Lecture 1
Chapter 1: Introduction to Statistics and Data Analysis

Measures of Location: The Sample Mean and Median

The aim of this lecture is to explain the following concepts :

- Measures of Location.
- The Sample Mean and Median.
- The Sample Range and Sample Standard Deviation.
- Histogram.

Definition 1 *Suppose that the observations in a sample are $x_1, x_2, ..., x_n$. The **sample mean**, denoted by $\bar{x}$, is*

$$\bar{x} = \sum_{i=1}^{n} \frac{x_1 + x_2 + ... + x_n}{n}$$

Definition 2 *Given that the observations in a sample are $x_1, x_2, ..., x_n$, arranged in increasing order of magnitude, the **sample median** is*

$$\bar{x} = \begin{cases} x_{(n+1)/2}, & if\ n\ is\ odd \\ \frac{1}{2}(x_{n/2} + x_{n/2+1}), & if\ n\ is\ even. \end{cases}$$

Definition 3 *A **trimmed mean** is computed by "trimming away" a certain percent of both the largest and the smallest set of values.*
For example, the 10% trimmed mean is found by eliminating the largest 10% and smallest 10% and computing the average of the remaining values.

Definition 4 *The **sample variance**, denoted by s^2, is given by*

$$s^2 = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n-1}$$

*The **sample standard deviation**, denoted by s, is the positive square root of s^2, that is,*

$$s = \sqrt{s^2}$$

Example 1 : *An engineer is interested in testing the "bias" in a pH meter. Data are collected on the meter by measuring the pH of a neutral substance (pH = 7.0). A sample of size 10 is taken, with results given by*
7.07 7.00 7.10 6.97 7.00 7.03 7.01 7.01 6.98 7.08.
Find sample variance and standard deviation.

Solution : *The sample mean $\bar{x}$ is given by*

$$\bar{x} = \frac{7.07 + 7.00 + 7.10 + ... + 7.08}{10} = 7.0250.$$

The sample variance s^2 is given by

$$s^2 = \frac{1}{9}[(7.07-7.025)^2+(7.00-7.025)^2+(7.10-7.025)^2+....+(7.08-7.025)^2] = 0.001939.$$

As a result, the sample standard deviation is given by

$$s = \sqrt{0.001939} = 0.044.$$

So the sample standard deviation is 0.0440 with n-1 = 9 degrees of freedom.

Exercises :

1. The following measurements were recorded for the drying time, in hours, of a certain brand of latex paint.
3.4 2.5 4.8 2.9 3.6
2.8 3.3 5.6 3.7 2.8
4.4 4.0 5.2 3.0 4.8
Assume that the measurements are a simple random sample.
(a) What is the sample size for the above sample?
(b) Calculate the sample mean for these data.
(c) Calculate the sample median.
(d) Compute the 20% trimmed mean for the above data set.
Solution :

(a) sample size = 15.

(b) $\bar{x} = \frac{1}{15}(3.4 + 2.5 + 4.8 + ... + 4.8) = 3.787$

(c) Sample median is the 8th value, after the data is sorted from smallest to largest = 3.6.

(d) After trimming total 40% of the data (20% highest and 20% lowest), the data becomes:
2.9 3.0 3.3 3.4 3.6
3.7 4.0 4.4 4.8.
So. the trimmed mean is

$$\bar{x}_{tr20} = \frac{1}{9}(2.9 + 3.0 + ... + 4.8) = 3.678.$$

2. According to the journal Chemical Engineering, an important property of a fiber is its water absorbency. A random sample of 20 pieces of cotton fiber was taken and the absorbency on each piece was measured. The following are the absorbency values:
18.71 21.41 20.72 21.81 19.29 22.43 20.17
23.71 19.44 20.50 18.92 20.33 23.00 22.85
19.25 21.77 22.11 19.77 18.04 21.12
(a) Calculate the sample mean and median for the above sample values.
(b) Compute the 10% trimmed mean.
Solution :
Given sample size = 20.
(a) Mean=20.768 and Median=20.610.
(b) $\bar{x}_{tr10} = 20.743$.

