STA355: Midterm Exam (Solutions)

Instructor: Wenlong Mou

Oct 23rd, 2026

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 $X_n \xrightarrow{a.s.} 0$ and $\mathbb{E}|X_n| < +\infty$ for every n, then $\mathbb{E}[X_n] \to 0$. **Answer:** F.

Let $U \sim \text{Unif}(0,1)$ and $X_n = n\mathbb{1}_{U \le 1/n}$. Then $X_n \xrightarrow{a.s.} 0$, but $\mathbb{E}[X_n] = 1$ for all n.

(2) If $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{d} c$ for a deterministic constant $c \in \mathbb{R}$, then $X_n Y_n \xrightarrow{d} cX$. **Answer:** T.

Convergence in distribution to a constant implies convergence in probability, so $Y_n \xrightarrow{p} c$, and Slutsky's theorem applies.

(3) If $\sqrt{n}(X_n - \mu) \xrightarrow{d} \mathcal{N}(0, \sigma^2)$ with $\sigma^2 > 0$, and $g : \mathbb{R} \to \mathbb{R}$ is continuously differentiable with $g'(\mu) = 0$, then $\sqrt{n}\big(g(X_n) - g(\mu)\big) \xrightarrow{p} 0$. **Answer:** T.

By Taylor expansion, $g(X_n) - g(\mu) = o_p(|X_n - \mu|) = o_p(1/\sqrt{n})$, since $X_n - \mu = O_p(1/\sqrt{n})$.

(4) Let $X_1, X_2, \cdots, X_n \overset{\text{i.i.d.}}{\sim} \mathbb{P}$ with $\mathbb{E}_{\mathbb{P}}[X_1^2] < +\infty$. If $\widehat{T}_n$ is an unbiased estimator of $\mu(\mathbb{P}) = \mathbb{E}_{\mathbb{P}}[X_1]$ for every n, then $\widehat{T}_n$ is consistent for $\mu(\mathbb{P})$. **Answer:** F.

$\widehat{T}_n = X_1$ is unbiased for every n, but it does not converge in probability to $\mu(\mathbb{P})$ unless X_1 is almost surely constant.

(5) Let ϕ be a level-0.05 test of $H_0 : \theta^* \in \Theta_0$ versus $H_1 : \theta^* \in \Theta_1$. If ϕ fails to reject H_0 on the observed data, then H_0 is true with probability at least 0.95. **Answer:** F.

The parameter θ^* is not random, and the level only controls the type I error. Failing to reject H_0 may simply be due to low power.

(6) Let $X_1, X_2, \cdots, X_n \overset{\text{i.i.d.}}{\sim} F$ with $\text{var}_F(X_1) = \sigma^2 \in (0, +\infty)$. Let $\widehat{se}_{\text{bootstrap},B}$ be the Monte Carlo nonparametric bootstrap estimate of the standard error of $\bar{X}_n$, based on B bootstrap replications. For any fixed $n \ge 2$, $\widehat{se}_{\text{bootstrap},B} \xrightarrow{p} \sigma/\sqrt{n}$ as $B \to \infty$. **Answer:** F.

As $B \to \infty$, it converges to the idealized bootstrap standard error $\big(\frac{1}{n^2}\sum_{i=1}^n (X_i - \bar{X}_n)^2\big)^{1/2}$, which depends on the data and differs from $\sigma/\sqrt{n}$.

(7) Let $\{p_\theta : \theta \in \Theta\}$ be an identifiable parametric model, and $X \sim p_{\theta^*}$ for some $\theta^* \in \Theta$, with $\mathbb{E}_{\theta^*}\big[|\log p_\theta(X)|\big] < +\infty$ for all $\theta \in \Theta$. Then θ^* is the unique minimizer of $\theta \mapsto -\mathbb{E}_{\theta^*}[\log p_\theta(X)]$ over Θ. **Answer:** T.

$-\mathbb{E}_{\theta^*}[\log p_\theta(X)] = -\mathbb{E}_{\theta^*}[\log p_{\theta^*}(X)] + D_{\text{KL}}(\mathbb{P}_{\theta^*} \,\|\, \mathbb{P}_\theta)$, and the KL divergence is strictly positive for $\theta \neq \theta^*$ by identifiability.

(8) In a regular one-parameter model $\{p_\theta : \theta \in \Theta\}$, the score $\frac{\partial}{\partial\theta} \log p_\theta(X)$ has mean zero and variance $I(\theta)$ under $\mathbb{P}_\theta$, where $I(\theta)$ is the Fisher information.
Answer: T.

By the score identity (Lecture 6), the score has mean zero, so its variance equals its second moment, which is the Fisher information $I(\theta)$.

