

STA355: Midterm Exam

Instructor: Wenlong Mou

Oct 17th, 2025

Student name: _____

Student ID: _____

Student signature: _____

This exam contains 11 pages.

Total marks: 100 pts

Time Allowed: 150 minutes

Question 1. [24 points, 3 points each] Mark each statement with T (true) or F (false). No justification required.

(1) If $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{d} Y$, then $X_n + Y_n \xrightarrow{d} X + Y$. **Answer:** F.

(2) If X_1, X_2, \dots are i.i.d. random variables with $\mathbb{E}[X_i] = \mu$ and $\text{var}(X_i) = \sigma^2$. Suppose that g is a continuously differentiable function with $g'(\mu) \neq 0$, then the sample mean \bar{X}_n satisfies $\sqrt{n}(g(\bar{X}_n) - g(\mu)) \xrightarrow{d} \mathcal{N}(0, \sigma^2(g'(\mu))^2)$.

Answer: T.

(3) Let p be a pdf on \mathbb{R} that is symmetric about 0, i.e., $p(x) = p(-x)$ for all $x \in \mathbb{R}$. For $X_1, X_2, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} p$, the sample mean \bar{X}_n satisfies $\bar{X}_n \xrightarrow{p} 0$. **Answer:** F.

(4) Let $X_1, X_2, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} \mathbb{P}$, where \mathbb{P} is a one-dimensional distribution with cdf F . Then the empirical cdf \hat{F}_n satisfies $\sup_{x \in \mathbb{R}} |\hat{F}_n(x) - F(x)| \xrightarrow{p} 0$.

Answer: T.

(5) Let $X_1, X_2, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} \mathbb{P}$. Suppose that we use the nonparametric bootstrap for inference. Let $X_1^*, X_2^*, \dots, X_n^*$ be a bootstrap sample drawn, then $X_1^*, X_2^*, \dots, X_n^*$ are conditionally i.i.d. given the data. **Answer:** T.

(6) Let $X_1, X_2, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} p_\theta$, where $\theta \in \mathbb{R}^d$ is an unknown parameter. Then the MLE $\hat{\theta}_n$ always exists and is unique. **Answer:** F.

(7) Let $X_1, X_2, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} p_\theta$, where $\theta \in \mathbb{R}^d$ is an unknown parameter. Suppose that the MLE $\hat{\theta}_n$ exists and is unique, and the Fisher information $I(\theta)$ exists and is non-singular for every θ . Then for any unbiased estimator $\tilde{\theta}_n$, we have $\lim_{n \rightarrow \infty} n \cdot \text{var}_\theta(\tilde{\theta}_n) \geq \lim_{n \rightarrow \infty} n \cdot \text{var}_\theta(\hat{\theta}_n)$, for any $\theta \in \mathbb{R}^d$. **Answer:** T.

(8) Let $X_1, X_2, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} \mathbb{P}$. Suppose that $\hat{\theta}_n$ is an estimator for parameter $\theta \in \mathbb{R}$, and \hat{b}_n is the jackknife estimate of bias. Then $\hat{\theta}_n - \hat{b}_n$ is an unbiased estimator for θ . **Answer:** F.

Question 2. [22 points] Basic probabilistic convergence and empirical cdf

Part (a) [7 points]. Let F be a cdf on \mathbb{R} . Let the data be $X_1, X_2, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} F$. Given a collection of deterministic points $x_1, x_2, \dots, x_m \in \mathbb{R}$, find the asymptotic distribution of the random vector

$$\sqrt{n} \cdot [\widehat{F}_n(x_j) - F(x_j)]_{j=1,2,\dots,m},$$

where \widehat{F}_n is the empirical cdf.

Answer: For each $j = 1, 2, \dots, m$, note that $\widehat{F}_n(x_j) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{X_i \leq x_j}$. Let $Y_{i,j} = \mathbb{1}_{X_i \leq x_j}$, then $Y_{i,j} \stackrel{\text{i.i.d.}}{\sim} \text{Ber}(F(x_j))$ with $\mathbb{E}[Y_{i,j}] = F(x_j)$ and $\text{var}(Y_{i,j}) = F(x_j)(1 - F(x_j))$.

