Department of Applied Mathematics
PROBABILITY AND STATISTICS PRELIMINARY EXAMINATION
August 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

   Suppose that a random vector (X, Y) taking values in $\mathbb{R}^2$ is uniformly distributed within the region enclosed by the curve $y = x^2$ and the line $y = 1$.

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

   (b) Compute the covariance between X and Y.

   (c) Are X and Y independent? Justify your answer!

---

2. Probability: Problem 2

   In oncology, there are three main states of an individual. Initially, one is by default in the "healthy" state (denoted by H), but subject to the potential threat of a cancer. In the next "outbreak" state (denoted by O), a cancer is diagnosed and one goes through medical treatments. In the third "rehabilitation" state (denoted by R), one survives the cancer and gradually returns to activities of daily living. The state of an individual at time $t \geq 0$, $X(t)$, is modeled as a continuous-time Markov chain with the infinitesimal generator (or rate matrix)

$$Q = \begin{bmatrix} -\eta & \eta & 0 \\ 0 & -\lambda & \lambda \\ 0 & \gamma & -\gamma \end{bmatrix}, \tag{1}$$

   where $\eta, \lambda, \gamma > 0$ are given constants with $\gamma > \eta$. We assume that $X(0) = H$.

(a) Explain in words the meaning of "$\gamma > \eta$".

(b) Given $t > 0$, find the probability that $X(t) = O$ and the state R has not yet been visited, i.e.

$$\mathbb{P}\left(X(t) = O \text{ and } X(u) \in \{H, O\}\ \forall u \in [0, t)\right).$$

(c) Given $t > 0$, what is the probability that an individual has exactly *two* cancer outbreaks during the period $[0, t]$?

(d) Suppose that there are $N > 0$ individuals in a population. Each individual is healthy at time 0 and subject to the thread of a cancer as in (1) *independently of other individuals.* As $t \to \infty$, what are the limiting fractions of the population that are in the "healthy", "outbreak", and "rehabilitation" states?

---

3. Probability: Problem 3

In a casino, a gambler with initial wealth \$1,000 plans to participate in a game. In each round of the game, the gambler either loses \$1,000 (with probability $p \in (0, 1)$) or wins \$1,000 (with probability $1 - p$). For convenience, we express this as

$$\mathbb{P}(X_n = -1000) = p \quad \text{and} \quad \mathbb{P}(X_n = 1000) = 1 - p,$$

where X_n represents the payoff from the n^{th} round of the game. We assume that every round of the game is carried out independently. Suppose that the gambler follows the following stopping strategy: he will leave the game once he is bankrupt (i.e., his wealth becomes \$0) or his wealth reaches \$4,000, whichever comes first. Let N denote the total number of rounds the gambler participates in. Note that N is a random variable depending on the stopping strategy of the gambler.

(a) The wealth of the gambler after the n^{th} round, denoted by W_n for $n = 0, 1, , ..., N$, 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 $\mathbb{E}[N]$.

(c) Show that $\mathbb{E}[X_1 + X_2 + \cdots + X_N] = \mathbb{E}[X_1]\mathbb{E}[N]$.

(d) Suppose that the gambler (still with initial wealth \$1,000) follows a slightly different strategy: he will leave the game once he is bankrupt (i.e., his wealth becomes \$0) or his wealth reaches \$5,000. In terms of maximizing his expected wealth when leaving the game (i.e., $\mathbb{E}[W_N]$) , which strategy (the new one or the original one) is better? Justify your answer!
(**Hint:** Use part (c)).

Statistics Section

4. Statistics: Problem 4

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.

---

5. Statistics: Problem 5

Let X and Y be two independent random variables with $X \sim$ exponential(λ) and $Y \sim$ exponential(μ). (Recall that if $Z \sim$ exponential(λ), the density is $f(z) = \lambda \exp(-\lambda z)$ and the CDF is $F(z) = 1 - \exp(-\lambda z)$.)

It is impossible to obtain direct observations of X and Y. Instead, we only observe the random variables Z and W where: $Z = \min\{X, Y\}$ and $W = 1$ when $Z = X$, and 0 otherwise. (This situation arises frequently in medical trials, where we say that the variables X and Y are censored.)

(a) Find the joint distribution function of Z and W.

(b) Prove that Z and W are independent.

(c) Assume that $\{(Z_i, W_i)\}_{i=1}^{i=n}$ are n independent and identically distributed bivariate observations. Find the MLE of λ and μ. (For partial credit, outline every step in the process of finding the MLE.)

---

6. Statistics: Problem 6

The jackknife is a general technique for reducing bias in an estimator (Quenouille, 1956; Miller, 1974). A one-step jackknife estimator is defined as follows. Let $X_1, ..., X_n$ be a random sample, and let $T_n = T_n(X_1, ..., X_n)$ be some estimator of a parameter θ. In order to "jackknife" T_n, we calculate the n statistics $T_n^{(i)}$, $i = 1, ..., n$, where $T_n^{(i)}$ is calcuated just as T_n but using the $n-1$ observations with X_i removed from the sample. The jackknife estimator of θ, denoted by T_n^* is given by

$$T_n^* = nT_n - \frac{n-1}{n}\sum_{i=1}^{n} T_n^{(i)}$$

(In general, T_n^* will have a smaller bias than T_n.)

Now to be specific, let $X_1, ..., X_n$ be independent and identically distributed Bernoulli(θ) random variables. The object is to estimate θ^2.

(a) Find the MLE of θ^2, and show that it is a biased estimator of θ^2.

(b) Derive the one-step jackknife estimator based on the MLE.

(c) Show that the one-step jackknife estimator is an unbiased estimator of θ^2. (In general, jackknifing only reduces bias; in this special case, however, it removes it entirely.)

(d) Is this jackknife estimator the best unbiased estimator of θ^2? If yes, prove it. If no, find the best unbiased estimator.