Department of Applied Mathematics
PROBABILITY AND STATISTICS PRELIMINARY EXAMINATION
January 2022

Instructions:

Do two of three problems in each section (Stat and Prob).
Place an **X** on the lines next to the problem numbers
that you are **NOT** submitting for grading.

Prob	
1.	____
2.	____
3.	____

Please do not write your name anywhere on this exam.
You will be identified only by your student number.
Write this number **on each page** submitted for grading.
Show all relevant work.

Stat	
4.	____
5.	____
6.	____
Total	____

Student Number ___________________________

Probability Section

1. Probability: Problem 1

   Consider a random vector (X, Y) taking values in $(0, \infty) \times (0, \infty)$ with the joint probability density function
   $$f(x, y) = e^{-(x+y)} \left[1 + \alpha(2e^{-x} - 1)(2e^{-y} - 1)\right],$$
   where $\alpha \in [-1, 1]$ is a given constant.

   (a) What is the probability density function of Y?

   (b) What is the probability density function of X given Y?

   (c) Are X and Y independent?

   (d) Compute the correlation coefficient between X and Y.

---

2. Probability: Problem 2

   A continuous-time Markov chain X_t is used to model the state of a financial market, which alternates between "*bull*" (the good state) and "*bear*" (the bad state). A statistical analysis shows that "bull" turns into "bear" with a rate $\lambda > 0$, while "bear" turns into "bull" with a different rate $\eta > 0$. Suppose that $X_0 = bull$.

   (a) Write down the infinitesimal generator (or rate matrix) of the continuous-time Markov chain.

   (b) Let T be the time spent for the market to change to *bear* and go back to *bull*. Find the probability density function of T.

A passively managed mutual fund adjusts its portfolio only when the market state changes. It charges a management fee Z_i in the event of the i^{th} market state change, where $\{Z_i\}_{i\in\mathbb{N}}$ are i.i.d. Uniform$(0, 100)$ that are independent of the Markov chain X.

(c) For any $t > 0$, let N_t denote the total number of state changes of the market up to time t and C_t denote the total management fee accumulated up to time t. Show that $\mathbb{E}[C_t] = \mathbb{E}[Z_1]\mathbb{E}[N_t]$.

(d) For any $t > 0$, compute $\mathbb{E}[C_t \mid X_t = bull]$.

---

3. Probability: Problem 3

Let $\{X_i\}_{i\in\mathbb{N}}$ be i.i.d. random variables with $\mathbb{P}(X_i = 1) = p$ and $\mathbb{P}(X_i = -1) = 1 - p$ for some $p \in (0, 1)$. Consider a discrete-time process M defined by

$$M_0 := 0 \quad \text{and} \quad M_t := \sum_{i=1}^{t} X_i \quad \forall t \in \mathbb{N}.$$

Let τ be the first time M reaches either -1 or 3.

(a) If we only focus on the process M up to time τ, we may assume without loss of generality that $M_s := M_\tau$ for $s \geq \tau$. Then, the evolution of M up to time τ can be described using a Markov chain with finite states. Write down the transition matrix P of this Markov chain. Which states are recurrent? Which states are transient?

(b) Find $\mathbb{E}[\tau]$.

(c) Your answer in (b) should be a finite number. Hence, we can apply Wald's equation and get $\mathbb{E}[M_\tau] = \mathbb{E}[X_1]\mathbb{E}[\tau]$. From this, find the probability that M reaches 3 before it reaches -1.

(d) In the case where $\mathbb{E}[M_\tau] = 0$, we re-scale the process M as follows: for any $n \in \mathbb{N}$, define

$$W_t^{(n)} := \frac{1}{\sqrt{2^n}} M_{2^n t}, \quad \forall t \in \mathcal{D}_n := \left\{ \frac{k}{2^n} : k \in \mathbb{N} \cup \{0\} \right\}.$$

Assume that the limiting process

$$W_t := \lim_{n\to\infty} W_t^{(n)}, \quad \forall t \in \bigcup_{n\in\mathbb{N}} \mathcal{D}_n = \left\{ \frac{k}{2^m} : k \in \mathbb{N} \cup \{0\},\ m \in \mathbb{N} \right\}$$

is well-defined. For any fixed $t \in \bigcup_{n\in\mathbb{N}} \mathcal{D}_n$, find the distribution of W_t.
(**Comment:** A Brownian motion emerges as the continuous extension of W_t to all $t \geq 0$).

