WPI Mathematical Sciences Ph.D. General Comprehensive Exam

MA 541 Probability and Mathematical Statistics II

May 2026

Note: Please show clear, logical reasoning for each solution. Even if you cannot solve a problem completely, clearly outline your approach. For example, if you run out of time before finishing a calculation, demonstrate the steps you would take to complete it and, if applicable, how you would use the result in subsequent steps.

Problem 1. Let $X_1, \ldots, X_n \overset{iid}{\sim}$ Bernoulli(p). Let $\bar{X}$ denote the sample mean.

(a) Is $\bar{X}^2$ an unbiased estimator of p^2? Explain.

(b) Find the best unbiased estimator (BUE), T, of p^3.

(c) Within the class of all estimators of p^3, what can you say about T in (b)?

Problem 2. Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim}$ Bernoulli(p). Consider the prior distribution: $p \sim$ Beta(a, b) with known constants $a, b > 0$. Find the Bayes estimator for p, and write it into a format of weighted sum of data information (the sample mean) and the prior information (the prior mean). Note the kernel function of Beta(a, b) is $f(p) = C(a, b)\, p^{a-1}(1-p)^{b-1}$.

Problem 3. Let $X_1, \ldots, X_n$ be i.i.d. with density

$$f_\theta(x) = \frac{1}{2} e^{-|x-\theta|}, \qquad x \in \mathbb{R},$$

where $\theta \in \mathbb{R}$ is unknown.

(1) Show that the sample median is a maximum likelihood estimator (MLE) of θ, and derive the asymptotic distribution of the sample median $\widehat{\theta}_n$.

Hint. Consider the following fact: Let $X_1, \ldots, X_n$ be i.i.d. from a continuous distribution with median θ, and suppose the density f exists and is continuous with $f(\theta) > 0$. Let $\widehat{\theta}_n$ be the sample median. Then

$$\sqrt{n}(\widehat{\theta}_n - \theta) \xrightarrow{d} N\left(0, \frac{1}{4f(\theta)^2}\right), \qquad n \to \infty.$$

(2) Consider testing

$$H_0 : \theta = 0 \quad \text{versus} \quad H_1 : \theta > 0.$$

Derive an asymptotic level-α test based on the sample median $\widehat{\theta}_n$.

(3) Now consider local alternative

$$\theta = \frac{h}{\sqrt{n}}, \qquad h > 0.$$

Find the limiting power of the test in Part (2), as $n \to \infty$.

Problem 4. Let $X_1, \ldots, X_n$ be i.i.d. random variables with density

$$f(x;\theta) = e^{-(x-\theta)}, \qquad x > \theta,$$

where $\theta \in \mathbb{R}$.
Consider testing

$$H_0 : \theta \leq 0 \qquad \text{versus} \qquad H_1 : \theta > 0.$$

(a) Find the uniformly most powerful level-α test. Show all the steps clearly.

(b) Derive the power function of the test. Is the test unbiased? Justify your answer.

Problem 5. Say you are studying a finite population of size N, which is composed of smokers and non-smokers. Let X be the number of smokers in a simple random sample (with replacement) of size n. If there are S smokers in the population, then X is drawn from the hypergeometric distribution with probability mass function

$$Pr_S(X = x) = \frac{\binom{S}{x}\binom{N-S}{n-x}}{\binom{N}{n}} \quad \text{if } max(0, n+S-N) \leq x \leq min(n, S) \text{ and 0 otherwise}$$

(a) Show that there exists a uniformly most powerful test of $H_0 : S \leq s_0$ versus $H_A : S > s_0$, and explain how you would construct it for a given level α.

(b) How would you use the test from part A, along with an analogous test of $H_0' : S \geq s_1$ versus $H_A' : S < s_1$, to construct a $1 - \alpha$ confidence interval for S.

Problem 6. Let $X_1, ..., X_n$ be i.i.d. random variables from $N(\mu, 1)$, and let $\theta = P(X < c)$, where c is a known constant.

(a) Find the MLE of θ and derive its asymptotic distribution. You may leave your answer in terms of the standard normal CDF.

(b) Show that $\tilde{\theta} = \sum_{i=1}^{n} I(X_i < c)/n$ is a consistent estimator of θ. Derive its asymptotic distribution.

(c) Find the Cramer-Rao Lower Bound (CRLB) for θ and identify which of the above estimators achieve it, if any.