By the multivariate Central Limit Theorem, the vector

$$\sqrt{n} \cdot [\widehat{F}_n(x_j) - F(x_j)]_{j=1,2,\dots,m} \xrightarrow{d} \mathcal{N}(0, \Sigma),$$

where Σ is an $m \times m$ covariance matrix with entries

$$\Sigma_{j,k} = \text{cov}(\mathbb{1}_{X_1 \leq x_j}, \mathbb{1}_{X_1 \leq x_k}) = F(\min\{x_j, x_k\}) - F(x_j)F(x_k).$$

In particular, $\Sigma_{j,j} = F(x_j)(1 - F(x_j))$.

Rubrics: You get full points if you correctly identify the asymptotic normal distribution and compute the covariance matrix.

2 points for correctly explaining the CLT.

2 points for correctly finding the mean vector and diagonal entries of the covariance matrix.

Part (b) [7 points]. Let $X_1, X_2, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} \text{Ber}(p)$ for some $p \in (0, 1)$. Let $\hat{p}_n := \frac{1}{n} \sum_{i=1}^n X_i$ be the sample proportion of successes. Find the asymptotic distribution of

$$Z_n := n \cdot \left\{ \hat{p}_n \log \frac{\hat{p}_n}{p} + (1 - \hat{p}_n) \log \frac{1 - \hat{p}_n}{1 - p} \right\}.$$

Answer: By the Delta method, we can analyze Z_n as follows. Define $g(q) = q \log \frac{q}{p} + (1 - q) \log \frac{1 - q}{1 - p}$.

We have $g(p) = 0$, and computing derivatives:

$$g'(p) = \log \frac{p}{p} - \log \frac{1-p}{1-p} + 1 - 1 = \log \frac{p(1-p)}{p(1-p)} = 0.$$

Computing the second derivative:

$$g''(q) = \frac{1}{q} + \frac{1}{1-q}.$$

By the Central Limit Theorem, $\sqrt{n}(\hat{p}_n - p) \xrightarrow{d} \mathcal{N}(0, p(1-p))$. Using a second-order Taylor expansion and the delta method:

$$Z_n = n \cdot g(\hat{p}_n) \approx n \cdot \frac{1}{2} g''(p)(\hat{p}_n - p)^2 = \frac{n}{2p(1-p)}(\hat{p}_n - p)^2.$$

Therefore,

$$Z_n \xrightarrow{d} \frac{1}{2p(1-p)} \cdot [\sqrt{n}(\hat{p}_n - p)]^2 \xrightarrow{d} \frac{1}{2}\xi^2,$$

where $\xi \sim \mathcal{N}(0, 1)$. This is a chi-squared distribution with 1 degree of freedom scaled by $\frac{1}{2}$.

Rubrics: You get full points if you correctly apply Delta method and derive the asymptotic distribution.

3 points for correctly explaining the idea of using Taylor expansion and Delta method.

2 points for correctly finding the asymptotic distribution of \hat{p}_n .

Part (c) [8 points]. Let X_1, X_2, \dots, X_n be i.i.d. random variables drawn from a distribution with cdf F . Assume that F is continuous and strictly increasing everywhere. The sample median \hat{m}_n is defined as the middle order statistic (for simplicity, we will assume n is odd). Let $m = F^{-1}(1/2)$ be the population median. Show that $\hat{m}_n \xrightarrow{P} m$.

