High Dimensional Statistics

Homework 1

Javad Hezareh (404208723)

Fall 2025 (1404)

Problem 1

Let us define $S := \frac{\partial}{\partial\theta} \log f(X;\theta)$ where $f(X,\theta)$ is the distribution of the data X given our model parameter θ. For the expectation of the S with respect to X we have:

$$\begin{aligned}
\mathbb{E}_X[S] &= \mathbb{E}_X\left[\frac{\partial}{\partial\theta}\log f(X;\theta)\right] \\
&= \int \frac{\partial}{\partial\theta}\log f(X;\theta) f(X;\theta)dX \\
&= \int \frac{1}{f(X;\theta)}\frac{\partial}{\partial\theta} f(X;\theta) f(X;\theta)dX \\
&= \int \frac{\partial}{\partial\theta} f(X;\theta)dX \\
&= \frac{\partial}{\partial\theta}\int f(X;\theta)dX \\
&= \frac{\partial}{\partial\theta}1 = 0
\end{aligned}$$

In other words the Fisher information or $I(\theta)$ can be written as:

$$\begin{aligned}
I(\theta) &= \mathbb{E}_X\left[\left(\frac{\partial}{\partial\theta}\log f(X;\theta)\right)^2\right] \\
&= \mathbb{E}_X\left[(S - \mathbb{E}[S])^2\right] \\
&= \mathrm{Var}(S)
\end{aligned}$$

Now from Cauchy-Schwarz inequality we know that:

$$(\mathrm{Cov}(X,Y))^2 \leq \mathrm{Var}(X)\mathrm{Var}(Y)$$

We can use this inequality to bound the variance of our estimator $T = t(X)$. In other words we have:

$$\begin{aligned}
(\mathrm{Cov}(T,S))^2 &\leq \mathrm{Var}(T)\mathrm{Var}(S) \\
\Rightarrow \mathrm{Var}(T) &\geq \frac{(\mathrm{Cov}(T,S))^2}{\mathrm{Var}(S)} \qquad (\mathrm{Var}(S) > 0) \\
&= \frac{(\mathrm{Cov}(T,S))^2}{I(\theta)}
\end{aligned}$$

Now we need to compute $\text{Cov}(T, S)$:

$$\begin{aligned}
\text{Cov}(T, S) &= \mathbb{E}[TS] - \mathbb{E}[T]\mathbb{E}[S] \\
&= \mathbb{E}[TS] \quad (\text{since } \mathbb{E}[S] = 0) \\
&= \mathbb{E}\left[T \frac{\partial}{\partial \theta} \log f(X;\theta)\right] \\
&= \int t(X) \frac{\partial}{\partial \theta} \log f(X;\theta) f(X;\theta) dX \\
&= \int t(X) \frac{1}{f(X;\theta)} \frac{\partial}{\partial \theta} f(X;\theta) f(X;\theta) dX \\
&= \int t(X) \frac{\partial}{\partial \theta} f(X;\theta) dX \\
&= \frac{\partial}{\partial \theta} \int t(X) f(X;\theta) dX \\
&= \frac{\partial}{\partial \theta} \mathbb{E}[T] = \psi'(\theta)
\end{aligned}$$

Putting it all together we have:

$$\text{Var}(T) \geq \frac{(\psi'(\theta))^2}{I(\theta)}$$

Problem 2

For a vector X the maximization of $\theta^T X$ under the constraint $\|\theta\|_2 \leq 1$ is equal to $\|X\|_2$. To see this, we can use the Cauchy-Schwarz inequality as follows:

$$\begin{aligned}
\theta^T X &\leq \|\theta\|_2 \|X\|_2 \\
&\leq 1 \cdot \|X\|_2 \quad (\|\theta\|_2 \leq 1) \\
&= \|X\|_2
\end{aligned}$$

The equality holds when θ is in the same direction as X and has unit norm. Therefore we can write:

$$\max_{\|\theta\|_2 \leq 1} \theta^T X = \|X\|_2$$

The same thing happens for $\max_{\|\theta\|_2 \leq 1} |\theta^T X|$. To find an upper bound for this value, we can use ϵ-net over the unit sphere in $\mathbb{R}^d$. Let's denote this ϵ-net by $\mathcal{N}_\epsilon$. We know for any vector θ with $\|\theta\|_2 \leq 1$, there exists a vector $\theta' \in \mathcal{N}_\epsilon$ such that $\|\theta - \theta'\|_2 \leq \epsilon$. Therefore we can write:

$$\begin{aligned}
\theta^T X &= (\theta' + (\theta - \theta'))^T X \\
&= \theta'^T X + (\theta - \theta')^T X \\
&\leq \theta'^T X + \epsilon \|X\|_2
\end{aligned}$$

