SM-4331 Advanced Statistics Class Test 1
2019/20 Semester 2
18 February 2020
Time allowed: 90 minutes

Instructions:

- There are three (3) questions totalling 20 marks and one (1) bonus question for 5 marks. The total attainable marks is 20 only.
- Answer **ALL** questions on a separate answer sheet.
- Ensure that you have written your name and student number on your answer sheets that you are submitting.
- The use of calculators is allowed.

Question:	1	2	3	4	Total
Marks:	8	4	7	5	20

1. Let $X_1, \ldots, X_n$ be a random sample from an exponential distribution with the density function
$$f(x|\lambda) = \begin{cases} \lambda^{-1} e^{-x/\lambda} & x \geq 0 \\ 0 & x < 0 \end{cases}$$
where $\lambda > 0$ is an unknown parameter.
   (a) (3 marks) Find the MLE $\hat{\mu}$ for μ.
   (b) (3 marks) Find the Cramér-Rao lower bound for the variance of the unbiased estimators for λ.
   (c) (2 marks) Find the MLE for $\theta = \lambda^2$, and show that it is biased.

2. (a) (1 mark) Let X_n, Y_n, X and Y be random variables, g a continuous function, and c a real-valued constant. All of the following are results of Slutzky's theorem as $n \to \infty$, **except one**. Which one of the below statements is not necessarily true?
      A. If $X_n \xrightarrow{\text{P}} X$ and $Y_n \xrightarrow{\text{P}} Y$, $X_n + Y_n \xrightarrow{\text{P}} X + Y$.
      B. If $X_n \xrightarrow{\text{P}} X$ and $Y_n \xrightarrow{\text{P}} Y$, $X_n Y_n \xrightarrow{\text{P}} XY$.
      C. If $X_n \xrightarrow{\text{D}} X$, $g(X_n) \xrightarrow{\text{D}} g(X)$.
      D. If $X_n \xrightarrow{\text{D}} X$ and $Y_n \xrightarrow{\text{D}} Y$, $X_n + Y_n \xrightarrow{\text{D}} X + Y$.
   (b) (3 marks) Let $\{X_n\}$ be a sequence of random variables. Suppose that $\mathrm{E}(X_n) \to c$ and $\mathrm{Var}(X_n) \to 0$ as $n \to \infty$, where c is a fixed constant. Using Markov's inequality, show that X_n converges to c in probability.

3. A forester wants to estimate the total number of farm acres planted in trees for a state. Because the number of acres of trees varies considerably with the size of the farm, he decides to stratify on farm sizes. The $N = 240$ farms in the state are placed in one of four categories according to the size. A stratified random sample of $n = 40$ farms, selected by using proportional allocation, yields the results shown in Table 1.

n_h	Strata I	Strata II	Strata III	Strata IV
1	97	125	142	167
2	42	67	310	220
3	25	256	495	780
4	105	310	320	655
5	27	220	196	540
6	45	142	256	
7	53	155	440	
8	67	96	510	
9	125	47	396	
10	92	236		
11	86	352		
12	43	190		
13	59			
14	21			
N_h	86	72	52	30
$\sum x$	887	2196	3065	2362
$\sum x^2$	70131	501464	1178157	1405314

Table 1: Total farm acres of trees for the stratified sample of $n = 40$ farms.

(a) (2 marks) Calculate the sample mean and (unbiased) sample variance of the number of farm acres planted in trees within each strata.

(b) (2 marks) Using your answer to (a), calculate the estimate of the mean number of farm acres planted in trees from the stratified sample. Give an estimate of the variance of this estimator.

(c) (3 marks) Estimate the total number of acres of trees on farms in the state, and construct a suitable 95% confidence interval for this estimate.

——— *Bonus Question* ———

4. (a) (1 mark) State, without proof, the pmf $f(x)$ of a random variable X distributed according to $X \sim \mathrm{Bin}(n, p)$.

   (b) (1 mark) What is the mean and variance of $X \sim \mathrm{Bin}(n, p)$?

   (c) (2 marks) Let $Y_1, \dots, Y_n \overset{\text{iid}}{\sim} \mathrm{Bern}(p)$ be a random sample, and $\hat{p} = X/n$ be an estimator for p, where $X = \sum_{i=1}^{n} Y_i$. Using the CLT, find the distribution of X.

   (d) (1 mark) What can you say about the distribution of $X \sim \mathrm{Bin}(n, p)$ as $n \to \infty$?

——— *End of Paper* ———