Problem 1

(a) We have n i.i.d. samples $X_1, \dots, X_n$ sampled from $\texttt{Uni}(\theta, \theta+1)$. We first compute the CDF of $X_{(1)}$, the first-order statistic. For $\theta \le x \le \theta + 1$,

$$\begin{aligned} F_{X_{(1)}}(x) &= P(X_{(1)} \le x) \\ &= 1 - P(X_{(1)} > x) \\ &= 1 - \prod_{i=1}^{n} P(X_i > x) \\ &= 1 - (\theta + 1 - x)^n \\ \implies p_{X_{(1)}}(x) = \frac{\mathrm{d}F_{X_{(1)}}(x)}{\mathrm{d}x} &= n\,(\theta + 1 - x)^{n-1} \end{aligned}$$

Now, we compute the required expectation.

$$\begin{aligned} \mathbb{E}[(X_{(1)} - \theta)^2] &= \int_{\theta}^{\theta+1} (x-\theta)^2 p_{X_{(1)}}(x) dx \\ &= \int_{\theta}^{\theta+1} n(x-\theta)^2 \,(\theta+1-x)^{n-1}\, dx \end{aligned}$$

Let $t = \theta + 1 - x$, then:

$$\begin{aligned} \mathbb{E}[(X_{(1)} - \theta)^2] &= \int_{\theta}^{\theta+1} n(x-\theta)^2 \,(\theta+1-x)^{n-1}\, dx \\ &= \int_0^1 n(1-t)^2 t^{n-1} dt \\ &= n\int_0^1 (t^{n-1} + t^{n+1} - 2t^n) dt \\ &= \frac{2}{(n+1)(n+2)} \end{aligned}$$

(b) We propose three lemmas that would be helpful for proving the lower bound.

Lemma 1. *For distributions P and Q,*

$$\|P - Q\|_{TV} = \int_{\mathcal{B}} |p(x) - q(x)| dx = \frac{1}{2}\int_{\mathcal{X}} |p(x) - q(x)| dx \tag{1}$$

where $\mathcal{B} = \{x \in \mathcal{X} : p(x) \ge q(x)\}$.

Proof.

$$
\begin{aligned}
\|P-Q\|_{\mathrm{TV}} &= \sup_{\mathcal{A}\subset\mathcal{X}} |P(\mathcal{A})-Q(\mathcal{A})| \\
&= \sup_{\mathcal{A}\subset\mathcal{X}} \left|\int_{\mathcal{A}} p(x)-q(x)dx\right| \\
&= \frac{1}{2}\sup_{\mathcal{A}\subset\mathcal{X}} \left(\left|\int_{\mathcal{A}} p(x)-q(x)dx\right| + \left|\int_{\mathcal{X}\setminus\mathcal{A}} p(x)-q(x)dx\right|\right) \quad \text{As both terms are equal} \\
&\le \frac{1}{2}\sup_{\mathcal{A}\subset\mathcal{X}} \left(\int_{\mathcal{A}} |p(x)-q(x)|\,dx + \int_{\mathcal{X}\setminus\mathcal{A}} |p(x)-q(x)|\,dx\right) \\
&= \frac{1}{2}\sup_{\mathcal{A}\subset\mathcal{X}} \int_{\mathcal{X}} |p(x)-q(x)|\,dx \\
&= \frac{1}{2}\int_{\mathcal{X}} |p(x)-q(x)|\,dx \qquad (2) \\
&= \frac{1}{2}\int_{\mathcal{B}} p(x)-q(x)dx + \frac{1}{2}\int_{\mathcal{X}\setminus\mathcal{B}} q(x)-p(x)dx \\
&\le \frac{1}{2}\left|\int_{\mathcal{B}} p(x)-q(x)dx\right| + \frac{1}{2}\left|\int_{\mathcal{X}\setminus\mathcal{B}} p(x)-q(x)dx\right| \\
&\le \frac{1}{2}\cdot 2\sup_{\mathcal{A}\subset\mathcal{X}} \left|\int_{\mathcal{A}} p(x)-q(x)dx\right| \\
&= \|P-Q\|_{\mathrm{TV}} \qquad (3)
\end{aligned}
$$

Furthermore, note that:

$$
\begin{aligned}
0 = \int_{\mathcal{X}} p(x)-q(x)dx &= \int_{\mathcal{B}} p(x)-q(x)dx + \int_{\mathcal{X}\setminus\mathcal{B}} p(x)-q(x)dx \\
&= \int_{\mathcal{B}} |p(x)-q(x)|dx - \int_{\mathcal{X}\setminus\mathcal{B}} |p(x)-q(x)|dx \\
\implies \int_{\mathcal{B}} |p(x)-q(x)|dx &= \int_{\mathcal{X}\setminus\mathcal{B}} |p(x)-q(x)|dx = \frac{1}{2}\int_{\mathcal{X}} |p(x)-q(x)|dx \qquad (4)
\end{aligned}
$$

Using (2),(3) and (4), we conclude:

$$\|P-Q\|_{\mathrm{TV}} = \int_{\mathcal{B}} |p(x)-q(x)|dx = \frac{1}{2}\int_{\mathcal{X}} |p(x)-q(x)|dx \qquad (5)$$

□

Lemma 2. *For any distributions P and Q, the Hellinger distance is defined as $d_{hel}(P,Q)^2 = \int_{\mathcal{X}}(\sqrt{p(x)} - \sqrt{q(x)})^2 dx$. Then, we have:*

$$\|P-Q\|_{TV} \le d_{hel}(P,Q) \qquad (6)$$

Proof.