This inequality holds for any θ with $\|\theta\|_2 \leq 1$. Therefore it holds for the maximum of both sides:

$$\begin{aligned}
\max_{\|\theta\|_2 \leq 1} \theta^T X &\leq \max_{\theta' \in \mathcal{N}_\epsilon} \theta'^T X + \epsilon \|X\|_2 \\
\implies \|X\|_2 &\leq \max_{\theta' \in \mathcal{N}_\epsilon} \theta'^T X + \epsilon \|X\|_2 \\
\implies (1-\epsilon)\|X\|_2 &\leq \max_{\theta' \in \mathcal{N}_\epsilon} \theta'^T X \\
\implies \|X\|_2 &\leq \frac{1}{1-\epsilon} \max_{\theta' \in \mathcal{N}_\epsilon} \theta'^T X
\end{aligned}$$

Therefore for $\lambda > 0$ and $\epsilon = 1/2$ we have:

$$\begin{aligned}
e^{\lambda \mathbb{E}[\|X\|_2]} &\leq \mathbb{E}\left[e^{\lambda \|X\|_2}\right] && \text{(Jensen's ineqaulity)} \\
&\leq \mathbb{E}\left[e^{\frac{\lambda}{1-1/2} \max_{\theta' \in \mathcal{N}_{1/2}} \theta'^T X}\right] \\
&= \mathbb{E}\left[\max_{\theta' \in \mathcal{N}_{1/2}} e^{2\lambda \theta'^T X}\right] \\
&\leq \mathbb{E}\left[\sum_{\theta' \in \mathcal{N}_{1/2}} e^{2\lambda \theta'^T X}\right] && \left(0 \leq \alpha_t \implies \max_t \alpha_t \leq \sum_t \alpha_t\right) \\
&= \sum_{\theta' \in \mathcal{N}_{1/2}} \mathbb{E}\left[e^{2\lambda \theta'^T X}\right] \\
&\leq \sum_{\theta' \in \mathcal{N}_{1/2}} e^{2\lambda^2 \sigma^2} && (\theta'^T X \sim \mathrm{SubG}(\sigma)) \\
&\leq (1+4)^d e^{2\lambda^2 \sigma^2} && \left(|\mathcal{N}_{1/2}| \leq (1 + 2/\epsilon)^d\right)
\end{aligned}$$

Taking the logarithm of both sides we have:

$$\lambda\mathbb{E}[\|X\|_2] \leq d\log(5) + 2\lambda^2\sigma^2$$
$$\Rightarrow \mathbb{E}[\|X\|_2] \leq \frac{d\log(5)}{\lambda} + 2\lambda\sigma^2$$

To get the tightest bound, we can minimize the right-hand side with respect to λ:

$$\begin{aligned}
-\frac{d\log(5)}{\lambda^2} + 2\sigma^2 = 0 \\
\Rightarrow \lambda^2 &= \frac{d\log(5)}{2\sigma^2} \\
\Rightarrow \lambda &= \sqrt{\frac{d\log(5)}{2\sigma^2}} \qquad (\lambda > 0)
\end{aligned}$$

Plugging this value of λ back into the inequality we get:

$$\begin{aligned}
\mathbb{E}[\|X\|_2] &\leq \frac{d\log(5)}{\sqrt{\frac{d\log(5)}{2\sigma^2}}} + 2\sigma^2\sqrt{\frac{d\log(5)}{2\sigma^2}} \\
&= \sqrt{8\sigma^2 d\log(5)} \\
&\leq 4\sigma\sqrt{d} \qquad (\sqrt{8\log(5)} \approx 3.59 < 4)
\end{aligned}$$

To show a high-probability bound for $\|X\|_2$, we consider $\epsilon = 1/2$ then we ahve:

$$\|X\|_2 \leq 2\max_{\theta' \in \mathcal{N}_\epsilon} \theta'^T X$$

Therefore if we know $\|X\|_2 \geq \alpha$ then we would have $2\max_{\theta'\in\mathcal{N}_\epsilon} \theta'^T X \geq \alpha$ or $\max_{\theta'\in\mathcal{N}_\epsilon} \theta'^T X \geq \alpha/2$. Thus we can write:

$$\begin{aligned}
\mathbb{P}\left(\|X\|_2 \geq \alpha\right) &\leq \mathbb{P}\left(\max_{\theta'\in\mathcal{N}_\epsilon} \theta'^T X \geq \alpha/2\right) \\
&= \mathbb{P}\left(\bigcup_{\theta'\in\mathcal{N}_\epsilon} \theta'^T X \geq \alpha/2\right) \\
&\leq \sum_{\theta'\in\mathcal{N}_\epsilon} \mathbb{P}\left(\theta'^T X \geq \alpha/2\right) \\
&\leq \sum_{\theta'\in\mathcal{N}_\epsilon} e^{-\frac{\alpha^2}{8\sigma^2}} \qquad (\theta'^T X \sim \mathrm{SubG}(\sigma)) \\
&\leq (1+4)^d e^{-\frac{\alpha^2}{8\sigma^2}}
\end{aligned}$$

Setting the right-hand side to δ and solving for α we have:

$$\begin{aligned}
\delta &= 5^d e^{-\frac{\alpha^2}{8\sigma^2}} \\
\Rightarrow \frac{\alpha^2}{8\sigma^2} &= \log\left(\frac{5^d}{\delta}\right) \\
\Rightarrow \alpha &= \sigma\sqrt{8\log\left(\frac{5^d}{\delta}\right)}
\end{aligned}$$

Thus we have:

$$
\begin{aligned}
&\mathbb{P}\left(\|X\|_2 \geq \sigma\sqrt{8\log\left(\frac{5^d}{\delta}\right)}\right) \leq \delta \\
\implies &\mathbb{P}\left(\|X\|_2 \leq \sigma\sqrt{8\log\left(\frac{5^d}{\delta}\right)}\right) \geq 1-\delta \\
\implies &\mathbb{P}\left(\|X\|_2 \leq \sigma\sqrt{8d\log 5 + 8\log(1/\delta)}\right) \geq 1-\delta \\
\implies &\mathbb{P}\left(\|X\|_2 \leq \sigma\sqrt{8d\log 5} + \sigma\sqrt{8\log(1/\delta)}\right) \geq 1-\delta \quad (\sqrt{a+b} \leq \sqrt{a}+\sqrt{b}, a, b \geq 0) \\
\implies &\mathbb{P}\left(\|X\|_2 \leq 4\sigma\sqrt{d} + 2\sigma\sqrt{2\log(1/\delta)}\right) \geq 1-\delta \quad (\sqrt{8\log 5} < 4)
\end{aligned}
$$

Problem 3

We first show for the exponential function we have:

$$e^x \leq 1 + x + \frac{x^2/2}{1 - |x|/3} \quad \text{for } \; |x| < 3$$

To show this, we expand the exponential function as a Taylor series around 0:

$$\begin{aligned}
e^x &= \sum_{n=0}^{\infty} \frac{x^n}{n!} = 1 + x + \frac{x^2}{2} + \frac{x^3}{3!} + \frac{x^4}{4!} + \cdots \\
&= 1 + x + \frac{x^2}{2}\left(1 + \frac{x}{3} + \frac{x^2}{3 \times 4} + \cdots\right) \\
&\leq 1 + x + \frac{x^2}{2}\left(1 + \frac{|x|}{3} + \frac{|x|^2}{3 \times 4} + \cdots\right) \\
&\leq 1 + x + \frac{x^2}{2}\left(1 + \frac{|x|}{3} + \left(\frac{|x|}{3}\right)^2 + \cdots\right) \quad \text{for } \; |x| < 3 \\
&= 1 + x + \frac{x^2}{2}\left(\frac{1}{1 - |x|/3}\right) \quad \text{for } \; |x| < 3
\end{aligned}$$

Now we use our assumptions about the provided random variable X to bound its moment generating function (MGF). Let's set $|\lambda| < \frac{3}{K}$, then we have:

$$\left.\begin{aligned} |X| &\leq K \\ |\lambda| &< \frac{3}{K} \end{aligned}\right\} \Rightarrow |\lambda X| < 3$$

Then we can use the previous result to bound the MGF of X as follows:

$$\begin{aligned}
\mathbb{E}[e^{\lambda X}] &\leq \mathbb{E}\left[1 + \lambda X + \frac{(\lambda X)^2/2}{1 - |\lambda X|/3}\right] \\
&= 1 + \lambda \mathbb{E}[X] + \mathbb{E}\left[\frac{\lambda^2 X^2/2}{1 - |\lambda X|/3}\right] \\
&= 1 + 0 + \mathbb{E}\left[\frac{\lambda^2 X^2/2}{1 - |\lambda X|/3}\right] && (\mathbb{E}[X] = 0) \\
&\leq 1 + \mathbb{E}\left[\frac{\lambda^2 X^2/2}{1 - |\lambda| K/3}\right] && (|X| \leq K) \\
&= 1 + \frac{\lambda^2/2}{1 - |\lambda| K/3} \mathbb{E}[X^2] \\
&\leq \exp\left(\frac{\lambda^2/2}{1 - |\lambda| K/3} \mathbb{E}[X^2]\right) && (1 + x \leq e^x)
\end{aligned}$$