7. The following measurements were recorded for the drying time, in hours, of a certain brand of latex paint.
3.4 2.5 4.8 2.9 3.6
2.8 3.3 5.6 3.7 2.8
4.4 4.0 5.2 3.0 4.8
Compute the sample variance and sample standard deviation.
Solution : The sample variance s^2 is given by

$$s^2 = \frac{1}{15-1}[(3.4-3.787)^2+(2.5-3.787)^2+(4.8-3.787)^2+.....+(4.8-3.787)^2] = 0.94284.$$

As a result, the sample standard deviation is given by

$$s = \sqrt{0.9428} = 0.971.$$

8. Compute the sample variance and standard deviation for the water absorbency data of the above Q. No. 2.
Solution : The sample variance s^2 is given by

$$s^2 = \frac{1}{20-1}[(18.71-20.768)^2+(21.41-20.768)^2+....+(21.12-20.768)^2] = 0.94284.$$

As a result, the sample standard deviation is given by

$$s = \sqrt{2.5345} = 1.592.$$

Histogram :
A table listing relative frequencies is called a **relative frequency distribution**.
The information provided by a relative frequency distribution in tabular form is easier to grasp if presented **graphically**.
Using the midpoint of each interval and the corresponding relative frequency, we construct a **relative frequency histogram**.

Class Interval	Class Midpoint	Frequency, f	Relative Frequency
1.5–1.9	1.7	2	0.050
2.0–2.4	2.2	1	0.025
2.5–2.9	2.7	4	0.100
3.0–3.4	3.2	15	0.375
3.5–3.9	3.7	10	0.250
4.0–4.4	4.2	5	0.125
4.5–4.9	4.7	3	0.075

Figure 1: Relative Frequency Distribution of Battery Life

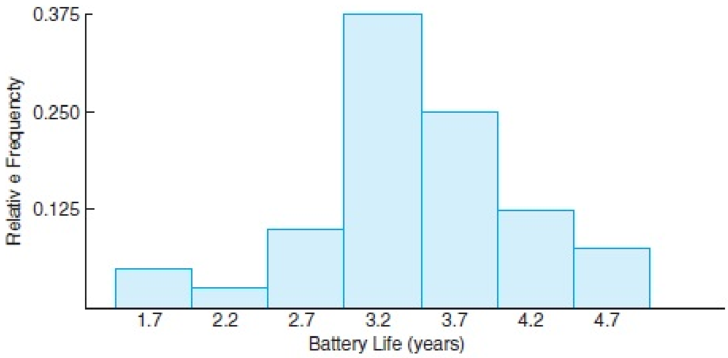

Figure 2: Relative frequency histogram

Lecture 2
Chapter 2 : Probability

2.1 Sample Space, 2.2 Events, 2.3 Counting Sample Points

The aim of this lecture is to explain the following concepts :

- Sample Space.
- Event.
- Counting Sample Points.

2.1 Sample Space :

Definition 1 *The set of all possible outcomes of a statistical experiment is called the **sample space** and is represented by the symbol S.*

Notes :

- Each outcome in a sample space is called an element or a member of the sample space, or simply a sample point.
- If the sample space has a finite number of elements, we may list the members separated by commas and enclosed in braces.
- Thus, the sample space S, of possible outcomes when a coin is flipped, may be written $S = \{H, T\}$, where H and T correspond to heads and tails, respectively.

Example 1 *: Consider the experiment of tossing a die. If we are interested in the number that shows on the top face, the sample space is $S_1 = \{1, 2, 3, 4, 5, 6\}$.*
If we are interested only in whether the number is even or odd, the sample space is simply $S_2 = \{even, odd\}$.

Example 2 *: An experiment consists of flipping a coin and then flipping it a second time if a head occurs. If a tail occurs on the first flip, then a die is tossed once. To list the elements of the sample space providing the most information, we construct the tree diagram of Figure 2.1. The sample space is* $S = \{HH, HT, T1, T2, T3, T4, T5, T6\}$.

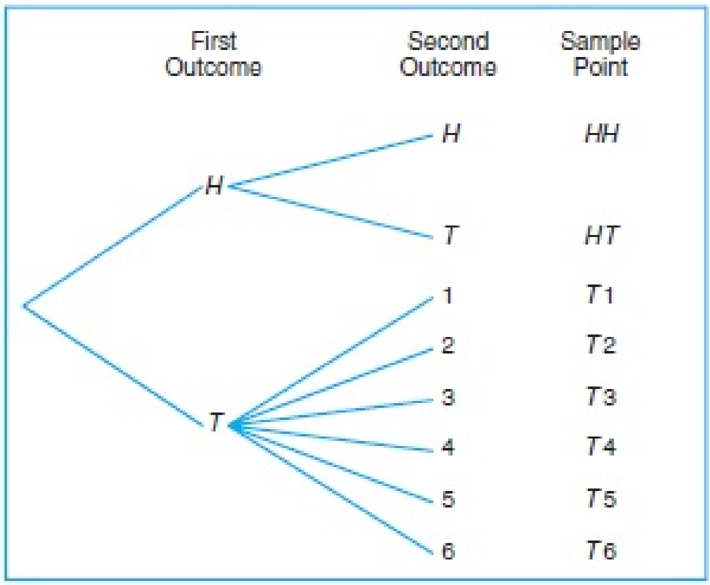

Figure 1: Tree diagram for Ex. 2

Example 3 *: Suppose that three items are selected at random from a manufacturing process. Each item is inspected and classified defective, D, or nondefective, N. To list the elements of the sample space providing the most information, we construct the tree diagram of Figure 2.2. The sample space is*