Question 2. [13 points] Delta method and plug-in estimation

Throughout this question, let $X_1, X_2, \cdots, X_n \overset{\text{i.i.d.}}{\sim} \text{Poi}(\lambda)$ with unknown $\lambda > 0$, i.e., $\mathbb{P}_\lambda(X_1 = k) = e^{-\lambda}\lambda^k/k!$ for $k = 0, 1, 2, \cdots$. We want to estimate $p_0 := \mathbb{P}_\lambda(X_1 = 0) = e^{-\lambda}$. Let $\bar{X}_n$ be the sample mean, and for a fixed $\alpha \in (0, 1)$, let $z_{1-\alpha/2}$ be the $(1 - \alpha/2)$-quantile of $\mathcal{N}(0, 1)$.

Part (a) [7 points]. Let $\widehat{p}_n := e^{-\bar{X}_n}$. Find the asymptotic distribution of $\sqrt{n}(\widehat{p}_n - p_0)$, and use it to construct an asymptotic $(1 - \alpha)$ confidence interval for p_0 that is computable from the data.

Answer: Since $\mathbb{E}_\lambda[X_1] = \text{var}_\lambda(X_1) = \lambda$, the central limit theorem gives

$$\sqrt{n}(\bar{X}_n - \lambda) \xrightarrow{d} \mathcal{N}(0, \lambda).$$

Let $g(x) = e^{-x}$, which is continuously differentiable with $g'(\lambda) = -e^{-\lambda} \neq 0$. By the Delta method,

$$\sqrt{n}(\widehat{p}_n - p_0) = \sqrt{n}\big(g(\bar{X}_n) - g(\lambda)\big) \xrightarrow{d} \mathcal{N}\big(0, \lambda e^{-2\lambda}\big).$$

Confidence interval: by the law of large numbers, $\bar{X}_n \xrightarrow{p} \lambda$, so $\widehat{\sigma}_n := \sqrt{\bar{X}_n} e^{-\bar{X}_n} \xrightarrow{p} \sqrt{\lambda} e^{-\lambda} > 0$. By Slutsky's theorem, $\sqrt{n}(\widehat{p}_n - p_0)/\widehat{\sigma}_n \xrightarrow{d} \mathcal{N}(0, 1)$. Therefore, the interval

$$C_n = \left[\widehat{p}_n - z_{1-\alpha/2} \cdot \frac{\sqrt{\bar{X}_n} e^{-\bar{X}_n}}{\sqrt{n}},\ \widehat{p}_n + z_{1-\alpha/2} \cdot \frac{\sqrt{\bar{X}_n} e^{-\bar{X}_n}}{\sqrt{n}}\right]$$

satisfies $\mathbb{P}_\lambda(C_n \ni p_0) \to 1 - \alpha$ as $n \to \infty$. (The event $\{\bar{X}_n = 0\}$, on which $\widehat{\sigma}_n = 0$, has probability $e^{-n\lambda} \to 0$, so it does not affect the asymptotic coverage.)

Rubrics: 4 points for the asymptotic distribution: 1 point for the CLT with the correct variance λ, 2 points for correctly applying the Delta method (including $g'(\lambda) = -e^{-\lambda}$), and 1 point for the final variance $\lambda e^{-2\lambda}$.

3 points for the confidence interval: 2 points for a correct interval, and 1 point for justifying the plug-in standard error (law of large numbers and Slutsky's theorem). Using the continuity of $x \mapsto \sqrt{x} e^{-x}$ (continuous mapping) without proof is fine. If the width of the interval depends on the unknown λ, give at most 1 point.

Any other asymptotically valid interval with correct justification also gets full points, e.g., mapping the interval $\big[\bar{X}_n \pm z_{1-\alpha/2}\sqrt{\bar{X}_n/n}\big]$ for λ through the decreasing function $x \mapsto e^{-x}$.

Part (b) [6 points]. Let $\widetilde{p}_n := \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{X_i=0}$ be the empirical frequency of zeros (i.e., the empirical cdf $\widehat{F}_n(0)$). Find the asymptotic distribution of $\sqrt{n}(\widetilde{p}_n - p_0)$, and show that its asymptotic variance is strictly larger than the one in Part (a), for every $\lambda > 0$.

Answer: The indicators $\mathbb{1}_{X_i=0}$ are i.i.d. $\mathrm{Ber}(p_0)$, with mean p_0 and variance $p_0(1-p_0)$. By the CLT,

$$\sqrt{n}(\widetilde{p}_n - p_0) \xrightarrow{d} \mathcal{N}\big(0, e^{-\lambda}(1-e^{-\lambda})\big).$$

Comparing with Part (a), we have

$$\frac{e^{-\lambda}(1-e^{-\lambda})}{\lambda e^{-2\lambda}} = \frac{e^{\lambda}-1}{\lambda} > 1, \quad \text{since } e^{\lambda} = 1 + \lambda + \frac{\lambda^2}{2} + \cdots > 1 + \lambda \text{ for } \lambda > 0.$$

So $\widetilde{p}_n$ has strictly larger asymptotic variance. When the Poisson model is correct, the model-based estimator $\widehat{p}_n$ is more efficient.