$$
\begin{aligned}
\|P-Q\|_{\mathrm{TV}} &= \frac{1}{2}\int_{\mathcal{X}} |p(x)-q(x)|dx && \text{Lemma 1}\\
&= \frac{1}{2}\int_{\mathcal{X}} |\sqrt{p(x)}-\sqrt{q(x)}|\cdot|\sqrt{p(x)}+\sqrt{q(x)}|dx\\
&\le \frac{1}{2}\sqrt{\int_{\mathcal{X}} |\sqrt{p(x)}-\sqrt{q(x)}|^2dx}\cdot\sqrt{\int_{\mathcal{X}} |\sqrt{p(x)}+\sqrt{q(x)}|^2dx} && \text{Cauchy-Schwarz inequality}\\
&= \frac{1}{2}d_{hel}(P,Q)\sqrt{2+\int_{\mathcal{X}}\sqrt{p(x)q(x)}dx}\\
&= \frac{1}{2}d_{hel}(P,Q)\sqrt{4-\left(2-\int_{\mathcal{X}}\sqrt{p(x)q(x)}dx\right)}\\
&= \frac{1}{2}d_{hel}(P,Q)\sqrt{4-\int_{\mathcal{X}}|\sqrt{p(x)}-\sqrt{q(x)}|^2}\\
&= d_{hel}(P,Q)\sqrt{1-\frac{1}{4}d_{hel}(P,Q)^2}\\
&\le d_{hel}(P,Q)
\end{aligned}
$$

□

Lemma 3. *For product distributions $P^n=\prod_{i=1}^n P_i$ and $Q^n=\prod_{i=1}^n Q_i$, the following result holds true:*

$$
d_{hel}(P^n,Q^n)^2 = 2-2\prod_{i=1}^n\left(1-\frac{1}{2}d_{hel}(P_i,Q_i)^2\right) \le 2-2\prod_{i=1}^n\left(1-d_{hel}(P_i,Q_i)^2\right) \tag{7}
$$

Proof. The proof follows directly from the definition of Hellinger distance and independence of the distributions. □

Let the two probability distributions be of the form $P_v = \mathtt{Uni}(\theta+v\delta,\theta+v\delta+1)$ for some $\delta>0$ and $v\in\{-1,+1\}$. So, $\widehat{\theta}(P_v)=\theta+v\delta$. Clearly, for semi-metric $\rho(x,y)=|x-y|$, we have $\rho(\widehat{\theta}(P_{+1}),\widehat{\theta}(P_{-1}))=2\delta$. So, for $\Phi(t)=t^2$, using Le Cam's method, we obtain:

$$
\begin{aligned}
\mathfrak{M}_n(\mathtt{Uni}(\theta,\theta+1),(\cdot)^2) &\ge \frac{1}{2}\Phi(\delta)\left[1-\left\|P_{+1}^n-P_{-1}^n\right\|_{\mathrm{TV}}\right]\\
&= \frac{1}{2}\delta^2\left[1-\left\|P_{+1}^n-P_{-1}^n\right\|_{\mathrm{TV}}\right]\\
&\ge \frac{1}{2}\delta^2\left[1-d_{hel}(P_{+1}^n,P_{-1}^n)\right] && \text{Lemma 2}\\
&\ge \frac{1}{2}\delta^2\left[1-\sqrt{2-2\left(1-d_{hel}(P_{+1},P_{-1})^2\right)^n}\right] && \text{Lemma 3}\\
&\ge \frac{1}{2}\delta^2\left[1-\sqrt{2-2\left(1-4\delta\right)^n}\right] && \text{Def. of Hellinger dist.}
\end{aligned}
$$

Let $\delta = 1/32n$,

$$\begin{aligned}
\mathfrak{M}_n(\texttt{Uni}(\theta, \theta+1), (\cdot)^2) &\geq \frac{1}{2}\delta^2 \left[1 - \sqrt{2 - 2\left(1 - 4\delta\right)^n}\right] \\
&= \frac{1}{2048n^2}\left[1 - \sqrt{2 - 2\left(1 - \frac{1}{8n}\right)^n}\right] \\
&\geq \frac{1}{2048n^2}\left[1 - \sqrt{2 - 2\left(1 - \frac{1}{8}\right)}\right] \qquad (1 + x/n)^n \geq 1 + x, \text{for } |x| \leq n \\
&= \frac{1}{4096n^2}
\end{aligned}$$

The numerical constant may be further improved by choosing a δ that maximizes the RHS. In part (a), we had a definite estimator of θ, namely the first-order statistic. As $n \to \infty$, we can see that the lower bound of error in both the cases declines as $\mathcal{O}\left(\frac{1}{n^2}\right)$.

■

Problem 2

(a) Without loss of generality, assume $a \geq b > 0$.

$$\left|\ln \frac{a}{b}\right| = \ln \frac{a}{b} \overset{(i)}{\leq} \frac{a}{b} - 1 = \frac{a-b}{b} = \frac{|a-b|}{\min\{a,b\}} \tag{8}$$

where (i) holds because $\ln(1+x) \leq x$.

(b) Let $\mathcal{B} = \{x \in \mathcal{X} : p_1(x) \geq p_2(x)\}$.

$$\begin{aligned}
m_1(z) - m_2(z) &= \int_{\mathcal{X}} q(z|x)[p_1(x) - p_2(x)]dx \\
&= \int_{\mathcal{B}} q(z|x)[p_1(x) - p_2(x)]dx + \int_{\mathcal{X}\setminus\mathcal{B}} q(z|x)[p_1(x) - p_2(x)]dx \\
&= \int_{\mathcal{B}} q(z|x)|p_1(x) - p_2(x)|dx - \int_{\mathcal{X}\setminus\mathcal{B}} q(z|x)|p_1(x) - p_2(x)|dx \\
&\leq \sup_{x\in\mathcal{X}} q(z|x) \int_{\mathcal{B}} |p_1(x) - p_2(x)|dx - \inf_{x\in\mathcal{X}} q(z|x) \int_{\mathcal{X}\setminus\mathcal{B}} |p_1(x) - p_2(x)|dx \\
&\overset{(4)}{=} \left(\sup_{x\in\mathcal{X}} q(z|x) - \inf_{x\in\mathcal{X}} q(z|x)\right) \int_{\mathcal{B}} |p_1(x) - p_2(x)|dx \\
&= \left(\sup_{x\in\mathcal{X}} q(z|x) - \inf_{x\in\mathcal{X}} q(z|x)\right) \|P_1 - P_2\|_{\text{TV}} \qquad \text{Lemma 1}
\end{aligned} \tag{9}$$