Problem 4

We first show that the L^p norm of a random variable is not less than the L^2 norm of that. In other words for a random variable X and $p \geq 2$ and using the definition of the L^p norm we have:

$$(\mathbb{E}[|X|^p])^{1/p} \geq \left(\mathbb{E}[|X|^2]\right)^{1/2}$$

Let's define a new random variable $Y = |X|^2$ and $r := p/2$. Now we can use Jensen's inequality since the function $f(x) = x^r$ is convex for $r \geq 1$ and $x \geq 0$. Thus we have:

$$\begin{aligned}
(\mathbb{E}[|X|^p])^{1/p} &= (\mathbb{E}[Y^r])^{1/p} \\
&\geq (\mathbb{E}[Y])^{r/p} && \text{(Jensen's inequality)} \\
&= \left(\mathbb{E}[|X|^2]\right)^{(p/2)/p} \\
&= \left(\mathbb{E}[|X|^2]\right)^{1/2}
\end{aligned}$$

Now for the random variable $Z = \sum_{i=1}^N a_i X_i$ we can use the previous result to bound its L^p norm as follows:

$$\begin{aligned}
\|Z\|_{L^p} &= \left(\mathbb{E}\left[\left|\sum_{i=1}^N a_i X_i\right|^p\right]\right)^{1/p} \\
&\geq \left(\mathbb{E}\left[\left|\sum_{i=1}^N a_i X_i\right|^2\right]\right)^{1/2} && \text{(from previous result)} \\
&= \left(\mathbb{E}\left[\left(\sum_{i=1}^N a_i X_i\right)^2\right]\right)^{1/2} && (|\alpha|^2 = \alpha^2) \\
&= \left(\mathbb{E}\left[\sum_{i=1}^N a_i^2 X_i^2\right] + \mathbb{E}\left[\sum_{i,j} a_i a_j X_i X_j\right]\right)^{1/2} \\
&= \left(\sum_{i=1}^N a_i^2 \mathbb{E}[X_i^2] + \sum_{i,j} a_i a_j \mathbb{E}[X_i]\mathbb{E}[X_j]\right)^{1/2} && \text{(independence of } X_i\text{s)} \\
&= \left(\sum_{i=1}^N a_i^2 \mathbb{E}[X_i^2]\right)^{1/2} && (\mathbb{E}[X_i] = 0) \\
&= \left(\sum_{i=1}^N a_i^2\right)^{1/2} && (\mathrm{Var}(X_i) = 1)
\end{aligned}$$

To show the upper bound, we can use the triangle inequality of the L^p norm as follows:

$$\begin{aligned}
\|Z\|_{L^p} &= \left(\mathbb{E}\left[\left|\sum_{i=1}^{N} a_i X_i\right|^p\right]\right)^{1/p} \\
&\leq \sum_{i=1}^{N} |a_i| \left(\mathbb{E}\left[|X_i|^p\right]\right)^{1/p} && \text{(Triangle inequality)} \\
&= \left(\mathbb{E}\left[|X_1|^p\right]\right)^{1/p} \sum_{i=1}^{N} |a_i| && \text{(identical distribution of } X_i\text{s)} \\
&\leq \left(\mathbb{E}\left[|X_1|^p\right]\right)^{1/p} \sqrt{N} \left(\sum_{i=1}^{N} a_i^2\right)^{1/2} && \text{(Cauchy-Schwarz inequality)}
\end{aligned}$$