Remark: the price of this efficiency gain is that $\widehat{p}_n$ relies on the model being correct. If the data come from a distribution $\mathbb{P}$ on $\{0, 1, 2, \cdots\}$ with mean μ that is not Poisson, then $\widehat{p}_n \xrightarrow{p} e^{-\mu}$, which in general differs from $\mathbb{P}(X_1 = 0)$ (e.g., $\mathbb{P} = \mathrm{Ber}(1/2)$). On the other hand, $\widetilde{p}_n \xrightarrow{p} \mathbb{P}(X_1 = 0)$ for every such $\mathbb{P}$.

Rubrics: 3 points for the asymptotic distribution.

3 points for the comparison: 1 point for the ratio (or difference) of the two asymptotic variances, and 2 points for proving the inequality.

Question 3. [9 points] Plug-in estimation of a collision probability

Let $X_1, \cdots, X_n$ be i.i.d. from a distribution on the finite set $\{1, \cdots, K\}$, where $K \geq 2$ is fixed and $\pi_k := \mathbb{P}(X_1 = k) > 0$ for every k. The probability that two independent draws from this distribution coincide is $\theta := \sum_{k=1}^K \pi_k^2$. Let $\widehat{\pi}_k := \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{X_i=k}$, and consider the plug-in estimator $\widehat{\theta}_n := \sum_{k=1}^K \widehat{\pi}_k^2$.

- Show that $\sqrt{n}(\widehat{\theta}_n - \theta) \xrightarrow{d} \mathcal{N}(0, v)$, where $v = 4\big(\sum_{k=1}^K \pi_k^3 - \theta^2\big)$.
- Show that $v = 0$ if and only if $\pi_1 = \pi_2 = \cdots = \pi_K$. In that case, show that $\widehat{\theta}_n - \theta = O_p(1/n)$.

[Hint: $\sum_{k=1}^K \pi_k \widehat{\pi}_k = \frac{1}{n}\sum_{i=1}^n \pi_{X_i}$.]

Answer:

- We decompose

$$\widehat{\theta}_n - \theta = \sum_{k=1}^K (\widehat{\pi}_k - \pi_k)^2 + 2\sum_{k=1}^K \pi_k(\widehat{\pi}_k - \pi_k).$$

For each k, the CLT gives $\sqrt{n}(\widehat{\pi}_k - \pi_k) = O_p(1)$, so $(\widehat{\pi}_k - \pi_k)^2 = O_p(1/n)$. Since K is fixed, the first term is $O_p(1/n)$, and $\sqrt{n}$ times it converges to 0 in probability. For the second term, by the hint,

$$2\sum_{k=1}^K \pi_k(\widehat{\pi}_k - \pi_k) = 2\Big(\frac{1}{n}\sum_{i=1}^n \pi_{X_i} - \theta\Big).$$

The variables π_{X_i} are i.i.d., with mean $\mathbb{E}[\pi_{X_1}] = \sum_k \pi_k \cdot \pi_k = \theta$ and variance $\mathbb{E}[\pi_{X_1}^2] - \theta^2 = \sum_k \pi_k^3 - \theta^2$. By the CLT and Slutsky's theorem,

$$\sqrt{n}(\widehat{\theta}_n - \theta) \xrightarrow{d} \mathcal{N}\Big(0, 4\big(\textstyle\sum_k \pi_k^3 - \theta^2\big)\Big).$$

(Alternatively, apply the multivariate CLT to the vector $(\widehat{\pi}_1, \cdots, \widehat{\pi}_K)$, whose covariance is $\mathrm{diag}(\pi) - \pi\pi^\top$, and the Delta method to $g(x) = \sum_k x_k^2$ with gradient 2π. This gives $4\pi^\top(\mathrm{diag}(\pi) - \pi\pi^\top)\pi = 4(\sum_k \pi_k^3 - \theta^2)$. Since this covariance matrix is singular, one can also work with the first $K-1$ coordinates, using $\widehat{\pi}_K = 1 - \sum_{k<K} \widehat{\pi}_k$, which gives the same answer.)

- We have $v = 4\,\mathrm{var}(\pi_{X_1})$, which is 0 if and only if π_{X_1} is almost surely constant. Since every value k has positive probability, this means $\pi_1 = \cdots = \pi_K = 1/K$. In that case $\theta = 1/K$, and since $\sum_k \widehat{\pi}_k = 1$, the second term in the decomposition vanishes:

$$\widehat{\theta}_n - \theta = \sum_{k=1}^K \Big(\widehat{\pi}_k - \frac{1}{K}\Big)^2 = O_p(1/n).$$

So the first-order Delta method gives a degenerate limit, and the estimation error is of the smaller order $1/n$.

