Question Paper

MANIPAL ACADEMY OF HIGHER EDUCATION

MANIPAL SCHOOL OF INFORMATION SCIENCES, MANIPAL
FIRST SEMESTER MASTER OF ENGINEERING - ME (BIG DATA AND DATA ANALYTICS / HEALTHCARE DATA ANALYTICS / MACHINE LEARNING) DEGREE EXAMINATION - FEBRUARY 2021

Applied Probability and Statistics [MCL 601]

Marks: 100   **Duration: 180 mins.**

WEDNESDAY, FEBRUARY 17, 2021

Answer all the questions.

1) Give one practical explanation for each of the following numbers without simplifying and without using the words "objects", "bins", "groups", "distinct" etc. Your answer should be clear, and feel free to explaining using more than a sentence if needed. [TLO 1.1, L3]   (10)

   (a) 10^{365}

   (b) $45C_4$

   (c) $(3 + 100 - 1)C_{100}$

   (d) $7!$

   (e) $(100 - 1)C_{3-1}$

2) A hospital administrator codes incoming patients suffering gunshot wounds according to whether they have insurance (coding 1 if they do and 0 if they do not) and according to their condition, which is rated as good (g), fair (f), or serious (s). Consider an experiment that consists of the coding of a random patient.   (10)

   (a) Give the sample space of this experiment clearly showing the outcomes. [TLO 2.1, L5]

   (b) Let A be the event that the patient is in serious condition. Specify the outcomes in A. [TLO 2.1, L4]

   (c) Let B be the event that the patient is uninsured. Specify the outcomes in B. [TLO 2.1, L5]

   (d) Give all the outcomes in the event $(B^c \text{ or } A)$. [TLO 2.1, L5]

3)   (10)

A small community organization consists of 20 families, of which 4 have one child, 8 have two children, 5 have three children, 2 have four children, and 1 has five children.

(a) Consider the experiment of randomly choosing a family. Give the sample space of this experiment. You may label the families with one child a A_1, A_2, A_3, A_4, two children as $B_1, B_2, \ldots, B_8$, etc. [TLO 2.1, L5]

(b) If one of the families is randomly chosen, what is the probability that it has 1 child? What is the probability that it has 5 children? [TLO 2.3, L4]

  Hint: *Any family is equally likely to be chosen.*

(c) Consider the experiment of randomly choosing a child. How many outcomes are there in the sample space of this experiment? [TLO 2.1, L5]

(d) If one of the children is randomly chosen, what is the probability that the child comes from a family having 1 child? What is the probability that the child comes from a family having 5 children? [TLO 2.3, L4]

4) Suppose we assume that 5% of people are drug-users. A test is 95% accurate, which (10) means that if a person is a drug-user then the test result is positive 95% of the time; and if the person is not a drug-user then the test result is negative 95% of the time. Let us define the following events:

$$U = \text{person is a drug user},$$
$$P = \text{test positive}.$$

(a) Explain what the following quantities mean: [TLO 3.1, L5]
  - $P(U \mid P^c)$
  - $P(P \mid U^c)$
  - $P(P \text{ and } U)$
  - $P(U^c \mid P)$

(b) What quantity above would be of interest if we are concerned about making a wrong accusation? [TLO 3.1, L6]

(c) What quantity above would be of interest if we are concerned about high safety standard when recruiting some one for a job? [TLO 3.1, L6]

(d) A randomly chosen person tests positive. Is the individual highly likely to be a drug-user? Show your reasoning clearly. [TLO 3.2, L6]

5) In a certain community, 36 percent of the families own a dog and 22 percent of the (10) families that own a dog also own a cat. In addition, 30 percent of the families own a cat.

(a) What is the probability that a randomly selected family owns a dog? [TLO 3.2, L4]

(b) What the probability that a randomly selected family owns a cat? [TLO 3.2, L4

(c) What probability does 22 percent in the problem statement indicate? [TLO 3.2, L5]