Similarly,

$$\begin{aligned}
m_1(z) - m_2(z) &= \int_{\mathcal{B}} q(z|x)|p_1(x) - p_2(x)|dx - \int_{\mathcal{X}\setminus\mathcal{B}} q(z|x)|p_1(x) - p_2(x)|dx \\
&\geq \inf_{x\in\mathcal{X}} q(z|x) \int_{\mathcal{B}} |p_1(x) - p_2(x)|dx - \sup_{x\in\mathcal{X}} q(z|x) \int_{\mathcal{X}\setminus\mathcal{B}} |p_1(x) - p_2(x)|dx \\
&= \left(\inf_{x\in\mathcal{X}} q(z|x) - \sup_{x\in\mathcal{X}} q(z|x)\right) \|P_1 - P_2\|_{\text{TV}} \qquad \text{Lemma 1}
\end{aligned} \tag{10}$$

From (9) and (10), we get:

$$\begin{aligned}
\left(\inf_{x\in\mathcal{X}} q(z|x) - \sup_{x\in\mathcal{X}} q(z|x)\right) \|P_1 - P_2\|_{\text{TV}} &\overset{(10)}{\leq} m_1(z) - m_2(z) \\
&\overset{(9)}{\leq} \left(\sup_{x\in\mathcal{X}} q(z|x) - \inf_{x\in\mathcal{X}} q(z|x)\right) \|P_1 - P_2\|_{\text{TV}} \\
\implies |m_1(z) - m_2(z)| &\leq \left|\left(\sup_{x\in\mathcal{X}} q(z|x) - \inf_{x\in\mathcal{X}} q(z|x)\right) \|P_1 - P_2\|_{\text{TV}}\right| \\
&= \sup_{x,x'} |q(z|x) - q(z|x')| \, \|P_1 - P_2\|_{\text{TV}}
\end{aligned} \tag{11}$$

Now, we upper bound the RHS of (11)[1].

$$\begin{aligned}
\sup_{x,x'} |q(z|x) - q(z|x')| &= \inf_{\hat{x}} \sup_{x,x'} |q(z|x) - q(z|\hat{x}) + q(z|\hat{x}) - q(z|x')| \\
&\leq 2 \inf_{\hat{x}} \sup_{x} |q(z|x) - q(z|\hat{x})| \\
&= 2 \inf_{\hat{x}} q(z|\hat{x}) \sup_{x} \left| \frac{q(z|x)}{q(z|\hat{x})} - 1 \right| \\
&= 2 \inf_{\hat{x}} q(z|\hat{x}) \max\{e^{\alpha} - 1, 1 - e^{-\alpha}\} && \frac{q(z|x)}{q(z|\hat{x})} \in [e^{-\alpha}, e^{\alpha}] \\
&= 2 \inf_{\hat{x}} q(z|\hat{x})(e^{\alpha} - 1) && e^{\alpha} - 1 \geq 1 - e^{-\alpha}\}
\end{aligned}$$

So, combining the above upper bound with (11),for some $c > 0$, we get,

$$|m_1(z) - m_2(z)| \leq c(e^{\alpha} - 1) \inf_{x \in \mathcal{X}} q(z|x) \|P_1 - P_2\|_{\mathrm{TV}} \tag{12}$$

(c) Now, we prove the required inequality.

$$\begin{aligned}
D(M_1||M_2) + D(M_2||M_1) &= \int m_1(z) \log \frac{m_1(z)}{m_2(z)} + m_2(z) \log \frac{m_2(z)}{m_1(z)} dz \\
&= \int (m_1(z) - m_2(z)) \log \frac{m_1(z)}{m_2(z)} dz \\
&\leq \int |m_1(z) - m_2(z)| \left| \log \frac{m_1(z)}{m_2(z)} \right| dz \\
&\overset{(8)}{\leq} \int |m_1(z) - m_2(z)| \frac{|m_1(z) - m_2(z)|}{\min\{m_1(z), m_2(z)\}} dz \\
&= \int \frac{|m_1(z) - m_2(z)|^2}{\min\{m_1(z), m_2(z)\}} dz \\
&\overset{(12)}{\leq} \int \frac{c^2(e^{\alpha} - 1)^2 [\inf_{x \in \mathcal{X}} q(z|x)]^2 \|P_1 - P_2\|_{\mathrm{TV}}^2}{\min\{m_1(z), m_2(z)\}} dz \\
&\overset{(i)}{\leq} c^2(e^{\alpha} - 1)^2 \|P_1 - P_2\|_{\mathrm{TV}}^2 \int \frac{[\inf_{x \in \mathcal{X}} q(z|x)]^2}{\inf_{x \in \mathcal{X}} q(z|x)} dz \\
&= c^2(e^{\alpha} - 1)^2 \|P_1 - P_2\|_{\mathrm{TV}}^2 \int \inf_{x \in \mathcal{X}} q(z|x) dz \\
&\leq c^2(e^{\alpha} - 1)^2 \|P_1 - P_2\|_{\mathrm{TV}}^2 \int q(z|x) dz \\
&= C(e^{\alpha} - 1)^2 \|P_1 - P_2\|_{\mathrm{TV}}^2
\end{aligned} \tag{13}$$

where (i) follows because $m_i(z) = \int q(z|x) p_i(x) dx \geq \inf_{x \in \mathcal{X}} q(z|x) \int p_i(x) dx = \inf_{x \in \mathcal{X}} q(z|x)$. So, $\min\{m_1(z), m_2(z)\} \geq \inf_{x \in \mathcal{X}} q(z|x)$.