Rubrics: 6 points for the first bullet point: 2 points for the decomposition, 1 point for showing that the quadratic term is negligible, 2 points for the CLT with the correct variance, and 1 point for the conclusion via Slutsky's theorem. For the multivariate Delta method route: 2 points for the multinomial covariance, 1 point for the gradient, 2 points for the quadratic form and its simplification, and 1 point for the conclusion. Ignoring the covariances between the $\widehat{\pi}_k$ (which gives the incorrect variance $4\sum_k(\pi_k^3 - \pi_k^4)$) gets at most 3 points.

3 points for the second bullet point: 2 points for characterizing $v = 0$, and 1 point for the $O_p(1/n)$ rate.

Question 4. [27 points] The bootstrap

Part (a) [7 points]. Let $X_1, X_2, \cdots, X_n$ be i.i.d. from an unknown distribution F on $\mathbb{R}$, and let $\widehat{M}_n$ be the sample median, which estimates the median of F. Describe, step by step, how to use the nonparametric bootstrap with B Monte Carlo replications to compute (i) an estimate of the standard error of $\widehat{M}_n$, and (ii) an approximate $(1-\alpha)$ confidence interval for the median of F based on the normal approximation. Make clear which quantities are computed from the original data, and which from the bootstrap samples.

Answer:

1. From the original data, compute the sample median $\widehat{M}_n$.
2. For each $b = 1, \cdots, B$, independently: draw a bootstrap sample $X^*_{b,1}, \cdots, X^*_{b,n}$ by sampling n times with replacement from $\{X_1, \cdots, X_n\}$ (i.e., i.i.d. from the empirical distribution $\widehat{F}_n$), and compute its median $\widehat{M}^{*,b}_n$.
3. Estimate the standard error by the standard deviation of the bootstrap medians:
$$\widehat{se}_{\text{bootstrap},B} = \sqrt{\frac{1}{B}\sum_{b=1}^{B}\left(\widehat{M}^{*,b}_n - \frac{1}{B}\sum_{j=1}^{B}\widehat{M}^{*,j}_n\right)^2}.$$
4. The normal interval is centered at the estimate from the original data:
$$\left[\widehat{M}_n - z_{1-\alpha/2}\cdot\widehat{se}_{\text{bootstrap},B},\ \widehat{M}_n + z_{1-\alpha/2}\cdot\widehat{se}_{\text{bootstrap},B}\right],$$
where $z_{1-\alpha/2}$ is the $(1-\alpha/2)$-quantile of $\mathcal{N}(0,1)$.

Only $\widehat{M}_n$ is computed from the original sample; the medians $\widehat{M}^{*,1}_n, \cdots, \widehat{M}^{*,B}_n$ and $\widehat{se}_{\text{bootstrap},B}$ come from the bootstrap samples. Given the data, the only randomness is the resampling, and B should be large (e.g., $B \geq 1000$), so that the Monte Carlo error is small.

Rubrics: 2 points for the resampling step (samples of size n, drawn with replacement from the data, independently for each b), 1 point for computing the median of each bootstrap sample, 2 points for the standard error formula (normalizing by B or by $B-1$ are both fine), 1 point for the normal interval centered at $\widehat{M}_n$, and 1 point for clearly distinguishing what is computed from the original data and what from the bootstrap samples.

If the interval is centered at the average of the bootstrap medians instead of $\widehat{M}_n$, or a percentile interval is given instead of the normal interval, deduct 1 point.

Part (b) [9 points]. An analyst models positive data $X_1, X_2, \cdots, X_n$ as i.i.d. exponential with mean $\mu > 0$ (density $\mu^{-1}e^{-x/\mu}$ for $x > 0$, and variance μ^2), and estimates μ by $\bar{X}_n$.

- Find the parametric bootstrap estimate (resampling from the exponential distribution with mean $\bar{X}_n$) and the nonparametric bootstrap estimate of the standard error of $\bar{X}_n$. (Here the idealized bootstrap procedures admit closed-form expressions, so no simulation is needed.)

- Suppose that the data actually come i.i.d. from a distribution on $(0, +\infty)$ with mean μ and variance $\sigma^2 \in (0, +\infty)$, which is *not* necessarily exponential. Let $\widehat{se}_{\text{par}}$ be the parametric bootstrap estimate from the previous bullet point. For a fixed $\alpha \in (0, 1)$, find

$$\lim_{n\to\infty} \mathbb{P}\Big(\mu \in \big[\bar{X}_n - z_{1-\alpha/2} \cdot \widehat{se}_{\text{par}},\ \bar{X}_n + z_{1-\alpha/2} \cdot \widehat{se}_{\text{par}}\big]\Big),$$

  where $z_{1-\alpha/2}$ is the $(1-\alpha/2)$-quantile of $\mathcal{N}(0,1)$. When does this interval under-cover?