$$S = \{DDD, DDN, DND, DNN, NDD, NDN, NND, NNN\}$$

Suppose the experiment is to sample items randomly until one defective item is observed. The sample space for this case is $S = \{D, ND, NND, NNND, ...\}$

2.2 Events :

Definition 2 *An* ***event*** *is a subset of a sample space.*

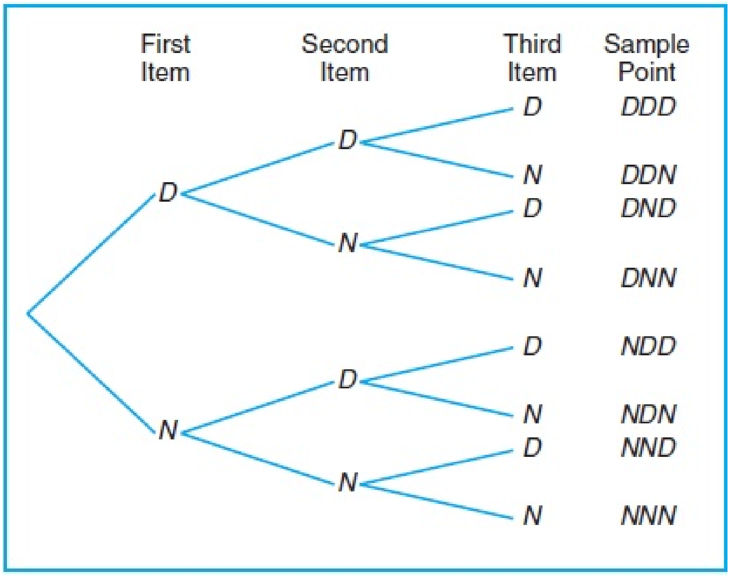

Figure 2: Tree diagram for Ex. 3

Example 4 *: The event A that the outcome when a die is tossed is divisible by 3. This will occur if the outcome is an element of the subset$S_1 = \{3, 6\}$ of the sample space S_1 in Example 1.*
In the event B that the number of defectives is greater than 1 in Example 3. This will occur if the outcome is an element of the subset $S = \{DDD, DDN, DND, NDD\}$ of the sample space S.

Definition 3 *The **complement** of an event A with respect to S is the subset of all elements of S that are not in A. We denote the complement of A by the symbol A'.*

Definition 4 *The **intersection** of two events A and B, denoted by the symbol $A \cap B$, is the event containing all elements that are common to A and B.*

Definition 5 *Two events A and B are **mutually exclusive**, or **disjoint**, if $A \cap B = \phi$, that is, if A and B have no elements in common.*

Definition 6 *The **union** of the two events A and B, denoted by the symbol $A \cup B$, is the event containing all the elements that belong to A or B or both.*

Exercises :

3. Which of the following events are equal?

(a) $A = \{1, 3\}$

(b) $B = \{x \mid x \textit{ is a number on a die}\}$

(c) $C = \{x \mid x^2 - 4x + 3 = 0\}$

(d) $D = \{x \mid x \textit{ is the number of heads when six coins are tossed}\}$

Solution :

(a) $A = \{1, 3\}$

(b) $B = \{1, 2, 3, 4, 5, 6\}$

(c) $C = \{x \mid x^2 - 4x + 3 = 0\} = \{x \mid (x-1)(x-3) \ = \ 0\} = \{1, 3\}$

(d) $D = \{0, 1, 2, 3, 4, 5, 6\}$

Clearly, A = C.

7. Four students are selected at random from a chemistry class and classified as male or female. List the elements of the sample space S_1, using the letter M for male and F for female. Define a second sample space S_2 where the elements represent the number of females selected.
Solution:

$$S_1 = \{MMMM, MMMF, MMFM, MFMM, FMMM, MMFF, MFMF, MFFM, \\ FMFM, FFMM, FMMF, MFFF, FMFF, FFMF, FFFM, FFFF\}$$

$$S_2 = \{0, 1, 2, 3, 4\}$$

2.3 Counting Sample Points :

Multiplication Rule : If an operation can be performed in n_1 ways, and if for each of these ways a second operation can be performed in n_2 ways, then the two operations can be performed together in $n_1 n_2$ ways.