Answer: Fix $\varepsilon > 0$. Since F is continuous and strictly increasing, $F(m - \varepsilon) < \frac{1}{2} < F(m + \varepsilon)$. Let

$$\delta = \min \left\{ \frac{1}{2} - F(m - \varepsilon), F(m + \varepsilon) - \frac{1}{2} \right\} > 0.$$

If $\sup_x |\hat{F}_n(x) - F(x)| < \delta$, then $\hat{F}_n(m - \varepsilon) < \frac{1}{2}$ and $\hat{F}_n(m + \varepsilon) > \frac{1}{2}$. Hence the empirical median (the 1/2-quantile of \hat{F}_n) satisfies $m - \varepsilon < \hat{m}_n < m + \varepsilon$, i.e., $|\hat{m}_n - m| < \varepsilon$.

Therefore,

$$\mathbb{P}(|\hat{m}_n - m| \geq \varepsilon) \leq \mathbb{P}\left(\sup_x |\hat{F}_n(x) - F(x)| \geq \delta\right) \rightarrow 0$$

by the Glivenko—Cantelli theorem (or by the DKW inequality). Thus $\hat{m}_n \xrightarrow{P} m$.

Rubrics: Any valid proof showing the convergence in probability will get full points.

3 points for correctly using the definition of convergence in probability and turning this into an ε - δ argument.

3 points for using the uniform convergence of empirical cdf to the true cdf.

If you only show pointwise convergence of empirical cdf, give 2 points.

Question 3. [28 points] The bootstrap

Part (a) [10 points]. Let $X_1, X_2, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} \text{Poi}(\lambda)$ for some $\lambda > 0$. The probability mass function of $\text{Poi}(\lambda)$ is given by $p_\lambda(k) = e^{-\lambda} \frac{\lambda^k}{k!}$ for $k = 0, 1, 2, \dots$. Let $\hat{\lambda}_n := \frac{1}{n} \sum_{i=1}^n X_i$ be the sample mean.

- Find the nonparametric bootstrap estimate of standard error for $\hat{\lambda}_n$. (in this case, the idealized bootstrap procedure admits a closed-form expression, so you do not need to do simulation).
- Find the parametric bootstrap estimate of standard error for $\hat{\lambda}_n$. (similarly, you can compute closed-form formula without the need to do simulation).

Answer:

- **Nonparametric bootstrap:** Let $X_1^*, X_2^*, \dots, X_n^*$ be a bootstrap sample drawn from the empirical distribution. The nonparametric bootstrap estimate of standard error is given by the standard deviation of the bootstrap sample mean:

$$\begin{aligned}\widehat{\text{SE}}_{\text{nonparam}} &= \sqrt{\text{var}\left(\frac{1}{n} \sum_{i=1}^n X_i^* \mid X_1, X_2, \dots, X_n\right)} \\ &= \frac{1}{\sqrt{n}} \sqrt{\text{var}(X_1^* \mid X_1, X_2, \dots, X_n)} = \frac{1}{\sqrt{n}} \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - \hat{\lambda}_n)^2}.\end{aligned}$$

- **Parametric bootstrap:** For the parametric bootstrap, we use the Poisson model with parameter $\hat{\lambda}_n$. Let the bootstrap sample $X_1^*, X_2^*, \dots, X_n^* \stackrel{\text{i.i.d.}}{\sim} \text{Poi}(\hat{\lambda}_n)$. The parametric bootstrap estimate of standard error is given by:

$$\begin{aligned}\widehat{\text{SE}}_{\text{param}} &= \sqrt{\text{var}\left(\frac{1}{n} \sum_{i=1}^n X_i^* \mid X_1, X_2, \dots, X_n\right)} \\ &= \frac{1}{\sqrt{n}} \sqrt{\text{var}(X_1^* \mid \hat{\lambda}_n)} = \frac{1}{\sqrt{n}} \sqrt{\hat{\lambda}_n}.\end{aligned}$$

Rubrics: Each case is worth 5 points. For each case, you get 3 points for correctly explaining the bootstrap procedure (i.e. how to compute variance of the bootstrap sample mean), and 2 points for correctly deriving the closed-form formula.

For each case, if you only give a final formula without explanation, give 3 points.