Answer:

- **Parametric bootstrap:** conditionally on the data, $X_1^*, X_2^*, \cdots, X_n^*$ are i.i.d. from the exponential distribution with mean $\bar{X}_n$, whose variance is $\bar{X}_n^2$. So the parametric bootstrap estimate of standard error is

$$\widehat{se}_{\text{par}} = \sqrt{\text{var}\left(\frac{1}{n}\sum_{i=1}^n X_i^* \mid X_1, X_2, \cdots, X_n\right)} = \sqrt{\frac{\bar{X}_n^2}{n}} = \frac{\bar{X}_n}{\sqrt{n}}.$$

  Nonparametric bootstrap: conditionally on the data, $X_1^*, X_2^*, \cdots, X_n^*$ are i.i.d. from the empirical distribution $\widehat{F}_n$, whose variance is $S_n^2 := \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X}_n)^2$. So the nonparametric bootstrap estimate of standard error is

$$\widehat{se}_{\text{np}} = \sqrt{\text{var}\left(\frac{1}{n}\sum_{i=1}^n X_i^* \mid X_1, X_2, \cdots, X_n\right)} = \frac{S_n}{\sqrt{n}}.$$

- By the CLT, $\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} \mathcal{N}(0, \sigma^2)$, and by the law of large numbers, $\bar{X}_n \xrightarrow{p} \mu > 0$. Slutsky's theorem gives $\sqrt{n}(\bar{X}_n - \mu)/\bar{X}_n \xrightarrow{d} \mathcal{N}(0, \sigma^2/\mu^2)$, so

$$\mathbb{P}\big(|\bar{X}_n - \mu| \le z_{1-\alpha/2} \cdot \widehat{se}_{\text{par}}\big) \to \mathbb{P}\Big(|Z| \le z_{1-\alpha/2} \cdot \frac{\mu}{\sigma}\Big) = 2\Phi\Big(z_{1-\alpha/2} \cdot \frac{\mu}{\sigma}\Big) - 1,$$

  where $Z \sim \mathcal{N}(0,1)$ and Φ is its cdf. This limit equals $1-\alpha$ if and only if $\sigma = \mu$, which holds under the exponential model. The interval based on the

parametric bootstrap under-covers if and only if $\sigma > \mu$, i.e., when the data are more dispersed than the exponential model allows. (It over-covers when $\sigma < \mu$.)

For comparison, since $S_n \xrightarrow{p} \sigma$, Slutsky's theorem gives $\sqrt{n}(\bar{X}_n - \mu)/S_n \xrightarrow{d} \mathcal{N}(0,1)$, so the interval based on the nonparametric bootstrap has asymptotic coverage exactly $1-\alpha$, whether or not the exponential model is correct.

Rubrics: 4 points for the first bullet point, 2 points for each standard error. For each one, if you only give a final formula without explaining the bootstrap distribution, give 1 point.

5 points for the second bullet point: 4 points for the limit $2\Phi(z_{1-\alpha/2}\mu/\sigma) - 1$ (with the CLT, the law of large numbers and Slutsky's theorem correctly used), and 1 point for the condition $\sigma > \mu$.

Part (c) [11 points]. Let $X_1, X_2, \cdots, X_n \overset{\text{i.i.d.}}{\sim} \text{Unif}([0,\theta])$ with unknown $\theta > 0$, and let $\widehat{\theta}_n = \max_i X_i$ be the MLE. The nonparametric bootstrap fails for $\widehat{\theta}_n$ (Lecture 4), so consider the parametric bootstrap: given the data, draw $X_1^*, \cdots, X_n^* \overset{\text{i.i.d.}}{\sim} \text{Unif}([0,\widehat{\theta}_n])$ and let $\widehat{\theta}_n^* = \max_i X_i^*$. For a fixed $\alpha \in (0,1)$, recall the (idealized) percentile and pivotal intervals for θ

$$C_n^{\text{perc}} = \big[\widehat{\theta}^*_{(\alpha/2)},\ \widehat{\theta}^*_{(1-\alpha/2)}\big], \qquad C_n^{\text{piv}} = \big[\widehat{\theta}_n - \widehat{H}^{-1}(1-\alpha/2),\ \widehat{\theta}_n - \widehat{H}^{-1}(\alpha/2)\big],$$

where $\widehat{\theta}^*_{(p)}$ is the p-quantile of the conditional distribution of $\widehat{\theta}_n^*$, and $\widehat{H}$ is the conditional cdf of $R_n^* = \widehat{\theta}_n^* - \widehat{\theta}_n$, both given the data.

- Compute the percentile interval C_n^{perc}, and show that $\mathbb{P}_\theta(C_n^{\text{perc}} \ni \theta) = 0$ for every n.
- Compute the pivotal interval C_n^{piv}, and show that $\mathbb{P}_\theta(C_n^{\text{piv}} \ni \theta) \to 1-\alpha$ as $n \to \infty$.

[Hint: first find the conditional cdf of $\widehat{\theta}_n^*$. You may use that $(2 - p^{1/n})^{-n} \to p$ as $n \to \infty$, for any $p \in (0,1)$.]