(d) What is the probability that a randomly selected family owns both a dog and a cat? [TLO 3.2, L4]
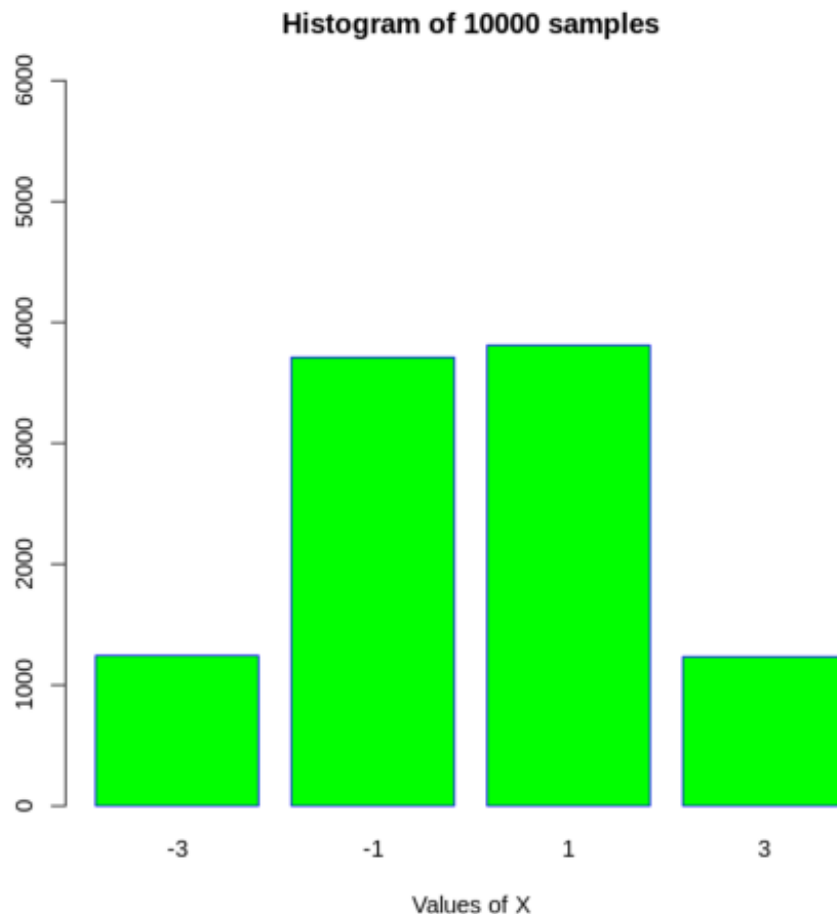
(e) What is the the conditional probability that a randomly selected family owns a dog given that it owns a cat? [TLO 3.2, L4]

6) (10)

Let X be a random variable that represents the difference between the number of heads and the number of tails obtained when a fair coin is tossed 3 times. Assume that the tosses are independent.

(a) What are the possible values of X? [TLO 4.1, L5]

(b) What are the associated probabilities for the values that X can take? [TLO 4.2, L4]

   Hint: *for example, if $X = 2$, clearly think in plain English how this could have happened using the words and/or, and then compute the probability.*

(c) The graph below shows a frequency histogram of the values that X can take from 10000 samples:

Histogram of 10000 samples

Values of X

What is the approximate probability that $P(X = \text{positive})$? [TLO 5.1, L4]

7) A person conducting telephone surveys must get 4 more completed surveys before their job is finished. On each randomly dialed number, there is a 90% chance of the participant rejecting the call. What is the probability that the person will finish their job at the 10th call? [TLO 4.2, L4] (10)

8) (10)

The expected number of typographical errors (that is, the number of mistyped letters) on a randomly chosen page of a certain magazine is 0.2. Let X be the random variable representing the number of typographical errors on a random page.

(a) What type of random variable is X? What is its associated parameter? Explain your reasoning clearly. [TLO 4.1, L4]

(b) What is the probability that the next page you will read has no typographical errors? [TLO 4.2, L4]

(c) What is the probability that the next page you will read has at least 2 typographical errors? [TLO 4.2, L4]

9) A machine produces items in batches. For each batch, the machine can be in control or out of control. Suppose the machine is in control for 99% of the production batches; The production defect rate is 0.01 when the machine is in control and 0.1 when the machine is out of control. Assume that the production size is 1000 items per batch. (10)

(a) If D is the number of defective items in a random production batch, compute the probability that the machine produces k defective items. Leave the answer as an expression involving k. [TLO 3.2, 4.2, L5]

Hint: *Starting with $P(D = k)$, use the law of total probability and the binomial random variable formula.*

(b) Suppose we want to update our *prior* knowledge of the probability that the machine is in-/out-of control. To that end, we select a small random sample of 5 items from a production batch and inspect how many among them are defective, which is represented by X. What is the probability that the machine is out of control given that 3 defective items were observed? [TLO 3.2, 4.2, 5.1, L5]

Hint: Use Bayes' theorem starting with $P(\text{Out of control} \mid X = 3)$.

(c) Suppose we want to interrupt the production process when we are suspicious that the production process is out of control based on the number of defective items we see in a sample of 5. Let us quantify our suspicion as a 40% or greater chance. Would you stop production if 3 defective items are observed? Explain clearly. [TLO 5.1, L6]

10) (10)

Say true or false; no explanation needed. Each question carries 2 points and an incorrect answer carries a negative point. The minimum score for this question is 0. [TLO 4.4, 4.5, 5.3, L4]

(a) The expected value of a random variable should always be one among the values that the random variable can take.

(b) Sample mean is a random variable.

(c) For a continuous random variable X modeling student scores in a test, it is possible that $F_X(40) > F_X(50)$, where F_X represents X's cumulative distribution function (CDF).

(d) For a continuous random variable X, its probability density function (PDF)

$$f_X(\text{input}) \leq 1$$

always for any input.

(e) The probability that a continuous random variable takes on any particular value is always zero.

-----End-----