PART "A" COMPREHENSIVE EXAMINATION
GENERAL PART A1
Probability and Statistics

June 2022 | 3 hours | 3 pages

Administered by
G. Dafni

Maximum marks: 100

Special Instructions: Do all questions. Read the problems carefully. Show all your work and justify your answers rigourously. Solutions must be clearly written in the examination booklet.
PERMITTED: approved calculators
NOT PERMITTED: books, notes, electronic devices or any material other than calculators. GOOD LUCK!

1.(15 marks)

(a) Let $f : [0, \infty) \to \mathbb{R}$ be a function and suppose **for every strictly decreasing sequence** $\{x_n\}$ converging to 0, the sequence $\{f(x_n)\}$ converges to 2. Prove that $\lim_{x \to 0^+} f(x) = 2$.

(b) Prove that $\displaystyle\sum_{n=1}^{\infty} \frac{1}{4n^2 - 2n}$ converges and find its sum.

(c) If $\{a_n\}$ is a sequence of positive real numbers with $\displaystyle\lim_{n\to\infty} \frac{a_{n+1}}{a_n} = \frac{1}{2}$, prove that $a_n \to 0$. Hint: you may use facts about infinite series.

2.(20 marks)

(a) Let V be the vector space of real polynomials of degree ≤ 2, with basis $B = \{1, x, x^2\}$. Let $T : V \mapsto V$ be the linear transformation sending a polynomial $p(x)$ to $p(x)+p'(x)+p''(x)$ (where $p'(x)$ and $p''(x)$ denote the first and second derivatives of the polynomial p.) Write down the matrix of T relative to the basis B.

(b) Is the matrix in part (a) diagonalizable? Explain.

(c) Let V_1 and V_2 be two vector subspaces of a vector space V. Which of the following subsets are always vector spaces? Justify your answers using short proofs or counterexamples.

(i) The intersection of V_1 and V_2.

(ii) The union of V_1 and V_2

(iii) The span of V_1 and V_2

(iv) The vectors that are in V_1 but not in V_2

(v) The complement of the span of V_1 and V_2.

3.(15 marks) Let X_1, X_2, X_3 be three independent random variables. Suppose that X_1 is uniformly distributed over the interval $(-1, 1)$, X_2 is exponentially distributed with mean $1/2$, and X_3 is exponentially distributed with mean $1/3$. Define

$$Z = \begin{cases} 0, & \text{if } X_1 + X_2 < 3, \\ -2, & \text{otherwise.} \end{cases}$$

(a) Compute $E[Z|X_1]$ and $\text{Var}(Z|X_1)$.

(b) Compute $\text{Var}\{(X_1 + X_2 + X_3)|X_2 > 1\}$.

4.(15 marks) Suppose a 3-state, discrete-time Markov chain $\{X_n\}_{n\geq 0}$ has states $1, 2, 3$, and $\mathbb{P}[X_0 = 1] = 0.35$, $\mathbb{P}[X_0 = 2] = 0.25$. The transition matrix of the Markov chain is

$$P = \begin{bmatrix} 0.1 & 0.2 & 0.7 \\ 0.3 & 0.4 & 0.3 \\ 0.2 & 0.5 & 0.3 \end{bmatrix}$$

Calculate $\mathbb{P}[\{(X_2 - X_1)^2 = 1\} \cup \{X_0 = 2\}]$.

5(a) (**10 marks**) A sample of size 5, say $X_1, \ldots, X_5$, is taken from a Bernoulli (θ) distribution, i.e.,

$$\mathbb{P}[X_i = 1] = \theta = 1 - \mathbb{P}[X_i = 0], \; 1 \leq i \leq 5.$$

Consider the problem of testing hypotheses

$$H_0 : \theta = 0.5 \text{ vs. } H_1 : \theta = 0.75.$$

For the non-randomized test given by the function

$$\varphi(X_1, \ldots, X_5) = \begin{cases} 1 & \text{i.e., reject } H_0, \text{ if } \sum_{i=1}^5 X_i > 4, \\ 0 & \text{i.e., accept } H_0, \text{ if } \sum_{i=1}^5 X_i \leq 4, \end{cases}$$

find the probabilities of type I and type II errors.

5(b) (**5 marks**) If it is required to achieve a size equal to 0.05, how should you modify the above test.

6(a) (**10 marks**) Consider the linear regression model:

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \; E[\epsilon_i] = 0, \mathrm{Var}[\epsilon_i] = \sigma^2, \mathrm{Cov}[\epsilon_i, \epsilon_j] = 0, 1 \leq i \neq j \leq n,$$

where $\beta_0, \beta_1, \sigma^2 > 0$ are unknown model parameters, but $-1 \leq x_i \leq 1$, $1 \leq i \leq n$, are non-random covariates that may be chosen by the experimenter, subject to $\sum_{i=1}^n (x_i - \bar{x})^2 > 0$, where $\bar{x} = \sum_{i=1}^n x_i/n$. Let $\hat{\beta}_0, \; \hat{\beta}_1$ be the *least-squares estimators* of β_0, β_1. Find $\mathrm{Var}_{\sigma^2}(\hat{\beta}_1)$ and the choices of $x_1, \ldots, x_n$ that *minimize* $\mathrm{Var}_{\sigma^2}(\hat{\beta}_1)$ for every $\sigma^2 > 0$.

6(b) (**10 marks**) Consider the linear model:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \end{bmatrix} + \underline{\varepsilon},$$

where $E[\epsilon_i] = 0$, $\mathrm{Var}[\epsilon_i] = \sigma^2$, $\mathrm{Cov}[\epsilon_i, \epsilon_j] = 0$, $1 \leq i \neq j \leq 5$, and θ_j, $1 \leq j \leq 4$, $\sigma^2 > 0$ are unknown model parameters. Note that the design matrix above is not of full column rank. Find the BLUE (*best linear unbiased estimator*) for $(\theta_2 - \theta_3)$ and the variance of the BLUE.