Answer: Conditionally on the data, $X_1^*, X_2^*, \cdots, X_n^*$ are i.i.d. $\text{Unif}([0,\widehat{\theta}_n])$, so

$$\mathbb{P}\big(\widehat{\theta}_n^* \le t \mid X_1, \cdots, X_n\big) = \prod_{i=1}^n \mathbb{P}\big(X_i^* \le t \mid X_1, \cdots, X_n\big) = \Big(\frac{t}{\widehat{\theta}_n}\Big)^n, \qquad 0 \le t \le \widehat{\theta}_n.$$

This cdf is continuous and strictly increasing on $[0,\widehat{\theta}_n]$, so its p-quantile is $\widehat{\theta}^*_{(p)} = \widehat{\theta}_n \cdot p^{1/n}$. Let $a_n := (1-\alpha/2)^{1/n}$ and $b_n := (\alpha/2)^{1/n}$, so that $0 < b_n < a_n < 1$.

- **Percentile interval:** $C_n^{\text{perc}} = \big[\widehat{\theta}_n \cdot b_n,\ \widehat{\theta}_n \cdot a_n\big]$. With probability one, we have $0 < \widehat{\theta}_n < \theta$, so the upper endpoint satisfies $\widehat{\theta}_n \cdot a_n < \widehat{\theta}_n < \theta$. Hence $\theta \notin C_n^{\text{perc}}$ almost surely, and the coverage probability is 0 for every n. (The whole bootstrap distribution lies below $\widehat{\theta}_n$, which itself always underestimates θ. The percentile interval inherits this bias.)
- **Pivotal interval:** since $R_n^* = \widehat{\theta}_n^* - \widehat{\theta}_n$, its conditional p-quantile is $\widehat{H}^{-1}(p) = \widehat{\theta}_n(p^{1/n} - 1)$. Therefore,

  $$C_n^{\text{piv}} = \big[\widehat{\theta}_n - \widehat{\theta}_n(a_n - 1),\ \widehat{\theta}_n - \widehat{\theta}_n(b_n - 1)\big] = \big[\widehat{\theta}_n(2 - a_n),\ \widehat{\theta}_n(2 - b_n)\big].$$

  In the real world, $V_n := \widehat{\theta}_n/\theta$ satisfies $\mathbb{P}_\theta(V_n \le v) = v^n$ for $v \in [0,1]$. Since $1 < 2 - a_n < 2 - b_n$, both $1/(2-a_n)$ and $1/(2-b_n)$ lie in $(0,1)$, and

  $$\mathbb{P}_\theta\big(C_n^{\text{piv}} \ni \theta\big) = \mathbb{P}_\theta\Big(\frac{1}{2-b_n} \le V_n \le \frac{1}{2-a_n}\Big) = (2-a_n)^{-n} - (2-b_n)^{-n}.$$

  By the hint, $(2-a_n)^{-n} \to 1 - \alpha/2$ and $(2-b_n)^{-n} \to \alpha/2$. So the coverage probability converges to $1-\alpha$.

Remark: the pivotal interval reflects the bootstrap distribution around $\widehat{\theta}_n$, which corrects the downward bias of $\widehat{\theta}_n$. Using the exact pivot $\widehat{\theta}_n/\theta$ instead of $\widehat{\theta}_n - \theta$ gives the interval $\left[\widehat{\theta}_n(1-\alpha/2)^{-1/n},\ \widehat{\theta}_n(\alpha/2)^{-1/n}\right]$, whose coverage probability is exactly $1-\alpha$.

Rubrics: 2 points for the conditional cdf and quantiles of $\widehat{\theta}_n^*$.

4 points for the first bullet point: 2 points for the percentile interval, and 2 points for the zero-coverage argument.

5 points for the second bullet point: 2 points for the pivotal interval, 2 points for the exact coverage probability (using the distribution of $\widehat{\theta}_n/\theta$), and 1 point for the limit.

Question 5. [9 points] Asymptotics of an M-estimator

Let $X_1, X_2, \cdots, X_n$ be i.i.d. with $X_i = \theta^* + U_i$, where $U_i \sim \text{Unif}([-1,1])$ and $\theta^* \in \mathbb{R}$ is unknown. Consider the M-estimator with the quartic loss

$$\widetilde{\theta}_n \in \arg\min_{\theta\in\mathbb{R}} \frac{1}{n}\sum_{i=1}^{n}(X_i-\theta)^4.$$

- Show that θ^* is the unique minimizer of the population risk $R(\theta) := \mathbb{E}[(X_1-\theta)^4]$.
- Assuming that $\widetilde{\theta}_n$ is consistent, use the asymptotic normality theorem for M-estimators to find the asymptotic distribution of $\sqrt{n}(\widetilde{\theta}_n - \theta^*)$. Compare its asymptotic variance with that of the sample mean $\bar{X}_n$.

[You may use $\mathbb{E}[U_1^2] = 1/3$ and $\mathbb{E}[U_1^6] = 1/7$.]