■

[1] This proof follows a similar proof in Duchi, J. C., Jordan, M. I., & Wainwright, M. J. (2018). Minimax optimal procedures for locally private estimation. Journal of the American Statistical Association, 113(521), 182-201. In that work, the authors prove more rigorously that $c = \min\{2, e^{\alpha}\}$.

Problem 3

(a) We have:

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

Let the two probability distributions be of the form $P_v = \texttt{Bernoulli}(\frac{1}{2} + v\delta)$ for some $\delta > 0$ and $v \in \{-1, +1\}$. So, $\widehat{\theta}(P_v) = \frac{1}{2} + v\delta$. Clearly, for semi-metric $\rho(x, y) = |x - y|$, we have $\rho(\widehat{\theta}(P_{+1}), \widehat{\theta}(P_{-1})) = 2\delta$, i.e, the distributions are 2δ-separated. Let M_1^n and M_2^n be the marginal distributions obtained as follows:

$$M_i^n(Z) = \int_{\mathcal{X}^n} Q^n(Z|x_1, \ldots, x_n) P_i^n(x) dx$$

for $i \in \{-1, +1\}$. Without loss of generality, assume $D(M_{+1}||D_{-1}) \leq D(M_{+1}||D_{-1})$. So, for $\Phi(t) = |t|$, using Le Cam's method for α-differentially private case, we obtain:

$$\begin{aligned}
\mathfrak{M}_n(\texttt{Bernoulli}(\theta), |\cdot|, \alpha) &\geq \frac{1}{2}\delta \left[1 - \left\|M_{+1}^n - M_{-1}^n\right\|_{\text{TV}}\right] \\
&\overset{(i)}{\geq} \frac{1}{2}\delta \left[1 - \sqrt{\frac{D(M_{+1}^n||M_{-1}^n)}{2}}\right] \\
&\overset{(ii)}{\geq} \frac{1}{2}\delta \left[1 - \sqrt{\frac{nD(M_{+1}||M_{-1})}{2}}\right] \\
&\overset{(iii)}{\geq} \frac{1}{2}\delta \left[1 - \sqrt{\frac{n[D(M_{+1}||M_{-1}) + D(M_{-1}||M_{+1})]}{4}}\right] \\
&\overset{(13)}{\geq} \frac{1}{2}\delta \left[1 - \sqrt{\frac{nC(e^\alpha - 1)^2\|P_{+1} - P_{-1}\|_{\text{TV}}^2}{4}}\right] \\
&\overset{(iv)}{\geq} \frac{1}{2}\delta \left[1 - \sqrt{\frac{nC(2\alpha^2)(4\delta^2)}{4}}\right] \\
&= \frac{1}{2}\delta \left[1 - \sqrt{2nC\alpha^2\delta^2}\right]
\end{aligned}$$

Let $\delta = \sqrt{\frac{1}{8nC\alpha^2}}$, we get:

$$\mathfrak{M}_n(\texttt{Bernoulli}(\theta), |\cdot|, \alpha) \geq \frac{1}{2}\delta \left[1 - \sqrt{2nC\alpha^2\delta^2}\right] = \frac{1}{4}\sqrt{\frac{1}{8nC\alpha^2}} = \frac{c}{\sqrt{n\alpha^2}}$$

for some $c > 0$. The inequalities (i) follows from the Pinsker's inequality,
(ii) follows from the tensorization property of KL divergence,
(iii) follows because $2D(M_{+1}||M_{-1}) \leq D(M_{+1}||M_{-1}) + D(M_{-1}||M_{+1})$ by assumption and,
(iv) holds because $(e^\alpha - 1)^2 \leq 2\alpha^2$ by assumption and total variation distance between $\texttt{Bernoulli}(p)$ and $\texttt{Bernoulli}(q)$ is $|p - q| \implies \|P_{+1} - P_{-1}\|_{\text{TV}} = 2\delta$.

(b) We know that $X_i \overset{\text{i.i.d.}}{\sim}$ `Bernoulli`(θ). We define Z_i as follows[2]:

$$Z_i = \begin{cases} X_i & \text{with probability } \frac{1}{2}+\gamma \\ 1-X_i & \text{with probability } \frac{1}{2}-\gamma \end{cases}$$

for some $\gamma \in (0, \frac{1}{2})$. We now calculate the conditional expectation of Z_i given X_i.

$$\begin{aligned} \mathbb{E}[Z_i|X_i] = X_i\left(\frac{1}{2}+\gamma\right) + (1-X_i)\left(\frac{1}{2}-\gamma\right) &= 2\gamma X_i - \gamma + \frac{1}{2} \\ \implies \mathbb{E}\left[\frac{1}{2\gamma}\left(Z_i+\gamma-\frac{1}{2}\right)|X_i\right] &= X_i \\ \implies \mathbb{E}\left[\mathbb{E}\left[\frac{1}{2\gamma}\left(Z_i+\gamma-\frac{1}{2}\right)|X_i\right]\right] &= \mathbb{E}[X_i] = \theta \end{aligned}$$

So, the proposed estimator is:

$$\widehat{\theta}(Z^n) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{2\gamma}\left(Z_i+\gamma-\frac{1}{2}\right) \tag{14}$$

We now prove that $\mathbb{E}[|\widehat{\theta}(Z^n)-\theta|] \leq \frac{C}{\sqrt{n\alpha^2}}$.

