WPI Mathematical Sciences Ph.D. General Comprehensive Exam

MA 540 Probability and Mathematical Statistics I

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)**.** Consider the following joint density of (X, Y)

$$f(x,y) = \frac{1}{2\pi\sqrt{(1-\rho^2)}} \exp\{-\frac{1}{2(1-\rho^2)}(x^2 - 2\rho xy + y^2)\}, -\infty < x, y < \infty.$$

(a) Show that $\mathrm{Cor}(X, Y) = \rho$.

(b) Approximate $\mathrm{Cor}(g(X), h(Y))$ for any two differentiable functions, $g(x)$ and $h(y)$. [Hint: First prove a general formula using a first order Taylor's series expansion.]

(c) Use your formula in (b) to find an approximation of $\mathrm{Cor}(e^X, e^Y)$. Are you surprised? Explain.

Problem 2 (20 points)**.** Suppose $X_1, X_2, X_3, \cdots, X_n$ is a random sample from a population with common pdf

$$f(x) = \frac{2x}{\theta^2} \quad ; 0 \le x \le \theta,$$

Let $T = X_{(n)} = \max(X_1, \ldots, X_n)$ be the largest order statistic, and consider estimators of θ of the form

$$\hat{\theta} = CT,$$

where C is a constant.

(a) Find the value of C which makes $\hat{\theta}$ unbiased.

(b) Find the value of C which gives the best mean square error.

Problem 3 (20 points)**.** Assume that in a long DNA chain, mutations occur at an average rate of λ per unit of length, and that in x units of length, the number of mutations, K, is distributed as $Poisson(\lambda x)$, i.e., $P(K = k) = e^{-\lambda x}\frac{(\lambda x)^k}{k!}$. Show that the distance between two adjacent mutations (considered as a continuous random variable) follows an Exponential distribution with parameter λ, i.e., its CDF is $F_X(x) = 1 - e^{-\lambda x}$, and its PDF is $f_X(x) = \lambda e^{-\lambda x}$ for $x > 0$. [Hint: Consider what it means for no mutations to occur in x units of length in terms of K.]

Problem 4 (20 points). Let $X_1, \dots, X_n$ be i.i.d. random variables with continuous distribution function F and density f. Allow $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$ to denote the order statistics.

(a) Allow $D_k = X_{(k+1)} - X_{(k)}$, where k is some integer between 1 and $n-1$. Show that for $d > 0$,

$$f_{D_k}(d) = \frac{n!}{(k-1)!(n-k-1)!} \int_{-\infty}^{\infty} f(x) f(x+d) [F(x)]^{k-1} [1 - F(x+d)]^{n-k-1} \, dx.$$

(b) Now assume $X_i \sim \text{Unif}(0,1)$. Demonstrate that D_k's distribution does not depend of k by solving for the marginal density of D_k.

Problem 5 (20 points). Let $Z \sim Bernoulli(\pi)$, $Y \sim Poisson(\lambda)$, and

$$X = \begin{cases} 0 & \text{if } Z = 0 \\ Y & \text{if } Z = 1 \end{cases}$$

(a) What is the P.M.F. of X?

(b) Find E$[X]$ and Var(X)

(c) Find $Cov(X, Z)$

Problem 6 (20 points). Let $X_1, X_2, \dots$ be i.i.d. random variables with

$$\mathbb{P}(X_1 > x) = x^{-\alpha}, \qquad x \geq 1,$$

for some unknown $\alpha > 0$. Define the partial maxima: $M_n = \max_{1 \leq k \leq n} X_k$.

(a) Show that for every $t > 0$,

$$\lim_{n \to \infty} \mathbb{P}\left(\frac{M_n}{n^{1/\alpha}} \leq t \right) = \exp\left(-t^{-\alpha}\right).$$

(b) Let $Y_k = \log X_k$. Show that M_n admits the representation

$$M_n \overset{d}{=} \exp\left(\max_{1 \leq k \leq n} Y_k \right),$$

where $\max_{k \leq n} Y_k$ is the maximum of i.i.d. exponentials with a rate you must specify.

(c) Show that for any fixed $u > 0$, the number of exceedances

$$K_n(u) := \sum_{k=1}^{n} \mathbf{1}\{X_k > u n^{1/\alpha}\}$$

converges in distribution to a Poisson random variable with mean $u^{-\alpha}$.