Answer:

- Write $d = \theta - \theta^*$, so that $X_1 - \theta = U_1 - d$. Since U_1 is symmetric about 0, its odd moments vanish, and
$$R(\theta) = \mathbb{E}[(U_1-d)^4] = \mathbb{E}[U_1^4] + 6d^2\mathbb{E}[U_1^2] + d^4 = \mathbb{E}[U_1^4] + 2d^2 + d^4,$$
which is uniquely minimized at $d = 0$, i.e., at $\theta = \theta^*$.
- The loss $f(\theta, x) = (x-\theta)^4$ is twice continuously differentiable in θ, with
$$\frac{\partial}{\partial\theta}f(\theta,x) = -4(x-\theta)^3, \qquad \frac{\partial^2}{\partial\theta^2}f(\theta,x) = 12(x-\theta)^2.$$
At $\theta = \theta^*$, we have $X_1 - \theta^* = U_1$, so
$$\Sigma^* = \mathbb{E}\big[16U_1^6\big] = \frac{16}{7}, \qquad R''(\theta^*) = \mathbb{E}\big[12U_1^2\big] = 4 > 0.$$
Since $\widetilde{\theta}_n$ is consistent and θ^* is an interior point, the asymptotic normality theorem for M-estimators gives
$$\sqrt{n}(\widetilde{\theta}_n - \theta^*) \xrightarrow{d} \mathcal{N}\Big(0, \frac{\Sigma^*}{R''(\theta^*)^2}\Big) = \mathcal{N}\Big(0, \frac{16/7}{16}\Big) = \mathcal{N}\Big(0, \frac{1}{7}\Big).$$
By the CLT, $\sqrt{n}(\bar{X}_n - \theta^*) \xrightarrow{d} \mathcal{N}(0, \text{var}(U_1)) = \mathcal{N}(0, 1/3)$. Since $1/7 < 1/3$, the quartic M-estimator is asymptotically more efficient than the sample mean here: the quartic loss puts more weight on extreme observations, which helps when the noise is bounded. (For Gaussian noise $U \sim \mathcal{N}(0,\sigma^2)$, the same formula gives $\mathbb{E}[U^6]/(9(\mathbb{E}[U^2])^2) = 15\sigma^6/(9\sigma^4) = 5\sigma^2/3 > \sigma^2$, so the quartic loss would be worse than the sample mean.)

Rubrics: 3 points for the first bullet point (any valid argument works, e.g., the strict convexity of R together with $R'(\theta^*) = -4\mathbb{E}[U_1^3] = 0$).

6 points for the second bullet point: 1 point for the derivatives of the loss, 1 point for checking the conditions of the theorem (twice continuously differentiable loss, interior θ^*, $R''(\theta^*) > 0$), 2 points for $\Sigma^* = 16/7$ and $R''(\theta^*) = 4$, 1 point for the limit $\mathcal{N}(0, 1/7)$, and 1 point for the comparison with $\bar{X}_n$.

Question 6. [18 points] Estimation in the Cauchy location model

Throughout this question, let $X_1, X_2, \cdots, X_n \overset{\text{i.i.d.}}{\sim} p_{\theta^*}$ from the Cauchy location family, with density and cdf

$$p_\theta(x) = \frac{1}{\pi\big[1+(x-\theta)^2\big]}, \qquad F_\theta(x) = \frac{1}{2} + \frac{1}{\pi}\arctan(x-\theta), \qquad x \in \mathbb{R}.$$

The parameter θ^* is unknown, but lies in $(-R, R)$ for a known constant $R > 0$.

Part (a) [6 points].

- Is the sample mean $\bar{X}_n$ a consistent estimator of θ^*? Justify your answer. (You may use without proof that $\bar{X}_n$ has the same distribution as X_1 for every n.)
- Show that the sample median $\widehat{m}_n := \widehat{F}_n^{-1}(1/2)$ is a consistent estimator of θ^*.

Answer:

- No. For any $\varepsilon > 0$ and every n, we have

$$\mathbb{P}\big(|\bar{X}_n - \theta^*| > \varepsilon\big) = \mathbb{P}\big(|X_1 - \theta^*| > \varepsilon\big) = 1 - \frac{2}{\pi}\arctan(\varepsilon) > 0,$$

  which does not depend on n, and hence does not converge to 0. (The law of large numbers does not apply, since $\mathbb{E}|X_1| = +\infty$.)

- The cdf $F_{\theta^*}(x) = \frac{1}{2} + \frac{1}{\pi}\arctan(x-\theta^*)$ is continuous and strictly increasing on $\mathbb{R}$, with $F_{\theta^*}(\theta^*) = 1/2$. So $\theta^* = F_{\theta^*}^{-1}(1/2)$ is the population median, and F_{θ^*} is continuous and strictly increasing at θ^*. By the consistency theorem for sample quantiles (Lecture 3), we have $\widehat{m}_n = \widehat{F}_n^{-1}(1/2) \xrightarrow{p} \theta^*$.