Statistics Section

4. Statistics: Problem 4

Let $X_1, X_2, \ldots, X_n$ be a random sample from a Uniform$(\theta, 2\theta)$ distribution, where $\theta > 0$.

(a) Find the method of moments (MOM) estimator of θ, $\hat{\theta}_{MOM}$. (Recall that MOM estimators are obtained by equating the sample moments with theoretical moments, and solving for θ).

(b) Find the MLE of θ, $\hat{\theta}_{MLE}$, and find a constant k such that $E_\theta(k\hat{\theta}_{MLE}) = \theta$

(c) Which of these two estimators can be improved using sufficiency, and how?

---

5. Statistics: Problem 5

Let $X_1, X_2, \ldots, X_n$ be a random sample from the continuous distribution with probability density function (pdf)

$$f(x;\theta) = \frac{2\theta(1-x)}{(2x-x^2)^{1-\theta}} I_{(0,1)}(x).$$

Here, $\theta > 0$ and $I_{(0,1)}(x)$ is the indicator function that takes on the value 1 when $0 < x < 1$ and is 0 otherwise.

(a) Find the distribution of $Y_i = -\ln(2X_i - X_i^2)$.

(b) Find the maximum likelihood estimator (MLE) for θ. Show that it is an asymptotically unbiased estimator for θ.

(c) Find the uniformly minimum variance unbiased estimator (UMVUE) for θ.

(d) Is the UMVUE an efficient estimator of θ? Justify.

---

6. Statistics: Problem 6

Let $T_1, T_2, \ldots, T_n$ be iid, continuous, non-negative random variables (representing lifetimes for example) from a distribution with pdf $f(t) = f(t;\theta)$ and cdf $F(t) = F(t;\theta)$. Let $C_1, C_2, \ldots, C_n$ be iid continuous random variables from a distribution with pdf $g(t)$ and cdf $G(t)$, with fixed, known parameters. Suppose we observe $(X_1, \Delta_1), (X_2, \Delta_2), \ldots, (X_n, \Delta_n)$ where

$$X_i = \min(T_i, C_i), \qquad \text{for } i = 1, 2, \ldots, n,$$

and Δ_i is the indicator random variable taking the value 1 if $T_i \leq C_i$ and is 0 otherwise.

Assume that the X_i and C_i are independent, for each i.

(a) Write down $h(\vec{X}, \vec{\Delta})$, the joint density of $\{(X_1, \Delta_1), (X_2, \Delta_2), \ldots, (X_n, \Delta_n)\}$.

(b) The "hazard function" is defined as:

$$h(t) = \lim_{u \to 0} \frac{P(t \leq T < t + u | T \geq t)}{u} = \frac{f(t)}{1 - F(t)}.$$

Consider the "joint" hazard function,

$$h(x, \delta) = \lim_{u \to 0} \frac{P(x \leq X < x + u, \Delta = \delta | X \geq x)}{u}.$$

Give an interpretation of this function specifically when $\delta = 1$.

(c) Suppose now that the lifetimes $T_1, T_2, \ldots, T_n$ are iid exponential random variables with rate λ. Find the MLE (maximum likelihood estimator) of λ based on the observations $(X_1, \Delta_1), (X_2, \Delta_2), \ldots, (X_n, \Delta_n)$.

(d) Estimate the Cramér-Rao lower bound for the variance of all unbiased estimators of λ based on $(X_1, \Delta_1), (X_2, \Delta_2), \ldots, (X_n, \Delta_n)$.