Now we show the upper bound of the L^p norm of Z. From the definition of the ψ-norm of a random variable we can write:

$$\|Z\|_{L^p} \leq C_0\sqrt{p}\,\|Z\|_{\psi} \qquad (\star)$$

And as X_is are Sub-Gaussian random variables, then Z is also a Sub-Gaussian random variable and we have the above inequality holds for it. We also have the following ineqaulity for the ψ-norm of the summation of Sub-Gaussian random variables:

$$\left\|\sum_i Y_i\right\|_{\psi}^2 \leq C_1 \sum_i \|Y_i\|_{\psi}^2$$

And we know $Y_i = a_i X_i$ therefore $\|a_i X_i\|^2 = a_i^2 \|X_i\|_{\psi}^2$. Thus we can write:

$$\begin{aligned}
\|Z\|_{\psi}^2 &= \left\|\sum_{i=1}^{N} a_i X_i\right\|_{\psi}^2 \\
&\leq C_1 \sum_{i=1}^{N} \|a_i X_i\|_{\psi}^2 \\
&= C_1 \sum_{i=1}^{N} a_i^2 \|X_i\|_{\psi}^2 \\
&\leq C_1 \sum_{i=1}^{N} a_i^2 K^2 && (K = \max_i \|X_i\|_{\psi}) \\
&= C_1 K^2 \sum_{i=1}^{N} a_i^2
\end{aligned}$$

Now using $(\star)$ we can write:

$$\begin{aligned}
\|Z\|_{L^p} &\leq C_0\sqrt{p}\,\|Z\|_{\psi} \\
&\leq C_0\sqrt{p}\sqrt{C_1 K^2 \sum_{i=1}^{N} a_i^2} \\
&= C_0 K \sqrt{C_1 p} \left(\sum_{i=1}^{N} a_i^2\right)^{1/2} = CK\sqrt{p}\left(\sum_{i=1}^{N} a_i^2\right)^{1/2}
\end{aligned}$$

Problem 5

We assume $X_1, \dots, X_n$ are zero-mean Sub-Gaussian random variables.

(a) As the exponential function is increasing, we can write:

$$e^{\lambda Z} = e^{\lambda \max_t X_t} = \max_t e^{\lambda X_t}$$

Now for the MGF of Z we have:

$$\begin{aligned}
\mathbb{E}[e^{\lambda Z}] &= \mathbb{E}[\max_t e^{\lambda X_t}] \\
&\leq \mathbb{E}\left[\sum_{t=1}^{n} e^{\lambda X_t}\right] && \left(0 \leq \alpha_t \implies \max_t \alpha_t \leq \sum_t \alpha_t\right) \\
&= \sum_{t=1}^{n} \mathbb{E}[e^{\lambda X_t}] \\
&\leq \sum_{t=1}^{n} e^{\lambda^2\sigma^2/2} && (X_i \sim \mathrm{SubG}(\sigma), \mathbb{E}(X_i) = 0) \\
&= ne^{\lambda^2\sigma^2/2}
\end{aligned}$$

We know $e^{\alpha x}$ is convex for all $\alpha \in \mathbb{R}$. Therefore using Jensen's inequality we have:

$$\begin{aligned}
e^{\lambda \mathbb{E}[Z]} &\leq \mathbb{E}[e^{\lambda Z}] \\
&\leq ne^{\lambda^2\sigma^2/2}
\end{aligned}$$

Now taking the logarithm of both sides we get:

$$\begin{aligned}
\lambda \mathbb{E}[Z] &\leq \log(n) + \frac{\lambda^2\sigma^2}{2} \\
(\lambda \geq 0) \Rightarrow \mathbb{E}[Z] &\leq \frac{\log(n)}{\lambda} + \frac{\lambda\sigma^2}{2}
\end{aligned}$$

We've consider $\lambda \geq 0$. Now we can minimize the right-hand side with respect to λ to get the tightest bound. Taking the derivative and setting it to zero we have:

$$\begin{aligned}
-\frac{\log(n)}{\lambda^2} + \frac{\sigma^2}{2} &= 0 \\
\Rightarrow \lambda^2 &= \frac{2\log(n)}{\sigma^2} \\
\Rightarrow \lambda &= \sqrt{\frac{2\log(n)}{\sigma^2}} && (\lambda \geq 0)
\end{aligned}$$

Plugging this value of λ back into the inequality we get:

$$\begin{aligned}
\mathbb{E}[Z] &\leq \frac{\log(n)}{\sqrt{\frac{2\log(n)}{\sigma^2}}} + \frac{\sigma^2}{2}\sqrt{\frac{2\log(n)}{\sigma^2}} \\
&= \sigma\sqrt{2\log(n)}
\end{aligned}$$

(b) We can use Chernoff bound to bound the tail probability of Z as follows:

$$\begin{aligned}\mathbb{P}(Z \geq t) &\leq \inf_{\lambda>0} e^{-\lambda t}\mathbb{E}[e^{\lambda Z}] \\ &\leq \inf_{\lambda>0} ne^{\lambda^2\sigma^2/2-\lambda t} && \text{(from part (a))}\end{aligned}$$