Example 5 *: How many sample points are there in the sample space when a pair of dice is thrown once?*
Solution : *The first die can land face-up in any one of $n_1 = 6$ ways. For each of these 6 ways, the second die can also land face-up in $n_2 = 6$ ways. Therefore, the pair of dice can land in $n_1 n_2 = (6)(6) = 36$ possible ways.*

Generalized Multiplication Rule : If an operation can be performed in n_1 ways, and if for each of these a second operation can be performed in n_2 ways, and for each of the first two a third operation can be performed in n_3 ways, and so forth, then the sequence of k operations can be performed in $n_1 n_2 ...... n_k$ ways.

Example 6 *: Sam is going to assemble a computer by himself. He has the choice of chips from two brands, a hard drive from four, memory from three, and an accessory bundle from five local stores. How many different ways can Sam order the parts?*
Solution : *Since* $n_1 = 2, n_2 = 4, n_3 = 3, and n_4 = 5,$ *there are* $n_l \times n_2 \times n_3 \times n_4 = 2 \times 4 \times 3 \times 5 = 120$ *different ways to order the parts.*

Definition 7 *A* ***permutation*** *is an arrangement of all or part of a set of objects.*

Notes :

- For any non-negative integer n, $n!$, called **n factorial**, is defined as $n! = n(n-1)(n-2).......(2)(1)$, with special case $0! = 1$.

- The number of permutations of n objects is $n!$.

- The number of permutations of n distinct objects taken r at a time is

$$^{n}P_{r} = \frac{n!}{(n-r)!}$$

.

Example 7 *: In one year, three awards (research, teaching, and service) will be given to a class of 25 graduate students in a statistics department. If each student can receive at most one award, how many possible selections are there?*
Solution : *Since the awards are distinguishable, it is a permutation problem. The total number of sample points is*

$$^{25}P_{3} = \frac{25!}{(25-3)!} = \frac{25!}{22!} = (25)(24)(23) = 13,800$$

.

Notes :

- The number of permutations of n objects arranged in a circle is $(n-1)!$.
- The number of distinct permutations of n things of which n_1 are of one kind, n_2 of a second kind,....., n_k of a kth kind is

$$\frac{n!}{n_1!\; n_2!\; ....\; n_k!}$$

.

Example 8 ***:*** *In a college football training session, the defensive coordinator needs to have 10 players standing in a row. Among these 10 players, there are 1 freshman, 2 sophomores, 4 juniors, and 3 seniors. How many different ways can they be arranged in a row if only their class level will be distinguished?*

Solution : *We find that the total number of arrangements is*

$$\frac{10!}{1!\; 2!\; 4!\; 3!} = 12,600$$

.

Notes :

- The number of ways of partitioning a set of n objects into r cells with n_1 elements in the first cell, n_2 elements in the second, and so forth, is

$$\binom{n}{n_1,\; n_2, ..., n_r} = \frac{n!}{n_1!\; n_2!\; ....\; n_r!}, \; where \; n_1 + n_2 + ... + n_r = n$$

.

- The number of combinations of n distinct objects taken r at a time is

$$\binom{n}{r} = \frac{n!}{r!(n-r)!}$$

Lecture 3
2.4 Probability of an Event
2.5 Additive Rules

The aim of this lecture is to explain the following concepts :

- Probability of an Event.
- Additive Rules.

2.4 Probability of an Event :

Definition 1 *The probability of an event A is the sum of the weights of all sample points in A. Therefore,* $0 \leq P(A) \leq 1, P(\phi) = 0, and P(S) = 1$*. Furthermore, if* $A_1, A_2, A_3, ...$ *is a sequence of mutually exclusive events, then* $P(A_1 \cup A_2 \cup A_3 \cup ......) = P(A_1) + P(A_2) + P(A_3) + ...$.

Example 1 ***:*** *A coin is tossed twice. What is the probability that at least 1 head occurs ?*
Solution : *The sample space for this experiment is* $S = \{HH, HT, TH, TT\}$*. If the coin is balanced, each of these outcomes is equally likely to occur. If A represents the event of at least 1 head occurring, then* $A = \{HH, HT, TH\}$ *and* $P(A) = \frac{1}{4} + \frac{1}{4} + \frac{1}{4} = \frac{3}{4}$

Note : If an experiment can result in any one of N different equally likely outcomes, and if exactly n of these outcomes correspond to event A, then the probability of event A is $P(A) = \frac{n}{N}$.