$$\mathrm{Var}\left[\widehat{\theta}(Z^n)\right] = \mathrm{Var}\left[\frac{1}{n}\sum_{i=1}^{n}\frac{1}{2\gamma}\left(Z_i+\gamma-\frac{1}{2}\right)\right] = \mathrm{Var}\left[\frac{1}{2n\gamma}\sum_{i=1}^{n}Z_i\right] = n\cdot\frac{1}{4n^2\gamma^2}\mathrm{Var}[Z_i] \leq \frac{1}{16\gamma^2 n} \tag{15}$$

where the last inequality follows because variance of a Bernoulli r.v. is upper bounded by 1/4. Applying Chebyshev's inequality for $\widehat{\theta}(Z^n)$, we get:

$$P(|\widehat{\theta}(Z^n)-\mathbb{E}[\widehat{\theta}(Z^n)]| > t) \leq \frac{\mathrm{Var}[\widehat{\theta}(Z^n)]}{t^2} \implies P(|\widehat{\theta}(Z^n)-\theta| > t) \overset{(15)}{\leq} \frac{1}{16n\gamma^2t^2}$$

Let $\delta = \frac{1}{16n\gamma^2t^2} \implies t = \frac{1}{4\gamma\sqrt{n\delta}}$, we obtain

$$P\left(|\widehat{\theta}(Z^n)-\theta| > \frac{1}{4\gamma\sqrt{n\delta}}\right) \leq \delta \implies P\left(|\widehat{\theta}(Z^n)-\theta| \leq \frac{1}{4\gamma\sqrt{n\delta}}\right) > 1-\delta$$

So w.h.p,

$$|\widehat{\theta}(Z^n)-\theta| \leq \frac{1}{4\gamma\sqrt{n\delta}} \implies \mathbb{E}[|\widehat{\theta}(Z^n)-\theta|] \leq \mathcal{O}\left(\frac{1}{\gamma\sqrt{n}}\right) \tag{16}$$

Now, we need to show the proposed channel is α-differentially private. Let $\alpha = \ln\frac{1+2\gamma}{1-2\gamma} \geq 0$ as $\gamma > 0$. We have assumed $\alpha \leq \frac{1}{2} \implies \gamma \leq \frac{1}{8}$. By definition,

$$\frac{Q(Z=z|X=x)}{Q(Z=z|X=x')} \leq \max_{x,x'}\frac{Q(Z=z|X=x)}{Q(Z=z|X=x')} = \frac{1/2+\gamma}{1/2-\gamma} = e^{\alpha} \lesssim e^{\gamma} \tag{17}$$

where $\lesssim$ means that $e^{\alpha} \leq Ke^{\gamma}$ for some universal constant K. So, from (16) and (17), we get:

$$\mathbb{E}[|\widehat{\theta}(Z^n)-\theta|] \leq \mathcal{O}\left(\frac{1}{\gamma\sqrt{n}}\right) = \mathcal{O}\left(\frac{1}{\alpha\sqrt{n}}\right) = \frac{C}{\sqrt{n\alpha^2}}$$

[2] http://www.gautamkamath.com/CS860notes/lec3.pdf

(c) We use the estimator defined in (14) and conduct the required experiments. We sample 70% of the data each time, pass through the α-differentially private channel for different α. Each experiment is repeated 100 times and the mean and standard deviation is reported in Table 1. The estimated value of θ and errors is plotted in Figure 1 and 2 respectively.

Table 1: Variation of $\widehat{\theta}(Z^n)$ and the estimation error with α. Here, $\theta = 0.29956$.

α	$\widehat{\theta}(Z^n)$	$\lvert\widehat{\theta}(Z^n) - \theta\rvert$
0.50000	0.30027 ± 0.00840	0.00682 ± 0.00467
0.25000	0.29852 ± 0.01790	0.01439 ± 0.01048
0.12500	0.30070 ± 0.02760	0.02263 ± 0.01630
0.06250	0.29454 ± 0.06492	0.05269 ± 0.03838
0.03125	0.28649 ± 0.13722	0.11233 ± 0.08004
0.01562	0.31393 ± 0.24966	0.20521 ± 0.14300
0.00781	0.26055 ± 0.52886	0.42192 ± 0.32101
0.00391	0.26004 ± 1.14442	0.95893 ± 0.62580
0.00195	0.49795 ± 2.13065	1.73291 ± 1.25526
0.00098	-0.07685 ± 4.90111	4.14733 ± 2.63868

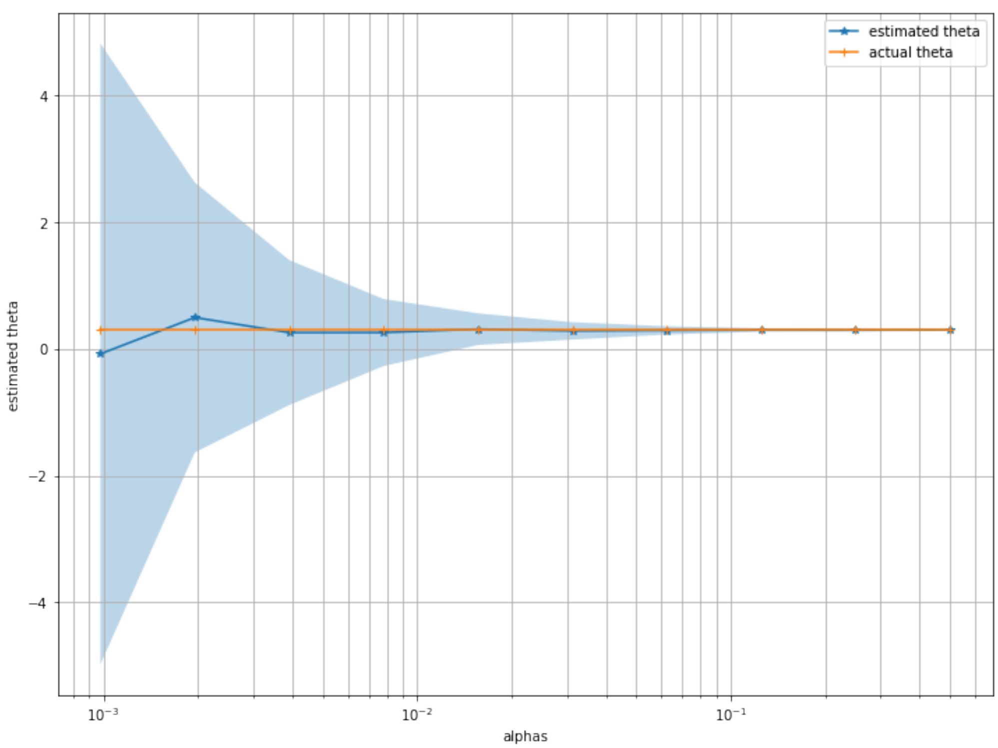

Figure 1: Estimated and actual value of θ vs α. The shaded area shows the uncertainty in estimation.

