STA355: Midterm Exam

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 $\mathbb{E}[X_n] \to 0$.

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

(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 $\sqrt{n}\big(g(X_n) - g(\mu)\big) \xrightarrow{p} 0$.

(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 $\mu(\mathbb{P})$.

(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.

(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 $n \geq 2$, $\widehat{se}_{\text{bootstrap},B} \xrightarrow{p} \sigma/\sqrt{n}$ as $B \to \infty$.

(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.

(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.

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.

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 $\lambda > 0$.

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

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.

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?

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 $p \in (0,1)$.]

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 $\mathbb{E}[U_1^6] = 1/7$.]

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^*.

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 $\log(1+u^2) \leq 2\sqrt{|u|}$ for all $u \in \mathbb{R}$.]