2.5 Additive Rules :

Theorem 0.1 *If A and B are two events, then*

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

.

Corrolary 0.1 *If $A_1, A_2, ..., A_n$ are mutually exclusive, then*

$$P(A_1 \cup A_2 \cup .... \cup A_n) = P(A_1) + P(A_2) + ... + P(A_n)$$

.

Corrolary 0.2 *If $A_1, A_2, ..., A_n$ is a partition of sample space S, then*

$$P(A_1 \cup A_2 \cup .... \cup A_n) = P(A_1) + P(A_2) + ... + P(A_n) = P(S) = 1$$

.

Theorem 0.2 *For three events A, B, and C, then*

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$$

.

Theorem 0.3 *If A and A' are complementary events, then*

$$P(A) + P(A') = 1$$

.

Exercises :

50. An experiment involves tossing a pair of dice, one green and one red, and recording the numbers that come up. If x equals the outcome on the green die and y the outcome on the red die, Assuming that all elements of S are equally likely to occur. Let A be the event A that the sum is greater than 8; B be the event that a 2 occurs on either die; and C be the event that a number greater than 4 comes up on the green die. Find

(a) the probability of event A.

(b) the probability of event C.

(c) the probability of event $A \cap C$.

Solution :

(a) $P(A) = \frac{5}{18}$.

(b) $P(C) = \frac{1}{3}$.

(c) $P(A \cap C) = \frac{7}{36}$.

53. The probability that an American industry will locate in Shanghai, China, is 0.7, the probability that it will locate in Beijing, China, is 0.4, and the probability that it will locate in either Shanghai or Beijing or both is 0.8. What is the probability that the industry will locate

(a) in both cities?

(b) in neither city?

Solution : Consider the events
S: industry will locate in Shanghai.
B: industry will locate in Beijing.

(a) $P(S \cap B) = P(S) + P(B) - P(S \cup B) = 0.7 + 0.4 - 0.8 = 0.3$.

(b) $P(S' \cap B') = 1 - P(S \cup B) = 1 - 0.8 = 0.2$

58. A pair of fair dice is tossed. Find the probability of getting

(a) a total of 8.

(b) at most a total of 5.

Solution :

(a) Here $|S| = 36$
Let A be the event of obtaining a total of 8 = {(2,6), (3,5), (4,4), (5,3), (6,2)}. i.e. $|A| = 5$
Hence the probability of obtaining a total of 8 is P(A)= $\frac{5}{36}$.

(b) Ten of the 36 elements total at most 5. Hence the probability of obtaining a total of at most is $\frac{10}{36} = \frac{5}{18}$.

59. In a poker hand consisting of 5 cards, find the probability of holding

(a) 3 aces.

(b) 4 hearts and 1 club.

Solution :

(a) $P(3\ aces) = \frac{\binom{4}{3}\binom{48}{2}}{\binom{52}{5}} = \frac{94}{54145}$

(b) $P(4 hearts and 1 club) = \frac{\binom{13}{4}\binom{13}{1}}{\binom{52}{5}} = \frac{143}{39984}$

65. Let A be the event that the component fails a particular test and B be the event that the component displays strain but does not actually fail. Event A occurs with probability 0.20, and event B occurs with probability 0.35.

(a) What is the probability that the component does not fail the test?

(b) What is the probability that the component works perfectly well (i.e., neither displays strain nor fails the test)?

(c) What is the probability that the component either fails or shows strain in the test?

Solution : P(A) = 0.2 and P(B) = 0.35
(a) $P(A') = 1 - 0.2 = 0.8$
(b) $P(A' \cap B') = 1 - P(A \cup B) = 1 - 0.2 - 0.35 = 0.45$
(c) $P(A \cup B) = 0.2 + 0.35 = 0.55$.

68. Interest centers around the nature of an oven purchased at a particular department store. It can be either a gas or an electric oven. Consider the decisions made by six distinct customers.

(a) Suppose that the probability is 0.40 that at most two of these individuals purchase an electric oven. What is the probability that at least three purchase the electric oven?

(b) Suppose it is known that the probability that all six purchase the electric oven is 0.007 while 0.104 is the probability that all six purchase the gas oven. What is the probability that at least one of each type is purchased?

Solution : (a) 1-0.40 = 0.60.
(b) The probability that all six purchasing the electric oven or all six purchasing the gas oven is 0.007 + 0.104 = 0.111.
So the probability that at least one of each type is purchased is 1-0.111 = 0.889.