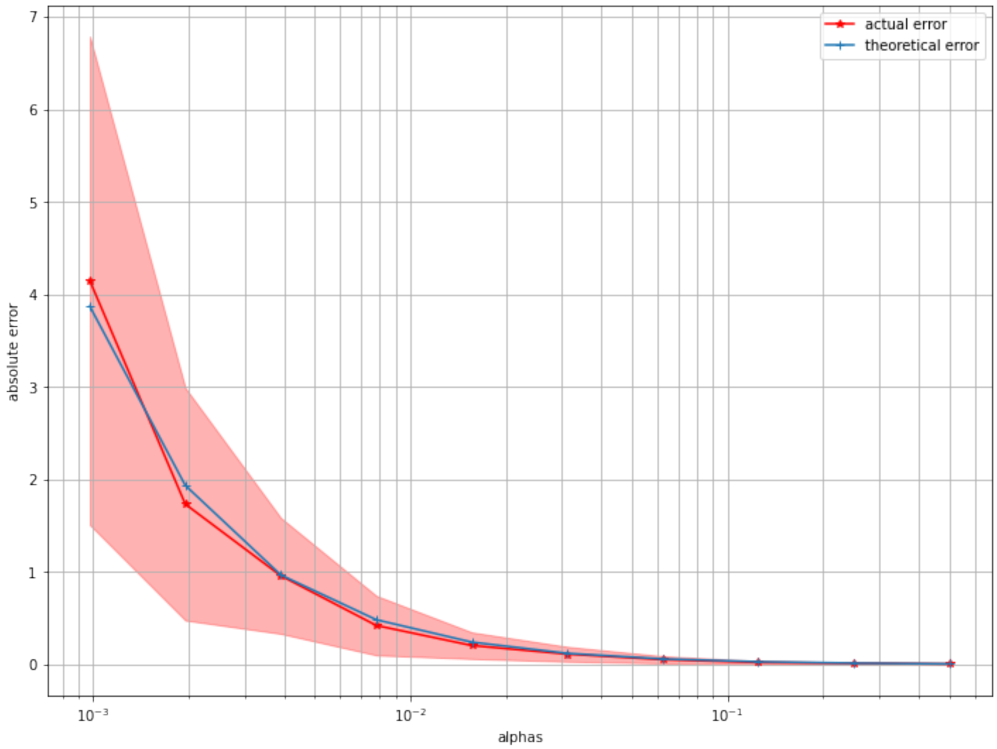

Figure 2: Actual and theoretical error vs α. The shaded area shows the uncertainty in error. Here theoretical error is taken to be $1/\sqrt{n\alpha^2}$ where $n = 70000$.

$\blacksquare$

Problem 4

(a) For a fixed matrix $X \in \mathbb{R}^{n\times d}$, we have:

$$Y = X\theta^s + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \sigma^2 I_{n\times n})$$

and we need estimate $\theta^s = \theta_{\min} S$ where $\theta_{\min} > 0$ and $S \in \mathcal{S}_k = \{s | s \in \{-1, 0, +1\}^d, \|s\|_1 = k\}$. Note that, for estimating θ^s, it is sufficient to find $\widehat{S}$, the estimate of S. Furthermore, note that $S \to Y \to \widehat{S}$ forms a Markov Chain. So, applying Fano's inequality, we obtain:

$$P(S \neq \widehat{S}) \geq 1 - \frac{I(S,Y) + \ln 2}{\ln |\mathcal{S}_k|}$$

We need to upper bound the second term on RHS to lower bound the LHS. Precisely, if we can show:

$$\frac{I(S,Y) + \ln 2}{\ln |\mathcal{S}_k|} \leq \frac{1}{2}$$

then, we can ensure that $P(S \neq \widehat{S}) \geq 1/2$. For any $S \in \mathcal{S}_k$, there can be exactly k non-zero coordinates and each non-zero coordinate can take two values from $\{-1, +1\}$. So, $|\mathcal{S}_k| = 2^k \binom{d}{k}$.

Proposition 4. *If S and S' are sampled independently and uniformly from $\mathcal{S}_k$, we have:*

$$I(S,Y) \leq \mathbb{E}_{S,S'}\left[\mathbb{E}_{Y|S}\left[\ln \frac{p(y|s)}{p(y|s')}\right]\right] = \mathbb{E}_{S,S'}\left[D(P_s || P_{s'})\right]$$

where P_s is the conditional distribution of Y given $S = s$.