To find the tightest bound, we minimize the right-hand side with respect to λ:

$$\begin{aligned}ne^{\lambda^2\sigma^2/2-\lambda t}(\sigma^2\lambda - t) &= 0 \\ \Rightarrow \sigma^2\lambda - t &= 0 \\ \Rightarrow \lambda &= \frac{t}{\sigma^2} && (\lambda > 0)\end{aligned}$$

Now for the tail bound we have:

$$\begin{aligned}\mathbb{P}(Z \geq t) &\leq ne^{\frac{t^2}{2\sigma^2}-\frac{t^2}{\sigma^2}} \\ &= ne^{-\frac{t^2}{2\sigma^2}}\end{aligned}$$

We can simply rewrite this and set $t = \sqrt{2\sigma^2\log(n/\delta)}$ to get:

$$\begin{aligned}\mathbb{P}\left(Z \geq \sqrt{2\sigma^2\log\left(\frac{n}{\delta}\right)}\right) &\leq ne^{-\log(n/\delta)} \\ &= \delta\end{aligned}$$

Problem 6

We define new random variables $Y_t^i = \mathbb{I}(X_t = i)$. Therefore, Y_t^i are Bernoulli random variables with parameter p_i. We know Y_t^i are Sub-Gaussian with parameter $1/2$. Using the Hoeffding bound for the average of these random variables we have:

$$\begin{aligned}
\mathbb{P}\left(|\hat{p}_i - p_i| \geq \epsilon\right) &= \mathbb{P}\left(\left|\frac{1}{n}\sum_{t=1}^{n} Y_t^i - p_i\right| \geq \epsilon\right) \\
&\leq 2\exp\left(-\frac{n^2\epsilon^2}{2n(1/2)^2}\right) && \text{(Hoeffding Bound)} \\
&= 2\exp\left(-2n\epsilon^2\right) && (\star)
\end{aligned}$$

For probability vectors like p and $\hat{p}$ we know that:

$$\|p - \hat{p}\|_\infty \leq \alpha \implies \|p - \hat{p}\|_1 \leq 2\alpha$$

Therefore the event of $\|p - \hat{p}\|_\infty \leq \alpha$ is a subset of $\|p - \hat{p}\|_1 \leq 2\alpha$ and we can write:

$$\mathbb{P}\left(\|p - \hat{p}\|_\infty \leq \alpha\right) \leq \mathbb{P}\left(\|p - \hat{p}\|_1 \leq 2\alpha\right)$$

And for the complement event we have:

$$\mathbb{P}\left(\|p - \hat{p}\|_1 \geq 2\alpha\right) \leq \mathbb{P}\left(\|p - \hat{p}\|_\infty \geq \alpha\right)$$

We cound bound the right-hand side using union bound as follows:

$$\begin{aligned}
\mathbb{P}\left(\|p - \hat{p}\|_\infty \geq \alpha\right) &= \mathbb{P}\left(\max_i |p_i - \hat{p}_i| \geq \alpha\right) \\
&= \mathbb{P}\left(\bigcup_i |p_i - \hat{p}_i| \geq \alpha\right) \\
&\leq \sum_i \mathbb{P}\left(|p_i - \hat{p}_i| \geq \alpha\right) \\
&\leq \sum_i 2\exp\left(-2n\alpha^2\right) && \text{(Using } \star\text{)} \\
&= 2m\exp\left(-2n\alpha^2\right)
\end{aligned}$$

Therefore for $\epsilon = 2\alpha$ we have:

$$\mathbb{P}\left(\|p - \hat{p}\|_1 \geq \epsilon\right) \leq \mathbb{P}\left(\|p - \hat{p}\|_\infty \geq \epsilon/2\right) \leq 2m\exp\left(-n\epsilon^2/2\right)$$

Setting $\delta = 2m\exp\left(-n\epsilon^2/2\right)$ and solving for ϵ we get:

$$\begin{aligned}
\delta &= 2m\exp\left(-n\epsilon^2/2\right) \\
\Rightarrow \epsilon &= \sqrt{\frac{2}{n}\log\left(\frac{2m}{\delta}\right)}
\end{aligned}$$

Therefore we have:

$$\mathbb{P}\left(\|p - \hat{p}\|_1 \geq \sqrt{\frac{2}{n}\log\left(\frac{2m}{\delta}\right)}\right) \leq \delta$$

Problem 7

(a) The Bernoulli random variable is a bounded random variable and $X \in [0,1]$. As we know bounded random variables are Sub-Gaussian with parameter $(b-a)/2$, we can say that X is Sub-Gaussian with parameter $1/2$.