Part (b) [8 points]. Suppose that $X_1, X_2, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} \mathbb{P}$, where \mathbb{P} is a distribution on \mathbb{R} . Derive the nonparametric bootstrap estimate of the skewness of the sample mean \bar{X}_n

$$\mu_3(\bar{X}_n) = \frac{\mathbb{E}[(\bar{X}_n - \mathbb{E}\bar{X}_n)^3]}{(\text{var}(\bar{X}_n))^{3/2}}.$$

Describe how to compute the estimate in practice.

Answer: To compute the nonparametric bootstrap estimate of the skewness of the sample mean \bar{X}_n , we can follow these steps:

1. Resampling: Generate B bootstrap samples from the original data. Each bootstrap sample is created by sampling n observations from the empirical distribution of the original data with replacement. Denote the b -th bootstrap sample as $X_1^{*(b)}, X_2^{*(b)}, \dots, X_n^{*(b)}$ for $b = 1, 2, \dots, B$.
2. Compute Bootstrap Means: For each bootstrap sample, compute the sample mean $\bar{X}_n^{*(b)}$ as $\bar{X}_n^{*(b)} = \frac{1}{n} \sum_{i=1}^n X_i^{*(b)}$.
3. Estimate Moments: Calculate the empirical third central moment $\hat{\mu}_3$ and the empirical variance $\hat{\sigma}^2$ of the bootstrap means:

$$\begin{aligned}\hat{\mu}_3 &= \frac{1}{B} \sum_{b=1}^B (\bar{X}_n^{*(b)} - \bar{X}_n)^3, \\ \hat{\sigma}^2 &= \frac{1}{B} \sum_{b=1}^B (\bar{X}_n^{*(b)} - \bar{X}_n)^2,\end{aligned}$$

4. Compute Skewness: Finally, compute the bootstrap estimate of skewness:

$$\hat{\mu}_3(\bar{X}_n) = \frac{\hat{\mu}_3}{(\hat{\sigma}^2)^{3/2}}.$$

This gives us the nonparametric bootstrap estimate of the skewness of the sample mean.

Rubrics: Any valid procedure showing how to compute the bootstrap estimate of skewness will get full points.

If you managed to compute the conditional second and third moments of the bootstrap sample mean explicitly without using simulation, you also get full points.

3 points for correctly explaining the bootstrap resampling procedure.

3 points for correctly computing the empirical moments and the skewness formula.

If the main idea is correct but the implementation is flawed, give 6 points.

Part (c) [10 points]. Let $X_1, X_2, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} \text{Ber}(\theta)$ for some $\theta \in [1/2, 1]$. In such a case, a natural estimator for θ is $\hat{\theta}_n := \max\{\bar{X}_n, 1/2\}$, where $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ is the sample mean. Suppose that the true parameter is $\theta = 1/2$.

- Find the asymptotic distribution of $\sqrt{n}(\hat{\theta}_n - \theta)$.
- Let $\hat{\theta}_n^*$ be the bootstrap version of $\hat{\theta}_n$ based on a bootstrap sample drawn from the empirical distribution. Compute the conditional probability

$$\mathbb{P}(\hat{\theta}_n^* = 1/2 \mid X_1, X_2, \dots, X_n).$$

Does it converge to $\mathbb{P}(\hat{\theta}_n = 1/2)$ in probability?

Answer:

- **Asymptotic distribution:** When $\theta = 1/2$, \bar{X}_n is asymptotically normal:

$$\sqrt{n}(\bar{X}_n - 1/2) \xrightarrow{d} \mathcal{N}(0, 1/4)$$

For $\hat{\theta}_n = \max\{\bar{X}_n, 1/2\}$, we have

$$\sqrt{n}(\hat{\theta}_n - 1/2) = \max\{\sqrt{n}(\bar{X}_n - 1/2), 0\}$$

Thus, the limiting distribution is the positive part of a normal:

$$\sqrt{n}(\hat{\theta}_n - 1/2) \xrightarrow{d} \max\{Z, 0\}, \quad Z \sim \mathcal{N}(0, 1/4).$$