Proof.

$$\begin{aligned}
I(S,Y) &= \mathbb{E}_{S,Y}\left[\ln \frac{p(s,y)}{p(s)p(y)}\right] \\
&= \mathbb{E}_{S,Y}\left[\ln \frac{p(y|s)}{p(y)}\right] \\
&= \mathbb{E}_{S}\left[\mathbb{E}_{Y|S}\left[\ln \frac{p(y|s)}{p(y)}\right]\right] \\
&= \mathbb{E}_{S}\left[\mathbb{E}_{Y|S}\left[\ln \frac{p(y|s)}{\mathbb{E}_{S'}[p(y|s')]}\right]\right] \\
&= \mathbb{E}_{S}\left[\mathbb{E}_{Y|S}[\ln p(y|s)] - \ln[\mathbb{E}_{S'}[p(y|s')]]\right] \\
&\overset{(i)}{\leq} \mathbb{E}_{S}\left[\mathbb{E}_{Y|S}[\ln p(y|s)] - \mathbb{E}_{S'}[\ln[p(y|s')]]\right] \\
&\overset{(ii)}{=} \mathbb{E}_{S}\left[\mathbb{E}_{S'}[\mathbb{E}_{Y|S}[\ln p(y|s)]] - \mathbb{E}_{S'}[\ln[p(y|s')]]\right] \\
&= \mathbb{E}_{S}\left[\mathbb{E}_{S'}\left[\mathbb{E}_{Y|S}\left[\ln \frac{p(y|s)}{p(y|s')}\right]\right]\right] \\
&= \mathbb{E}_{S,S'}\left[\mathbb{E}_{Y|S}\left[\ln \frac{p(y|s)}{p(y|s')}\right]\right] \\
&= \mathbb{E}_{S,S'}\left[D(P_s || P_{s'})\right]
\end{aligned}$$

where (i) holds due Jensen's inequality and (ii) holds because S and S' are independent.

$\square$

So, from Proposition 4, we get:

$$\begin{aligned}
I(S,Y) &\leq \mathbb{E}_{S,S'}\left[D(\mathcal{N}(X\theta^s,\sigma^2 I)||\mathcal{N}(X\theta^{s'},\sigma^2 I))\right] \\
&\overset{(i)}{=} \frac{\theta_{\min}^2}{2\sigma^2}\mathbb{E}_{S,S'}\left[\left\|X(s-s')\right\|_2^2\right] \\
&= \frac{\theta_{\min}^2}{2\sigma^2}\mathbb{E}_{S,S'}\left[\|Xs\|_2^2+\left\|Xs'\right\|_2^2-2\langle Xs,Xs'\rangle\right] \\
&= \frac{\theta_{\min}^2}{2\sigma^2}\mathbb{E}_{S,S'}\left[\|Xs\|_2^2+\left\|Xs'\right\|_2^2\right]-2\mathbb{E}_{S,S'}\left[\langle Xs,Xs'\rangle\right] \\
&\overset{(ii)}{=} \frac{\theta_{\min}^2}{2\sigma^2}\cdot 2\mathbb{E}_S\left[\|Xs\|_2^2\right] \\
&\overset{(iii)}{=} \frac{\theta_{\min}^2}{\sigma^2}\cdot\frac{kn}{d}\cdot\left\|n^{-1/2}X\right\|_{\texttt{Fr}}^2
\end{aligned} \tag{18}$$

The statement (i) holds because the KL divergence between $\mathcal{N}(\theta_1,\sigma^2 I)$ and $\mathcal{N}(\theta_2,\sigma^2 I)$ is $\frac{1}{2\sigma^2}\|\theta_1-\theta_2\|_2^2$,
(ii) holds because $\mathbb{E}_{S,S'}[\|Xs\|_2^2]=\mathbb{E}_S[\|Xs\|_2^2]$ as S and S' are independent. Similarly, $\mathbb{E}_{S,S'}[\|Xs'\|_2^2]=\mathbb{E}_{S'}[\|Xs'\|_2^2]=\mathbb{E}_S[\|Xs\|_2^2]$. Further $\mathbb{E}_{S,S'}\left[\langle Xs,Xs'\rangle\right]=0$ as $\mathbb{E}_S[s]=0$. This can be established because S is uniformly sampled and for every $s\in\mathcal{S}_k,\exists -s\in\mathcal{S}_k$,
(iii) holds because:

$$\mathbb{E}_S[\|Xs\|_2^2]=\mathbb{E}_S\left[\sum_j |s_jX_j|^2\right]=\sum_j |X_j|^2\mathbb{E}_S\left[s_j^2\right]=\frac{k}{d}\sum_j |X_j|^2=\frac{k}{d}\|X\|_{\texttt{Fr}}^2=\frac{kn}{d}\left\|n^{-\frac{1}{2}}X\right\|_{\texttt{Fr}}^2$$

where X_j are the columns of X. Suppose $n\leq c\frac{(d/k)\ln\binom{d}{k}\sigma^2}{\left\|n^{-1/2}X\right\|_{\texttt{Fr}}^2\theta_{\min}^2}$, for some $c>0$,

$$\begin{aligned}
\frac{I(S,Y)+\ln 2}{\ln|\mathcal{S}_k|} &\overset{(18)}{\leq} \frac{\frac{\theta_{\min}^2}{\sigma^2}\frac{kn}{d}\left\|n^{-1/2}X\right\|_{\texttt{Fr}}^2+\ln 2}{\ln 2^k\binom{d}{k}} \\
&\leq \frac{c\ln\binom{d}{k}+\ln 2}{\ln\binom{d}{k}+k\ln 2}\leq\frac{1}{2}
\end{aligned}$$

where the last inequality can be satisfied if c,d and k are chosen appropriately. Now, as $\frac{I(S,Y)+\ln 2}{\ln|\mathcal{S}_k|}\leq\frac{1}{2}\implies P(S\neq\widehat{S})\geq\frac{1}{2}$.