72. Prove that

$$P(A' \cap B') = 1 + P(A \cap B) - P(A) - P(B)$$

. **Solution :**
$P(A' \cap B')$
$= 1 - P(A \cup B)$
$= 1 - (P(A) + P(B) - P(A \cap B))$
$= 1 + P(A \cap B) - P(A) - P(B)$

Lecture 4
2.6 Conditional Probability, Independence, and the Product Rule

The aim of this lecture is to explain the following concepts :

- Conditional Probability.
- Independence.
- The Product Rule

Definition 1 *The* ***conditional probability*** *of B, given A, denoted by $P(B|A)$, is defined by*

$$(B|A) = \frac{P(A \cap B)}{P(A)} \quad , \textit{ provided } P(A) > 0$$

.

Definition 2 *Two events A and B are* ***independent*** *if and only if*

$$P(B|A) = P(B) \textit{ or } P(A|B) = P(A),$$

assuming the existences of the conditional probabilities. Otherwise, A and B are dependent.

The Product Rule, or the Multiplicative Rule :

Theorem 0.1 *If in an experiment the events A and B can both occur, then*

$$P(A \cap B) = P(A)P(B|A), \quad \textit{provided } P(A) > 0$$

.

Theorem 0.2 *Two events A and B are independent if and only if*

$$P(A \cap B) = P(A)P(B)$$

. Therefore, to obtain the probability that two independent events will both occur, we simply find the product of their individual probabilities.

Theorem 0.3 *If, in an experiment, the events $A_1, A_2, ..., A_k$ can occur, then*

$$P(A_1 \cap A_2 \cap .... \cap A_k) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2).....P(A_k|A_1 \cap A_2 \cap ..... \cap A_{k-1}).$$

If the events $A_1, A_2, ..., A_k$ are independent, then

$$P(A_1 \cap A_2 \cap .... \cap A_k) = P(A_1)P(A_2).....P(A_k).$$

Definition 3 *A collection of events $A = \{A_1, ..., A_n\}$ are* ***mutually independent*** *if for any subset of $A, A_{i1}, ..., A_{ik}$, for $k \leq n$, we have*

$$P(A_{i1} \cap ..... \cap A_{ik}) = P(A_{i1})......P(A_{ik}).$$

Exercises :

74. A class in advanced physics is composed of 10 juniors, 30 seniors, and 10 graduate students. The final grades show that 3 of the juniors, 10 of the seniors, and 5 of the graduate students received an A for the course. If a student is chosen at random from this class and is found to have earned an A, what is the probability that he or she is a senior?

Solution :

$P(S|A) = 10/18 = 5/9.$

75. A random sample of 200 adults are classified below by sex and their level of education attained.

Education	*Male*	*Female*
Elementary	38	45
Secondary	28	50
College	22	17

If a person is picked at random from this group, find the probability that

(a) the person is a male, given that the person has a secondary education;

(b) the person does not have a college degree, given that the person is a female.

Solution : Consider the events:
M: a person is a male;
S: a person has a secondary education;
C: a person has a college degree.

(a) $P(M|S) = \dfrac{28}{78} = \dfrac{14}{39}$.

(b) $P(C'|M') = \dfrac{95}{112}$.

77. In the senior year of a high school graduating class of 100 students, 42 studied mathematics, 68 studied psychology, 54 studied history, 22 studied both mathematics and history, 25 studied both mathematics and psychology, 7 studied history but neither mathematics nor psychology, 10 studied all three subjects, and 8 did not take any of the three. Randomly select a student from the class and find the probabilities of the following events.

(a) A person enrolled in psychology takes all three subjects.

(b) A person not taking psychology is taking both history and mathematics.

Solution :

(a) $P(M \cap P \cap H) = \dfrac{10}{68} = \frac{5}{34}$.

(b) $P(H \cap M|P') = \dfrac{P(H \cap M \cap P')}{P(P')} = \dfrac{22-10}{100-68} = \dfrac{12}{32} = \dfrac{3}{8}$.

80. The probability that an automobile being filled with gasoline also needs an oil change is 0.25; the probability that it needs a new oil filter is 0.40; and the probability that both the oil and the filter need changing is 0.14.

(a) If the oil has to be changed, what is the probability that a new oil filter is needed?

(b) If a new oil filter is needed, what is the probability that the oil has to be changed?

Solution : Consider the events:
C: an oil change is needed,
F: an oil filter is needed.