(b), (c) For these parts, we show the provided $Q(p)$ satisfies the condition for minimum value of σ. To find the minimum value of σ that satisfies the subgaussianity constant we have:

$$\log\left(\mathbb{E}(e^{\lambda(X-p)})\right) \leq \frac{\sigma^2\lambda^2}{2} \implies \frac{2}{\lambda^2}\log\left(\mathbb{E}(e^{\lambda(X-p)})\right) \leq \sigma^2$$

To find the minimum value of σ^2 we should find the supremum value of the left-hand-side of the above inequality. Therefore:

$$\begin{aligned}
\sigma^2_{min} &= \sup_\lambda \frac{2}{\lambda^2}\log\left(\mathbb{E}(e^{\lambda(X-p)})\right) \\
&= \sup_\lambda \frac{2}{\lambda^2}\log\left(pe^{\lambda(1-p)} + (1-p)e^{-\lambda p}\right) \\
&= \sup_\lambda \frac{2}{\lambda^2}M(\lambda)
\end{aligned}$$

Where we define $M(\lambda) := \log\left(pe^{\lambda(1-p)} + (1-p)e^{-\lambda p}\right)$. Now for the optimal value $\lambda^* \neq 0$ we would have:

$$M'(\lambda^*) = \frac{2M(\lambda^*)}{\lambda^*}$$

One can easily show this by taking the gradient of our function and setting it to zero. For LHS we have:

$$\begin{aligned}
M'(\lambda) &= \frac{p(1-p)e^{-\lambda p}(e^\lambda - 1)}{e^{-\lambda p}(pe^\lambda + 1 - p)} \\
&= \frac{p(1-p)(e^\lambda - 1)}{pe^\lambda + 1 - p} \\
&= \frac{p(1-p)(t^2 - 1)}{pt^2 + 1 - p}
\end{aligned}$$

Where we defined $t^2 := e^\lambda$. For the RHS we can write:

$$\begin{aligned}
\frac{2M(\lambda)}{\lambda} &= \frac{2}{\lambda}\left(\log\left(e^{-\lambda p}(pe^\lambda + 1 - p)\right)\right) \\
&= \frac{2}{\lambda}\left[-\lambda p + \log(pe^\lambda + 1 - p)\right] \\
&= -2p + \frac{2}{\lambda}\log(pt^2 + 1 - p)
\end{aligned}$$

Now we have:

$$\begin{aligned}
&\frac{p(1-p)(t^2-1)}{pt^2+1-p} = -2p + \frac{2}{\lambda}\log(pt^2+1-p) \\
\implies &\frac{p(1-p)(t^2-1)}{pt^2+1-p} + 2p = \frac{2}{\lambda}\log(pt^2+1-p) \\
\implies &\frac{(t^2-1)(1+p)p+2p}{pt^2+1-p} = \frac{2}{\lambda}\log(pt^2+1-p) \\
\implies &p\frac{(t^2-1)(1+p)+2}{pt^2+1-p} = \frac{2}{2\log t}\log(pt^2+1-p) \\
\implies &\log(t)\cdot p\frac{(t^2-1)(1+p)+2}{pt^2+1-p} = \log(pt^2+1-p)
\end{aligned}$$

One candidate solution for this equation is $t^* = (1-p)/p$. One can simply check this would lead to the eqaulity:

$$\begin{aligned}
p(t^*)^2 + 1 - p &= p\frac{1+p^2-2p}{p^2} + 1 - p \\
&= \frac{1+p^2-2p}{p} + \frac{p-p^2}{p} \\
&= \frac{1-p}{p}
\end{aligned}$$

$$\begin{aligned}
((t^*)^2-1)(1+p)+2 &= \left(\frac{1+p^2-2p}{p^2} - 1\right)(1+p) + 2 \\
&= \frac{1-2p}{p^2}(1+p) + 2 \\
&= \frac{1+p-2p-2p^2}{p^2} + \frac{2p^2}{p^2} \\
&= \frac{1-p}{p^2}
\end{aligned}$$

Putting these in our eqaulity we would have:

$$\log(t^*)\frac{p\frac{1-p}{p^2}}{\frac{1-p}{p}} = \log(t^*)\times 1 = \log(p(t^*)^2+1-p)$$

