Department of Applied Mathematics
PROBABILITY AND STATISTICS PRELIMINARY EXAMINATION
January 2020

Instructions:

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

Prob
1. ____
2. ____
3. ____

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

Problem 1.

Let X and Y be continuous random variables with $Y \geq 0$ and joint p.d.f. $f(x,y)$, for $x \in \mathbb{R}$ and $y \geq 0$. Consider the random variable $Z := Y/(1+X^2)$.

(a) Show that

$$E(Z) = \int_{-\infty}^{+\infty} \int_0^{\infty} \frac{y\, f(x,y)}{1+x^2}\, dy\, dx,$$

using the conditional expectation $E(Z|X)$ and the conditional p.d.f. $f(y|x)$. You must use these two concepts to get full credit for this part!

Next, you will be guided to deduce the above expression using more basic principles.

(b) Determine the c.d.f. of Z in terms of $f(x,y)$.

(c) Deduce the p.d.f. of Z.

(d) Using part (c), determine $E(Z)$, and show the new identity is equivalent to the one in part (a).

Problem 2.

A machine is subject to a failure rate $f > 0$ and—upon failure—requires an exponential amount of time with rate $r > 0$ to be repaired.

(a) Specify the rate matrix Q of the natural continuous-time Markov process associated with the functioning of the machine. Use state 0 to denote that the machine is working, and state 1 to denote that it is being repaired.

(b) What's the probability that after a very long time the machine is working? Explain!

(c) If the machine is currently working, what's the probability it continues doing so without interruptions during the next t units of time? Explain!

(d) If the machine is currently working, what's the probability that it is working t units of time later? Justify!

Problem 3.

Let $r_0, r_1, r_2, \ldots$ be real numbers such that $r_i > 0$, and $\sum_{i=0}^{\infty} r_i = 1$.

Consider a discrete-time homogeneous Markov chain $X = (X_n)_{n \geq 0}$ with state space $S = \{0, 1, 2, \ldots\}$ and probability transitions $p(0, i) = r_i$ for $i \geq 0$, and $p(i, i) = p(i, i-1) = 1/2$ for $i > 0$. In what follows, T_0 denotes the *return time to state* 0 i.e. $T_0 := \min\{n \geq 1 \text{ such that } X_n = 0\}$.

(a) Is this chain irreducible? Explain!

(b) Determine $\mathbb{E}(T_0 \mid X_0 = i)$ for $i > 0$.

`Hint.` $\sum_{k=1}^{\infty} k(1-x)^{k-1} = \frac{1}{(1-x)^2}$ when $|x| < 1$.

(c) Conclude that the chain has a stationary distribution if and only if $\sum_{i=0}^{\infty} i \cdot r_i < +\infty$.

(d) Which states are recurrent if any? Explain!

Statistics Section

Problem 4.

1. Let $X_1, X_2, \ldots, X_n$ be a random sample from the exponential distribution with rate λ. Imagine that these are lifetimes of widgets and that we get to observe the first m widget failures $X_{(1)}, X_{(2)}, \ldots, X_{(m)}$ for some $m \leq n$. (Here, $X_{(1)}$ is the minimum value of the X's, $X_{(2)}$ is the next smallest, etcetera.)

   For notational simplicity, we will define $Y_i = X_{(i)}$ for $i = 1, 2, \ldots, m$.

   Let f and F denote the p.d.f. and c.d.f., respectively, for the exponential rate λ distribution.

   (a) Give a heuristic explanation (or otherwise prove) that the joint p.d.f. for $Y_1, Y_2, \ldots, Y_m$ is

   $$f(y_1, y_2, \ldots, y_m) = \frac{n!}{(n-m)!} [F(y_1)]^0 f(y_1) f(y_2) \cdots f(y_m) [1 - F(y_m)]^{n-m} I_{\{0 < y_1 < y_2 < \cdots < y_m\}}$$

   (b) Find the MLE (maximum likelihood estimator) of $1/\lambda$, the mean of the underlying exponential distribution.

   (c) Show that $S := (n-m)Y_m + \sum_{i=1}^{m} Y_i$ is complete and sufficient for λ.

(d) It is well known that Y_1, as the minimum of n exponentials with rate λ, has again an exponential distribution but with rate $n\lambda$. One can show that Y_2 has the same distribution as $E_1 + E_2$, where E_1 and E_2 are independent with $E_1 \sim exp(rate = n\lambda)$ and $E_2 \sim exp(rate = (n-1)\lambda)$, and, in general,

$$Y_i \stackrel{d}{=} E_1 + E_2 + \cdots + E_i$$

where $E_1, E_2, \ldots, E_i$ are independent and $E_i \sim exp(rate = (n-i+1)\lambda)$.
(Do not show this!)

Find the distribution of S.

(e) Finally, find the UMVUE (uniformly minimum variance unbiased estimator) of $1/\lambda$.

Problem 5.

Let $X_1, X_2, \ldots, X_n$ be a random sample from the uniform distribution on the interval $(0, \theta)$ for some $\theta > 0$ and $Y_1, Y_2, \ldots, Y_m$ a random sample from the uniform distribution on $(0, \lambda)$ for some $\lambda > 0$. Assume that the X's and Y's are mutually independent.

(a) Find the probability density function for $W = W(\vec{X}, \vec{Y}) = X_{(n)}/Y_{(m)}$, where $X_{(n)} = \max(X_1, X_2, \ldots, X_n)$ and $Y_{(m)} = \max(Y_1, Y_2, \ldots, Y_m)$.

(b) Consider deriving a test for

$$H_0 : \theta \leq \lambda \qquad \text{versus} \qquad H_1 : \theta > \lambda$$

based on the statistic W from part (a).

i. Restate the p.d.f. for W and the hypotheses in terms of the parameter $\rho := \theta/\lambda$.

ii. Which of the following critical regions is appropriate? (Choose one.)

$$C = \{(\vec{x}, \vec{y}) : W(\vec{x}, \vec{y}) < c\}$$

or

$$C = \{(\vec{x}, \vec{y}) : W(\vec{x}, \vec{y}) > c\}$$

iii. Give a fully simplified expression for $P(W \in C; \rho)$ for your chosen critical region C.

iv. For $n = 1$, $m = 2$, find the value of c to give a size $\alpha = 0.05$ and clearly state your test. (i.e. Reject H_0 in favor of H_1 if ...)

Problem 6.

Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution with pdf $f(x; \theta)$. Assume the distribution satisfies the "regularity conditions" needed for the Cramér-Rao lower bound to be applicable.

Let $L(\theta)$ be a likelihood for this model and let $\ell(\theta)$ be the log-likelihood.

(a) Find the asymptotic distribution of $\frac{\partial}{\partial\theta}\ell(\theta)$. Your answer should be in terms of the Fisher information.

(b) Suppose that $X_1, X_2, \ldots, X_n$ is a random sample from the distribution with pdf

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

Use part (a) to derive an approximate $100(1-\alpha)\%$ confidence interval for θ assuming n is large.

(c) Derive an exact $100(1-\alpha)\%$ confidence interval for θ in terms of a chi-squared critical value.