- **Bootstrap probability:** Let $X_1^*, X_2^*, \dots, X_n^*$ be a bootstrap sample drawn from the empirical distribution. The bootstrap estimator is $\hat{\theta}_n^* = \max\{\bar{X}_n^*, 1/2\}$, where $\bar{X}_n^* = \frac{1}{n} \sum_{i=1}^n X_i^*$. Conditionally on the data, $n\bar{X}_n^* \sim \text{Binomial}(n, \bar{X}_n)$. Therefore,

$$\begin{aligned} \mathbb{P}(\hat{\theta}_n^* = 1/2 \mid X_1, X_2, \dots, X_n) &= \mathbb{P}(\bar{X}_n^* \leq 1/2 \mid X_1, X_2, \dots, X_n) \\ &= \mathbb{P}\left(\text{Binomial}(n, \bar{X}_n) \leq n/2 \mid \bar{X}_n\right) \end{aligned} \quad (1)$$

$$\approx \Phi\left(\frac{\sqrt{n}(1/2 - \bar{X}_n)}{\sqrt{\bar{X}_n(1 - \bar{X}_n)}}\right), \quad (2)$$

where Φ is the cdf of standard normal distribution. As $n \rightarrow \infty$, $\bar{X}_n \xrightarrow{p} 1/2$ and $\sqrt{n}(1/2 - \bar{X}_n) \xrightarrow{d} \mathcal{N}(0, 1/4)$. So the argument of Φ converges in distribution to $\mathcal{N}(0, 1)$. This means that $\mathbb{P}(\hat{\theta}_n^* = 1/2 \mid X_1, X_2, \dots, X_n)$ does not converge to a constant in probability, while $\mathbb{P}(\hat{\theta}_n = 1/2) \rightarrow 1/2$ as $n \rightarrow \infty$.

Rubrics: 3 points for the first bullet point, 7 points for the second bullet point.

For the first bullet point: 2 points for correctly identifying the asymptotic distribution of \bar{X}_n .

For the second bullet point: you will get full credit as long as you get to equation (1) and explain why it does not converge in probability to a constant. 4 points for correctly deriving the expression, and 3 points for explaining the convergence behavior. For deriving the expression, you get 2 points for correctly explaining the bootstrap sampling distribution. If you use normal approximation as in (2) without deriving the exact expression, as long as the non-convergence argument is correct, you still get full credit.

Question 4. [26 points] Asymptotics of parametric estimation

Part (a) [10 points]. Let $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$ be i.i.d. samples, where $X_i \sim \mathcal{N}(0, I_d)$ are d -dimensional covariates and $Y_i \in \{0, 1\}$ are binary responses. Suppose the data are generated according to the logistic regression model:

$$\mathbb{P}(Y_i = 1 | X_i) = \frac{\exp(X_i^\top \theta^*)}{1 + \exp(X_i^\top \theta^*)}$$

for some true parameter $\theta^* \in \mathbb{R}^d$. Let $\hat{\theta}_n$ denote the maximum likelihood estimator (MLE) of θ^* based on the observed data.

- Write down the log-likelihood function for θ .
- Derive the asymptotic distribution of $\sqrt{n}(\hat{\theta}_n - \theta^*)$.

[You can express the asymptotic covariance matrix in expectations under the d -dimensional standard normal distribution.]

Answer: For each $i = 1, 2, \dots, n$, the conditional probability mass function of Y_i given X_i is:

$$\mathbb{P}(Y_i = 1 | X_i) = \left(\frac{\exp(X_i^\top \theta)}{1 + \exp(X_i^\top \theta)} \right) \quad \text{and} \quad \mathbb{P}(Y_i = 0 | X_i) = \left(\frac{1}{1 + \exp(X_i^\top \theta)} \right).$$

The log-likelihood function for θ based on a single pair (X_i, Y_i) is given by:

$$\ell(\theta; X_i, Y_i) = Y_i X_i^\top \theta - \log(1 + \exp(X_i^\top \theta)).$$