Rubrics: 3 points for each bullet point.

For the first bullet point: 1 point for the correct answer, and 2 points for the justification.

For the second bullet point: 2 points for verifying the assumptions of the quantile consistency theorem, and 1 point for the conclusion. A direct proof (e.g., using the Glivenko–Cantelli theorem as in Lecture 3) also gets full points.

Part (b) [12 points]. Let $\widehat{\theta}_n$ be the MLE over $\Theta = [-R, R]$.

- Show that $\widehat{\theta}_n$ is a consistent estimator of θ^*.
- Compute the Fisher information $I(\theta)$, and find the asymptotic distribution of $\sqrt{n}(\widehat{\theta}_n - \theta^*)$.

[You may use $\int_{-\infty}^{+\infty} \frac{u^2}{(1+u^2)^3} du = \frac{\pi}{8}$, and $\log(1+u^2) \le 2\sqrt{|u|}$ for all $u \in \mathbb{R}$.]

Answer:

- We verify the conditions for consistency of the MLE (Lecture 6).
  - The parameter space $\Theta = [-R, R]$ is compact, and $\theta \mapsto p_\theta(x)$ is continuous for every $x \in \mathbb{R}$.
  - Identifiability: the density p_θ is uniquely maximized at $x = \theta$, so $\theta_1 \neq \theta_2$ implies $p_{\theta_1} \neq p_{\theta_2}$.
  - Integrability: $\log p_\theta(x) = -\log \pi - \log\big(1 + (x-\theta)^2\big)$. For $\theta \in [-R, R]$, we have $|x - \theta| \le |x| + R$. Using the inequality $\log(1+u^2) \le 2\sqrt{|u|}$ with $u = |x| + R$, we have

$$\begin{aligned}\sup_{\theta \in \Theta} |\log p_\theta(x)| &\le \log \pi + \log\big(1 + (|x| + R)^2\big) \\ &\le \log \pi + 2\sqrt{|x| + R} \le \log \pi + 2\sqrt{R} + 2\sqrt{|x|}.\end{aligned}$$

  (The inequality follows from $1 + u^2 \le (1 + |u|)^2$ and $\log(1+y) \le \sqrt{y}$ for $y \ge 0$.) Moreover, we have

$$\mathbb{E}_{\theta^*}\big[\sqrt{|X|}\big] = \int_{-\infty}^{+\infty} \frac{\sqrt{|x|}}{\pi[1 + (x - \theta^*)^2]} dx < +\infty,$$

  since the integrand decays like $|x|^{-3/2}$ as $|x| \to \infty$. Hence the integrability condition holds:

$$\mathbb{E}_{\theta^*}\Big[\sup_{\theta \in \Theta} |\log p_\theta(X)|\Big] < +\infty.$$

  By the consistency theorem for the MLE, we have $\widehat{\theta}_n \xrightarrow{p} \theta^*$.

- The score function is

$$\frac{\partial}{\partial \theta} \log p_\theta(x) = \frac{2(x - \theta)}{1 + (x - \theta)^2}.$$

  By the change of variable $u = x - \theta$, for every $\theta \in \mathbb{R}$, we have

$$I(\theta) = \mathbb{E}_\theta\Big[\Big(\frac{2(X-\theta)}{1 + (X-\theta)^2}\Big)^2\Big] = \frac{4}{\pi} \int_{-\infty}^{+\infty} \frac{u^2}{(1+u^2)^3} du = \frac{4}{\pi} \cdot \frac{\pi}{8} = \frac{1}{2}.$$

Since θ^* lies in the interior of Θ and $\widehat{\theta}_n$ is consistent, with probability tending to one, $\widehat{\theta}_n$ is an interior maximizer, and hence solves the score equation $\frac{1}{n}\sum_{i=1}^{n} \frac{\partial}{\partial\theta} \log p_\theta(X_i) = 0$. The log-likelihood is twice continuously differentiable in θ, and $I(\theta^*) = 1/2 > 0$. By the asymptotic normality of the MLE,

$$\sqrt{n}(\widehat{\theta}_n - \theta^*) \xrightarrow{d} \mathcal{N}\big(0, I(\theta^*)^{-1}\big) = \mathcal{N}(0, 2).$$

Rubrics: 6 points for the first bullet point: 1 point for stating the conditions of the consistency theorem, 1 point for compactness and continuity, 1 point for identifiability, and 3 points for the integrability condition (any valid integrable upper bound works). A bound that grows linearly in $|x|$ is not integrable, since $\mathbb{E}|X| = +\infty$, and gets at most 1 of these 3 points.

6 points for the second bullet point: 1 point for the score function, 3 points for $I(\theta) = 1/2$, and 2 points for the asymptotic distribution with a brief justification. Any correct computation of $I(\theta)$ gets full points.