(a) $P(F|C) = \dfrac{P(F \cap C)}{P(C)} = \dfrac{0.14}{0.25} = 0.56.$

(b) $P(C|F) = \dfrac{P(C \cap F)}{P(F)} = \dfrac{0.14}{0.40} = 0.35.$

89. A town has two fire engines operating independently. The probability that a specific engine is available when needed is 0.96.

(a) What is the probability that neither is available when needed?

(b) What is the probability that a fire engine is available when needed?

Solution : Let A and B represent the availability of each fire engine.

(a) $P(A' \cap B') = P(A')P(B') = (0.04)(0.04) = 0.0016.$

(b) $P(A \cup B) = 1 - P(A' \cap B') = 1 - 0.0016 = 0.9984.$

91. Find the probability of randomly selecting 4 good quarts of milk in succession from a cooler containing 20 quarts of which 5 have spoiled, by using

(a) the first formula of Theorem 2.12 on page 68

(b) the formulas of Theorem 2.6 and Rule 2.3 on pages 50 and 54, respectively.

Solution :

(a) $P(Q_1 \cap Q_2 \cap Q_3 \cap Q_4) = P(Q_1)P(Q_2|Q_1)P(Q_3|Q_1 \cap Q_2)P(Q_4|Q_1 \cap Q_2 \cap Q_3) = (15/20)(14/19)(13/18)(12/17) = 91/323.$

(b) Let A be the event that 4 good quarts of milk are selected. Then $P(A) = \dfrac{\binom{15}{4}}{\binom{20}{4}} = \dfrac{91}{323}.$

Lecture 5
2.7 Bayes' Rule

The aim of this lecture is to explain the following concepts:

- Total Probability.
- Bayes' Rule

Theorem of total probability or the rule of elimination:

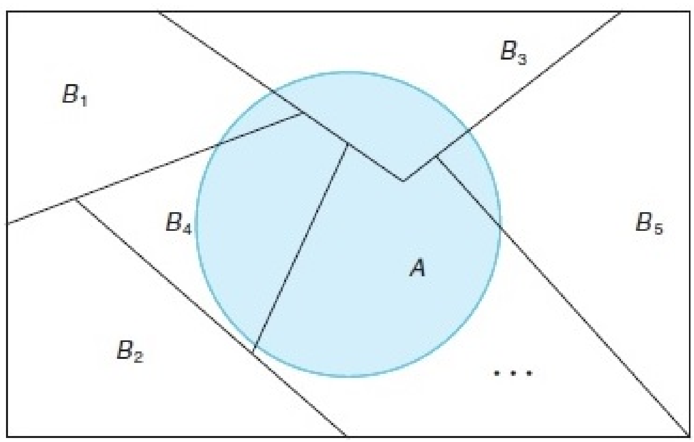

Figure 1: Partitioning the sample space S.

Theorem 0.1 *If the events $B_1, B_2, ..., B_k$ constitute a partition of the sample space S such that $P(B_i) \neq 0$ for i = 1,2,.....,k, then for any event A of S,*

$$P(A) = \sum_{i=1}^{k} P(B_i \cap A) = \sum_{i=1}^{k} P(B_i)P(A|B_i)$$

Theorem 0.2 *If the events $B_1, B_2, ..., B_k$ constitute a partition of the sample space S such that $P(B_i) \neq 0$ for i = 1,2,.....,k, then for any event A of S,*

$$P(B_r|A) = \frac{P(B_r \cap A)}{\sum_{i=1}^{k} P(B_i \cap A)} = \frac{P(B_r)P(A|B_r)}{\sum_{i=1}^{k} P(B_i)P(A|B_i)} \quad for\ r = 1, 2, ..., k.$$

Exercises:

95. In a certain region of the country it is known from past experience that the probability of selecting an adult over 40 years of age with cancer is 0.05. If the probability of a doctor correctly diagnosing a person with cancer as having the disease is 0.78 and the probability of incorrectly diagnosing a person without cancer as having the disease is 0.06, what is the probability that an adult over 40 years of age is diagnosed as having cancer?

Solution :

Consider the events:

C: an adult selected has cancer,

D: the adult is diagnosed as having cancer.

P(C) = 0.05,

$P(D|C) = 0.78$,

$P(C') = 0.95$

and $P(D|C') = 0.06$.

So, $P(D) = P(C \cap D) + P(C' \cap D)$

= (0.05)(0.78) + (0.95)(0.06)

= 0.096.