Therefore, the log-likelihood function for the entire dataset is:

$$L_n(\theta) = \sum_{i=1}^n \ell(\theta; X_i, Y_i) = \sum_{i=1}^n \left[Y_i X_i^\top \theta - \log(1 + \exp(X_i^\top \theta)) \right].$$

The score function based on a single observation is:

$$\nabla_\theta \ell(\theta; X_i, Y_i) = X_i \left(Y_i - \frac{\exp(X_i^\top \theta)}{1 + \exp(X_i^\top \theta)} \right).$$

Consequently, the Fisher information matrix $I(\theta)$ is given by:

$$I(\theta) = \mathbb{E} \left[\nabla_\theta \ell(\theta; X_i, Y_i) \nabla_\theta \ell(\theta; X_i, Y_i)^\top \right] = \mathbb{E} \left[X_i X_i^\top \cdot \frac{\exp(X_i^\top \theta)}{(1 + \exp(X_i^\top \theta))^2} \right].$$

By the asymptotic normality of MLE, we have:

$$\sqrt{n}(\hat{\theta}_n - \theta^*) \xrightarrow{d} \mathcal{N}(0, I(\theta^*)^{-1}).$$

Rubrics: Each bullet point is worth 5 points. For each bullet point, you will lose 2 points for major calculation mistakes, as long as your approach is correct.

Part (b) [8 points]. Consider a Gaussian mixture model with two components $X_1, X_2, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} \frac{1}{2}\mathcal{N}(-\theta^*, 1) + \frac{1}{2}\mathcal{N}(\theta^*, 1)$, for some $\theta^* \in [0, 1]$. Let $\hat{\theta}_n$ be the MLE for the model parameter θ^* .

- Show that the MLE $\hat{\theta}_n$ is consistent for θ^* .
- Compute the Fisher information $I(0)$ at $\theta^* = 0$.

Answer:

- **Consistency:** The log-likelihood based on data point X_i takes the form

$$\begin{aligned}\ell(\theta; X_i) &= \log \left(\frac{1}{2\sqrt{2\pi}} \left[e^{-\frac{(X_i+\theta)^2}{2}} + e^{-\frac{(X_i-\theta)^2}{2}} \right] \right) \\ &= -\frac{1}{2} \log(8\pi) - \frac{X_i^2 + \theta^2}{2} + \log(e^{X_i\theta} + e^{-X_i\theta}),\end{aligned}$$

which is continuous in θ . The parameter space $[0, 1]$ is compact. Furthermore, we have

$$\begin{aligned}\mathbb{E} \left[\sup_{\theta \in [0, 1]} |\log \ell(\theta, X_i)| \right] &\leq \sup_{\theta \in [0, 1]} \left\{ \frac{1}{2} \log(8\pi) + \frac{1 + \mathbb{E}_{\theta^*} |X_i|^2 + \theta^2}{2} + 1 + \mathbb{E}_{\theta^*} |X_i\theta| \right\} \\ &\leq 20 < +\infty.\end{aligned}$$

So the uniform law of large numbers applies. Furthermore, for any $\theta \neq \theta^*$, we have $\mathbb{P}_\theta \neq \mathbb{P}_{\theta^*}$, so the model is identifiable. By the consistency of MLE, we have $\hat{\theta}_n \xrightarrow{p} \theta^*$.

- **Fisher information at $\theta^* = 0$:** The score function is given by

$$\frac{\partial}{\partial \theta} \ell(\theta; X_i) = -\theta + X_i \cdot \frac{e^{X_i\theta} - e^{-X_i\theta}}{e^{X_i\theta} + e^{-X_i\theta}}.$$

Therefore, the Fisher information at $\theta = 0$ is given by

$$I(0) = \mathbb{E} \left[\frac{\partial}{\partial \theta} \ell(0; X_i)^2 \right] = 0.$$

Rubrics: The first bullet point is worth 5 points, the second bullet point is worth 3 points.

