CGS602A: Basic Statistics, Data Analysis and Inference Assignment #1:

Max marks:130

Due on/before: midnight, 3-Oct-2019.

- 1. (a) What is the sample space and the number of outcomes in the sample space for the following experiment? Toss a coin until a Head appears and record the number of tosses required.
 - (b) The Poisson distribution is defined by the expression below where θ is a parameter of the distribution.

$$p(x) = \begin{cases} \frac{\theta^x e^{-\theta}}{x!} & x = 0, 1, 2, ..\\ 0 & \text{otherwise} \end{cases}$$

Calculate the mean and variance of the distribution.

(c) Let Y be a random variable and m be the median Y - that is $P(Y \le m) \ge 0.5$ and $P(Y \ge m) \ge 0.5$. For real numbers a and b such that $m \le a \le b$ or $m \ge a \ge b$ show that $E(|Y - a|) \le E(|Y - b|)$. Assume $E(Y) < \infty$.

[5,10,10=25]

2. The winning number(s) in a lottery is chosen as follows. Numbers from 01 to 44 are written on identical tiles and put in a box which is shaken thoroughly just before drawing a tile.

Six tiles are drawn from the box and the winning number is constructed from the tiles. For each case below calculate the total number of possible numbers.

- (a) The tiles are drawn one-by-one without replacement and the winning number is the number formed by the sequence of tiles in the order drawn.
- (b) Same as part (a) except that the tiles are drawn with replacement.
- (c) Same as part (a) except that all numbers formed from the six tiles drawn are considered winners that is they are unordered.
- (d) Same as part (c) except that the tiles are drawn with replacement.
- (e) If the box contained n tiles and r were drawn write expressions in terms of n and r for each of the four ways mentioned above.

 $[5 \times 4, 10 = 30]$

- 3. (a) Argue for the correctness of the following two consequences of the probability axioms: If $A \subseteq B$ then $P(A) \leq P(B)$ and $P(A \cup B) = P(A) + P(B) P(A \cap B)$ where A, B are events.
 - (b) Argue that if events A, B are independent events then A^c and B^c are also independent.
 - (c) Assume that the incidence of Corona virus infection in the population is approx. 1%. We arrive at this as follows: current total infections approx. 60 lakhs. An estimate suggests that the actual infections are 20 times those detected so 12 million are infected. Assuming population is 1.3 billion or 1300 million we get approx. 1% incidence in the population.

The RT-PCR test is the main reliable test and assume its sensitivity is 0.95. Sensitivity means True Positive Rate that is 95% of the time if the person is infected the test will correctly detect that. Let the

24-Sep-2020

false positive rate be 0.01 - that is the test is positive 1% of the time when a persons does not have the disease.

Suppose in random testing a person gets a positive test what is the probability that he/she is actually infected. So, if I is the event the person is infected and E is the event the test is positive you must calculate P(I|E).

[(5,5),5,10=25]

- 4. (a) Let a box have n pairs of shoes and you randomly choose 2r shoes from the box. Calculate the probability that there is no matching pair in the shoes chosen.
 - (b) You and a friend each toss a fair coin n times. Find the probability that both of you have the same number of Heads.
 - (c) You want to hire one new employee. N applicants turn up. Your HR department gives them a test and orders them from 1 to N for suitability where 1 is best. You adopt the following protocol:
 - 1. Each candidate is seen in sequence in random order and a decision is made whether to hire the candidate.
 - 2. Having rejected k-1 candidates (k > 1) the kth candidate can be hired only if he/she is better than the previous (k-1) candidates.

Assuming the i^{th} candidate in the sequence was hired what is the probability the best candidate was hired?

(d) A fair die is rolled n times what is the probability that one of the faces never came up even once? For different values of n try to simulate this by doing the experiment many times on a computer (for the same value of n) and check if your answer is close to what you get through your simulation - this is your first Python program.

[10, 10, 10, (10, 10) = 50]