Therfore we have

$$\lambda^* = 2\ln t^* = 2\ln\left(\frac{1-p}{p}\right).$$

Plugging this in our function for σ^2_{min} we have:

$$\begin{aligned}
\sigma^2_{min} &= \frac{2}{{\lambda^*}^2} M(\lambda^*) \\
&= \frac{1}{\left[\ln\left(\frac{1-p}{p}\right)\right]^2} \ln\left(e^{-\lambda^* p}(pe^{\lambda^*} + 1 - p)\right) \\
&= \frac{1}{\left[\ln\left(\frac{1-p}{p}\right)\right]^2} \left[-\lambda^* p + \ln\left(pe^{\lambda^*} + 1 - p\right)\right] \\
&= \frac{1}{\left[\ln\left(\frac{1-p}{p}\right)\right]^2} \left[-2p\ln\left(\frac{1-p}{p}\right) + \ln\left(\frac{1-p}{p}\right)\right] \\
&= \frac{1-2p}{\ln\left(\frac{1-p}{p}\right)}
\end{aligned}$$

(d)

Problem 8

(a) We want to show the output of the $\textsc{Naive}(\epsilon, \delta)$ which is $\max_a \hat{p}_a$ is (ϵ, δ)-PAC algorithm. In other words we want to show:

$$\mathbb{P}\left(\max_a \hat{p}_a - \max_a p_a \geq \epsilon\right) \leq \delta$$

We know $\forall a \in A$ we have $p_a \leq \max_a p_a$. Therefore we can write:

$$\mathbb{P}\left(\max_a \hat{p}_a - \max_a p_a \geq \epsilon\right) \leq \mathbb{P}\left(\max_a (\hat{p}_a - p_a) \geq \epsilon\right)$$

To bound the right-hand side we can use union bound as follows:

$$\begin{aligned}
\mathbb{P}\left(\max_a (\hat{p}_a - p_a) \geq \epsilon\right) &= \mathbb{P}\left(\bigcup_a (\hat{p}_a - p_a) \geq \epsilon\right) \\
&\leq \sum_a \mathbb{P}\left(\hat{p}_a - p_a \geq \epsilon\right) \\
&= \sum_a e^{-2m\epsilon^2} && \text{(Hoeffding bound)} \\
&= ne^{-2m\epsilon^2}
\end{aligned}$$

We can use the Hoeffding bound since $\hat{p}_a$ is the average of m i.i.d. bounded random variables in $[0, 1]$, thus each of which is Sub-Gaussian with parameter $1/2$. Now setting $\delta = ne^{-2m\epsilon^2}$ and solving for m we get:

$$\begin{aligned}
\delta &= ne^{-2m\epsilon^2} \\
\Rightarrow m &= \frac{1}{2\epsilon^2} \log\left(\frac{n}{\delta}\right)
\end{aligned}$$

We need m samples per action to have the $\textsc{Naive}(\epsilon, \delta)$ algorithm be (ϵ, δ)-PAC. Therefore the sample complexity of the algorithm is $O\left(\frac{n}{\epsilon^2} \log\left(\frac{n}{\delta}\right)\right)$

(b) Let's set the output of the algorithm with $\hat{a}$. Then we want to show with low probability this output is different from the optimal solution wich is a_1 for this setup. Thus we want to show:

$$\mathbb{P}\left(\hat{a} \neq a_1\right) \leq \delta$$

According to the algorithm, to have this event happen the a_1 action should be removed from the set before one of the other actions. In other words:

$$\mathbb{P}\left(\hat{a} \neq a_1\right) = \mathbb{P}\left(\bigcup_i E_i\right) \leq \sum_i \mathbb{P}\left(E_i\right)$$

Where each E_i is the event of removing a_1 before a_i.

Determining the sample complexity of the algorithm is straight forward. For each action that gets eliminated at step i we have sampled

$$(t_n - t_{n+1}) + (t_{n-1} - t_n) + \cdots + (t_{n-i} - t_{n-i+1}) = t_{n-i} - t_{n+1} = t_{n-i}$$

times. At each step i we eliminate one action therefore till the end of iteration for those actions that were removed we have sampled:

$$\sum_{i=0}^{n-2} t_{n-i} = \sum_{i=2}^{n} t_i = \sum_{i=2}^{n} \left\lceil \frac{4}{\Delta_i^2} \log\left(\frac{n}{\delta}\right) \right\rceil$$

We have also sampled t_2 times for action a_1. Thus the total sample complexity of the algorithm would be:

$$t_2 + \sum_{i=2}^{n} t_i \in O\left(\log\left(\frac{n}{\delta}\right) \sum_{i=2}^{n} \frac{1}{\Delta_i^2} \right)$$