WPI Mathematical Sciences Ph.D. General Comprehensive Exam

MA 541 Probability and Mathematical Statistics II

January 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 (20 points)**.** Let $X_1, \ldots, X_n \overset{ind}{\sim} \text{Uniform}(\theta, \theta + \delta)$.

(a) Find consistent estimators of θ and δ.

(b) Use (a) to deduce unbiased estimators of θ and δ.

(c) Explain why δ is a useful parameter. What happens if the uniform distribution is incorrect?

Problem 2 (20 points)**.** Let $X_1, \ldots, X_n$ be i.i.d. random variables with probability density function

$$f(x \mid \theta) = \theta x^{\theta-1}, \qquad 0 < x < 1,\ \theta > 0.$$

(a) Show that $\sum_{i=1}^{n} (-\ln X_i)$ follows a Chi-square distribution with degrees of freedom $2n$.

(b) Consider testing

$$H_0 : \theta = \theta_0 \quad \text{versus} \quad H_1 : \theta \neq \theta_0.$$

Derive the level α likelihood ratio test and explicitly describe its acceptance region.

(c) Using the answer in part (b) to construct a $100(1-\alpha)\%$ confidence interval for θ.

Problem 3 (20 points)**.** In genetic or medical studies, individual-level data (i.e., data for each subject) are often inaccessible due to privacy concerns. However, for specific analyses, summary statistics, i.e., functions of individual data, are available (de-identified) and might be sufficient. Consider a simple linear regression model, where the response variables $Y_1, \ldots, Y_n$ satisfy $Y_i = \beta x_i + \epsilon_i$, $i = 1, \ldots, n$, where $x_1, \ldots, x_n$ are constant covariates, and $\epsilon_1, \ldots, \epsilon_n$ are i.i.d. $N(0, \sigma^2)$. If we aim to estimate the parameters (β, σ^2), do we need individual-level data $Y_1, \ldots, Y_n$ and $x_1, \ldots, x_n$, or would certain summary statistics suffice? Justify your answer and provide the appropriate summary statistics if they are sufficient.

Problem 4 (20 points). Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed continuous random variables with density

$$f(x) = x\theta^2 e^{-\theta x} \text{ for } x > 0.$$

Let $p = g(\theta) = P_\theta(X_i > 1)$. Consider the following two estimators:

$$\tilde{p}_n = \frac{\sum_{i=1}^n I(X_i > 1)}{n}; \quad \hat{p}_n = g(\hat{\theta}_n),$$

where $\hat{\theta}_n$ is the maximum likelihood estimator of θ.

(a) Find the asymptotic distribution of $\sqrt{n}(\tilde{p}_n - p)$.

(b) Find the asymptotic distribution of $\sqrt{n}(\hat{p}_n - p)$.

(c) Solve for the asymptotic relative efficiency of $\tilde{p}_n$ with respect to $\hat{p}_n$ and identify which estimator is more efficient.

Problem 5 (20 points). One technique for estimating the probability p of a rare event is to conduct repeated, independent Bernoulli (p) trials, and let X be the number of trials necessary to observe one instance of the event. (For instance, if the event first occurred on the 30th trial, $X = 30$.)

(a) Identify the distribution of X and use it to derive the MLE of p

(b) Use X to derive a level-α test of $H_0: \ p \geq p_0$ versus $H_A: \ p < p_0$; be sure to completely describe the rejection region as a function of p_0 and α.

(c) Is the test from (b) UMP? Justify your answer.

Problem 6 (20 points). Let X have the negative binomial distribution

$$P(X = x) = \binom{x-1}{r-1} p^r (1-p)^{x-r}, \quad x = r, r+1, \ldots,$$

where $r > 1$ is a known positive integer and $p \in (0,1)$.

(a) Find the sufficient and complete statistic for p.

(b) What is the UMVUE of p^k for $k \in \{1, 2, \ldots, r-1\}$? (Hint: use the identity: $(1-q)^{-a} = \sum_{n=0}^{\infty} \frac{\Gamma(n+a)}{\Gamma(a)\Gamma(n+1)} q^n$, for $a > 0$, $|q| < 1$.)