ST.JOSEPHS COLLEGE (AUTONOMOUS), BANGALORE- 27

M.Sc.Big Data Analytics I SEMESTER

SEMESTER EXAMINATION - OCTOBER 2019

BDA 1218 : PROBABILITY & STOCHASTIC PROCESS

Time: $2\frac{1}{2}$ **hrs** **Maximum marks: 70**

(This question paper has 2 printed pages and 1 part.)

1. (a) The probability of passing a test is 3/4. If three students take the test, what is the probability that (a) all three will fail and (b) at least one of the three will fail? [4]

   (b) If $P(A) = \frac{1}{2}, P(B) = \frac{1}{3}$ and $P(A^c \cap B^c) = \frac{5}{12}$, find $P(A|B)$. [2]

   (c) In some states, license plates have four characters: two letters followed by two numbers. If all sequences of four characters are equally likely, what is the probability that the license plate for a new car will contain no duplicate letters or numbers? [4]

2. There are three cabinets, 1, 2, and 3, each of which has two drawers. Each drawer contains one coin; A has two gold coins, B has two silver coins, and C has one gold and one silver coin. A cabinet is chosen at random, one drawer is opened, and a silver coin is found. What is the probability that the other drawer in that cabinet contains a silver coin? [10]

3. (a) Let X be the set of possible outcomes when a fair die is tossed. Then what is Var(X)? [6]

   (b) Let X, Y and Z be three random variables. What is the covariance between $X+Y$ and Z. [4]

4. (a) Give an example to highlight the difference between Type I and Type II errors[2]

   (b) Explain (preferably with illustrations) the different steps involved in testing a hypothesis [6]

   (c) What is the p-value [2]

5. It is claimed that sports-car owners drive on the average 18,000 miles per year. A consumer firm believes that the average mileage is probably lower. To check, the consumer firm obtained information from 40 randomly selected sports-car owners that resulted in a sample mean of 17,463 miles with a sample standard deviation of 1348 miles. What can we conclude about this claim? We can assume that random sample comes from a normal population. Use $\alpha = 0.01$ and $z_{0.01}$ is given as -2.33.

6. A problem of interest to sociologists is to determine the proportion society that has an upper- or lower-class occupation. Suppose that occupations are grouped into upper (U), middle (M), and lower (L) levels. U_1 will denote the event that a father's occupation is upper-level; U_2 will denote the event that a son's occupation is upper-level, etc.

	U_2	M_2	L_2
U_1	0.45	0.50	0.05
M_1	0.10	0.65	0.25
L_1	0.01	0.54	0.45

Such a table, is to be read in the following way: If a father is in U, the probability that his son is in U is 0.45, the probability that his son is in M is 0.50, etc. Find out what proportion of its people, in long run, will be in upper-class, middle-class and lower-class occupations? [10]

7. Attempt **any two** of the four questions below:

   (a) Discuss the Poisson distribution [5]

   (b) Discuss the Normal distribution [5]

   (c) Discuss the Markovian property with an illustration [5]

   (d) Distinguish between parametric and non-parametric tests. When would you use which? [5]