For the first bullet point: 2 points for correctly stating the conditions for consistency of MLE, 3 points for verifying the conditions (including 2 points for the dominance function condition and 1 point for others).

For the second bullet point: 1 point for correctly deriving the score function, 2 points for computing the Fisher information at $\theta = 0$.

Part (c) [8 points]. Under the same setting as in Part (b), suppose that we use method of moments (MoM) to estimate θ^* . The MoM estimator $\tilde{\theta}_n$ is defined as

$$\tilde{\theta}_n := \sqrt{\max\left\{\frac{1}{n} \sum_{i=1}^n X_i^2 - 1, 0\right\}}$$

Show that

$$\sup_{\theta^* \in [0,1]} \mathbb{E}_{\theta^*}[(\tilde{\theta}_n - \theta^*)^2] \leq \frac{c}{\sqrt{n}},$$

for some universal constant $c > 0$.

[Hint: you can start by computing the error for estimating $(\theta^*)^2$ first.]

Answer: Let $M_n = \frac{1}{n} \sum_{i=1}^n X_i^2$. Note that $\mathbb{E}_{\theta^*}[X_i^2] = (\theta^*)^2 + 1$, so $\mathbb{E}_{\theta^*}[M_n] = (\theta^*)^2 + 1$. The variance of M_n is given by:

$$\text{var}_{\theta^*}(M_n) = \frac{1}{n} \text{var}_{\theta^*}(X_i^2).$$

For $X_i \sim \frac{1}{2}\mathcal{N}(-\theta^*, 1) + \frac{1}{2}\mathcal{N}(\theta^*, 1)$, we compute:

$$\begin{aligned} \text{var}_{\theta^*}(X_i^2) &\leq \mathbb{E}_{\theta^*}[X_i^4] = \frac{1}{2}\mathbb{E}_{Z \sim \mathcal{N}(-\theta^*, 1)}[Z^4] + \frac{1}{2}\mathbb{E}_{Z \sim \mathcal{N}(\theta^*, 1)}[Z^4] \\ &= 3 + 6(\theta^*)^2 + (\theta^*)^4 \leq 10. \end{aligned}$$

Consequently, we have

$$\mathbb{E}_{\theta^*}[(M_n - 1 - (\theta^*)^2)^2] = \text{var}_{\theta^*}(M_n) \leq \frac{10}{n}.$$

When $M_n \geq 1$, we have $\tilde{\theta}_n^2 = M_n - 1$; when $M_n < 1$, we have $\tilde{\theta}_n = 0$, and thus $|\tilde{\theta}_n^2 - (\theta^*)^2| \leq |M_n^2 - 1 - (\theta^*)^2|$. Therefore,

$$\mathbb{E}_{\theta^*}[(\tilde{\theta}_n^2 - (\theta^*)^2)^2] \leq \mathbb{E}_{\theta^*}[(M_n - 1 - (\theta^*)^2)^2] \leq \frac{10}{n}.$$

By Cauchy–Schwarz inequality, we have

$$\begin{aligned} \mathbb{E}_{\theta^*}[(\tilde{\theta}_n - \theta^*)^2] &\leq \sqrt{\mathbb{E}_{\theta^*}[(\tilde{\theta}_n - \theta^*)^4]} \leq \sqrt{\mathbb{E}_{\theta^*}[(\tilde{\theta}_n - \theta^*)^2(\tilde{\theta}_n + \theta^*)^2]} \\ &= \sqrt{\mathbb{E}_{\theta^*}[(\tilde{\theta}_n^2 - (\theta^*)^2)^2]} \leq \frac{\sqrt{10}}{\sqrt{n}}, \end{aligned}$$

uniformly for all $\theta^* \in [0, 1]$. This completes the proof.

Rubrics: You get full points if you correctly derive the bound on the mean squared error of the MoM estimator.

4 points for correctly computing the variance of M_n .

You get 6 points if the overall idea is correct but there are some calculation mistakes.