(b) Here $X\in\{-1,+1\}^{n\times d}\implies\left\|n^{-1/2}X\right\|_{\texttt{Fr}}=d$. Suppose, we want $1-\delta$ level of confidence, i.e, $P(S \neq \widehat{S}) \leq \delta$, then, we need to ensure $\frac{I(S,Y)+\ln 2}{\ln|\mathcal{S}_k|} \geq 1-\delta$. So,

$$
\begin{aligned}
1-\delta \leq \frac{I(S,Y)+\ln 2}{\ln|\mathcal{S}_k|} &\overset{(18)}{\leq} \frac{\frac{\theta_{\min}^2}{\sigma^2}\frac{kn}{d}\left\|n^{-1/2}X\right\|_{\mathtt{Fr}}^2+\ln 2}{\ln 2^k\binom{d}{k}} \\
\implies n &\geq \frac{(1-\delta)\ln\left(2^k\binom{d}{k}\right)-\ln 2}{\frac{\theta_{\min}^2}{\sigma^2}\frac{k}{d}\left\|n^{-1/2}X\right\|_{\mathtt{Fr}}^2} \\
&\geq ((1-\delta)(k\ln 2+k\ln(d/k))-\ln 2)\cdot\frac{\sigma^2}{k\theta_{\min}^2} \\
&= \mathcal{O}\left(\frac{\sigma^2}{\theta_{\min}^2}\ln\frac{d}{k}\right)
\end{aligned}
$$

The term σ^2 corresponds to noise and $\theta_{\min}$ refers to the signal strength. So, $\theta_{\min}^2/\sigma^2$ refers to the signal-to-noise (SNR) value.

■

Problem 5

(a) We show that if the polynomial $p(x)$ changes sign in $[a-\epsilon, a+\epsilon], \forall \epsilon > 0$, then $p(x)$ has a root in $[a-\epsilon, a+\epsilon]$. The proof for the same follows directly from the intermediate value theorem which states that if $f(x)$ is a real valued continuous function in the interval $[a, b]$, then $\forall c \in [f(a), f(b)], \exists x \in [a, b]$ such that $f(x) = c$.

In this case, we fix $c = 0 \in [p(a-\epsilon), p(a+\epsilon)] \implies p(x) = 0$ for some $x \in [a-\epsilon, a+\epsilon]$. So, as $p(x)$ can have atmost d real roots, it can have atmost d sign changes.

(b) Suppose we are given any pair of set of real numbers and their corresponding image $\in \{\pm 1\}$ of cardinality $d+1$. We propose to construct $p(x)$ that shatters it, i.e., $p(x)$ satisfies the data. Let $S = \{a_0, \ldots, a_d\}$ such that $a_0 < a_1 < \cdots < a_d$ and $p(S) = \{\text{sign}(p(x)) | x \in S\}$. Let $S_r = \{(a_i, a_{i+1}) | \text{sign}(p(a_{i+1})) \neq \text{sign}(p(a_i)), i+1 \in [d]\}$. From part (a), we can deduce that $|S_r| \leq d$. Let $p(x)$ be defined as follows:

$$p(x) = A \prod_i (x - b_i), \text{ where } b_i = \frac{a_i + a_{i+1}}{2} \text{ for } (a_i, a_{i+1}) \in S_r \tag{19}$$

Clearly, degree of $p(x) \leq d$. Moreover, $p(b_i) = 0$ and by construction, there is *exactly one root* in $[a_i, a_{i+1}]$. So, sign of $p(x)$ is different at $x = a_i$ and $x = a_{i+1}$.

(c) Consider any set $S = \{a_0, \ldots, a_{d+1}\}$ and its corresponding image set $p(S) = \{+, -, +, \ldots\}$. The number of sign changes in $p(S)$ is $d+1$. However, $p(x) \in \mathcal{H}_d$ can have atmost d sign changes as proved in part (a). So, there is no $p(x) \in \mathcal{H}_d$ that shatters S.

From (b), we deduce that $\texttt{VCDim}(\mathcal{H}_d) \geq d$ and from (c), we get $\texttt{VCDim}(\mathcal{H}_d) < d+1$. So, $\texttt{VCDim}(\mathcal{H}_d) = d$.

■

Problem 6

(a) Each Boolean variable x can either be present as $x, \bar{x}$ or absent from the conjunction. So, for d Boolean variables, the total number of Boolean conjunctions is 3^d. So, $|\mathcal{H}^d_{\texttt{con}}| = 3^d \leq 3^d + 1$.

Note that, there is an expression $\phi \in \mathcal{H}^d_{\texttt{con}}$ in which *none* of the variables are present. We trivially assume its Boolean value to be 1.

(b) Suppose the $\texttt{VCDim}(\mathcal{H}^d_{\texttt{con}}) = k$. So, there exists $\mathcal{H} \subseteq \mathcal{H}^d_{\texttt{con}}$ that shatters a set of size of k. As $\mathcal{H} \subseteq \mathcal{H}^d_{\texttt{con}} \implies |\mathcal{H}| \leq |\mathcal{H}^d_{\texttt{con}}| \implies 2^k \leq 3^d \implies k \leq d \log 3$. So,

$$\texttt{VCDim}(\mathcal{H}^d_{\texttt{con}}) \leq d \log 3$$

(c) The cardinality of set of unit vectors is $d \leq d \log 3 \implies$ there could be as subset of $\mathcal{H}^d_{\texttt{con}}$ that shatters the set of unit vectors.

Proposition 5. *Let E be the set of unit vectors. Let $h(E)$ be its corresponding image set. We partition $[d]$ into E_0 and E_1 such that for $i \in \{0, 1\}$, we have:*

$$E_i = \{j | h(e_j) = i, j \in [d]\}$$

The conjunction $h = \wedge_{j \in E_0} \bar{x}_j$ shatters E where x_j is jth coordinate of the d-dimensional unit vector. If $E_0 = \Phi$, then $h = \phi$.

Proof. Suppose E_0 is non-empty. Consider any e_i such that $i \in E_1$. So, $\forall j \in E_0, e_{ij} = 0 \implies \wedge_{j \in E_0} \bar{e}_{ij} = 1$. Similarly, consider any e_k such that $k \in E_0 \implies e_{kk} = 1$. So, $\wedge_{j \in E_0} \bar{e}_{kj} = (\wedge_{j \in E_0 \setminus \{k\}} \bar{e}_{kj}) \wedge \bar{e}_{kk} = 0$. If E_0 is empty, $h = \phi$. So, the hypothesis trivially predicts 1 all the time which satisfies our condition. □

■