations,

$$\log\left(\frac{1}{4\delta}\right) \le n \inf\left\{\frac{(p_0 - p_1)^2}{2p^*(1-p^*)} : |f(p_0) - f(p_1)| > 2\varepsilon, p_0, p_1 \in (0,1)\right\}\,.$$

Let

$$p_0 = \frac{4}{3} - \frac{1}{3\gamma}\,.$$

It will be useful to note that

$$1 - \gamma p_0 = \frac{4}{3}(1-\gamma)\,, \tag{12}$$

and

$$1 - p_0 = \frac{1-\gamma}{3\gamma}\,. \tag{13}$$

Choose γ_0 so that for any $\gamma \geq \gamma_0$, $p_0 \geq 1/2$. We calculate

$$f'(p) = \frac{\gamma}{(1-\gamma p)^2}\,.$$

Note that both f and f' are increasing. Now, for $p_1 > p_0$, for some $z \in [p_0, p_1]$, we have

$$f(p_1) = f(p_0) + f'(z)(p_1 - p_0) \geq f(p_0) + f'(p_0)(p_1 - p_0)$$

and thus $f(p_1) > f(p_0) + 2\varepsilon$ if $f(p_0) + f'(p_0)(p_1 - p_0) > f(p_0) + 2\varepsilon$, or, equivalently, $p_1 - p_0 > 2\varepsilon / f'(p_0)$. Note that

$$p_0' := p_0 + 2\varepsilon / f'(p_0) < 1 \tag{14}$$

provided that

$$\varepsilon < \frac{1}{2}\,\frac{3}{32(1-\gamma)}\,. \tag{15}$$

Indeed, from the expression for f' and Eq. (12),

$$f'(p_0) = \left(\frac{3}{4}\right)^2 \frac{\gamma}{(1-\gamma)^2}\,,$$

and thus

$$\begin{aligned} p_0' - 1 = p_0 + 2\varepsilon/f'(p_0) - 1 &= \frac{4}{3} - \frac{1}{3\gamma} + \left(\frac{4}{3}\right)^2 \frac{2\varepsilon}{\gamma}(1-\gamma)^2 - 1 = \frac{4\gamma - 1 + \frac{32}{3}\varepsilon(1-\gamma)^2 - 3\gamma}{3\gamma} \\ &= \frac{1}{3\gamma}\left[\frac{32\varepsilon}{3}(1-\gamma)^2 - (1-\gamma)\right] \\ &= \frac{(1-\gamma)}{3\gamma}\left[\frac{32\varepsilon}{3}(1-\gamma) - 1\right] \\ &\leq -\frac{(1-\gamma)}{6\gamma} < 0\,, \end{aligned} \tag{16}$$

provided that Eq. (15) holds.

Since $p_1 \geq p_0 \geq 1/2$, $p^* = p_1$. Putting things together,

$$\begin{aligned} \log\left(\frac{1}{4\delta}\right) &\leq n \inf\left\{\frac{(p_0 - p_1)^2}{2p_1(1-p_1)} \,:\, p_1 \in (0,1) \text{ s.t. } p_1 - p_0 > 2\varepsilon/f'(p_0)\right\} \\ &= n\frac{\left(\frac{4}{3}\right)^4 \frac{4\varepsilon^2}{\gamma^2}(1-\gamma)^4}{2p_0'(1-p_0')} && \text{(assuming Eq. (15) so that Eq. (14) holds)} \\ &\leq n\frac{\left(\frac{4}{3}\right)^4 \frac{4\varepsilon^2}{\gamma^2}(1-\gamma)^4 6\gamma}{2p_0(1-\gamma)} && \text{(by Eq. (16) and } p_0' \geq p_0\text{)} \\ &\leq n\,6\left(\frac{4}{3}\right)^4 \frac{4\varepsilon^2}{\gamma_0}(1-\gamma)^3\,. && \text{(since } p_0 \geq 1/2 \text{ and } \gamma \geq \gamma_0\text{)} \end{aligned}$$

Reordering gives the result. □

Total for all questions: 170. Of this, up to 70 can be bonus marks. You can receive bonus marks by asking/upvoting questions, for a total of 70 bonus marks! You must ask at least one question in one of the Lecture Discussion Threads by the Assignment 4 deadline to receive 50 bonus marks. You can also receive 5 bonus marks for upvoting at least one question before 8am on the day of each lecture, for a maximum of 5 marks x 4 lectures = 20 marks for upvoting. Your assignment will be marked out of 170 minus the bonus marks you received.