96. Police plan to enforce speed limits by using radar traps at four different locations within the city limits. The radar traps at each of the locations L1, L2, L3, and L4 will be operated 40%, 30%, 20% ,and 30% of the time. If a person who is speeding on her way to work has probabilities of 0.2, 0.1, 0.5,

and 0.2, respectively, of passing through these locations, what is the probability that she will receive a speeding ticket?
Solution : Let $S_1, S_2, S_3,$ *and* S_4 represent the events that a person is speeding as he passes through the respective locations and
let R represent the event that the radar traps is operating resulting in a speeding ticket.
Then the probability that he receives a speeding ticket:
$P(R) = \sum_{i=1}^{4} P(R|S_i)P(S_i)$
$= (0.4)(0.2) + (0.3)(0.1) + (0.2)(0.5) + (0.3)(0.2)$
$= 0.27.$

97. Referring to Exercise 2.95, what is the probability that a person diagnosed as having cancer actually has the disease?
Solution : $P(C|D) = \dfrac{P(C \cap D)}{P(D)} = \dfrac{0.039}{0.096} = 0.40625.$

98. If the person in Exercise 2.96 received a speeding ticket on her way to work, what is the probability that she passed through the radar trap located at L2?
Solution : $P(S_2|R) = \dfrac{P(R \cap S_2)}{P(R)} = \dfrac{0.03}{0.27} = 1/9.$

RANDOM VARIABLE AND PROBABILITY DISTRIBUTIONS

LECTURE-6

1 Concept of Random Variable:-

Definition 1.1. A Random variable is a function that associates a real number with each element in the sample space.

Example 1.1.
Two balls are drawn in succession without replacement from an urn containing 4 red balls and 3 black balls. The possible outcomes and the values y of the random variable Y , where Y is the number of red balls, are

Sample Space	y
RR	2
RB	1
BR	1
BB	0

Definition 1.2. The random variable for which 0 and 1 are chosen to describe the two possible values is called a **Bernoulli random variable**.

Example 1.2.
Consider the simple condition in which components are arriving from the production line and they are stipulated to be defective or not defective. Define the random variable X by

$$X = \begin{cases} 1, & \text{if the component is defective,} \\ 0, & \text{if the component is not defective.} \end{cases}$$

Definition 1.3. If a sample space contains a finite number of possibilities or an un-ending sequence with as many elements as there are whole numbers, it is called a **discrete sample space**, (i.e. The set of possible outcomes is countable).

Definition 1.4. If a sample space contains an infinite number of possibilities equal to the number of points on a line segment, it is called a **continuous sample space**,(i.e. The set of possible outcomes is uncountable).

2 Discrete Probability Distributions:-

Definition 2.1. The set of ordered pairs $(x, f(x))$ is a **probability function, probability mass function, or probability distribution of the discrete random variable** X if, for each possible outcome x,

1. $f(x) \geq 0$,

2. $\sum_{x} f(x) = 1$,

3.$P(X = x) = f(x)$.

Definition 2.2. The cumulative distribution function $F(x)$ of a discrete random variable X with probability distribution $f(x)$ is
$F(x) = P(X \leq x) = \sum_{t \leq x} f(t), \text{for} -\infty < x < \infty$.The probability distribution function(pdf) can be obtained by directly the cumulating distribution function(cdf) provided 'X' is a continuous random variable.

3 Continuous Probability Distributions

Definition 3.1.

The function $f(x)$ is a probability density function (pdf) for the continuous random variable X, defined over the set of real numbers, if

1.$f(x) \geq 0,$ for all $x \in \mathbb{R}$,

2. $\int_{-\infty}^{\infty} f(x)dx = 1$,

3.$P(a < X < b) = \int_{a}^{b} f(x)dx.$

Definition 3.2. The **cumulative distribution function** $F(x)$ of a continuous random variable X with density function $f(x)$ is

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt, \; for - \infty < x < \infty,$$

where $P(a < X < b) = F(b) - F(a)$ and $f(x) = \frac{dF(x)}{dx}$.

LECTURE-7

Problem 3.3. 7).The total number of hours, measured in units of 100 hours, that a family runs a vacuum cleaner over a period of one year is a continuous random variable X that has the density function

$$f(x) = \begin{cases} x, & 0 < x < 1, \\ 2 - x, & 1 \leq x < 2, \\ 0, & \text{elsewhere} . \end{cases}$$

Find the probability that over a period of one year, a family runs their vacuum cleaner
(a) less than 120 hours,
(b) between 50 and 100 hours.
sols:-a)

$$\begin{aligned} P(X < 1.2) &= \int_{-\infty}^{1.2} f(x)dx \\ &= \int_{-\infty}^{0} f(x)dx + \int_{0}^{1} f(x)dx + \int_{1}^{1.2} f(x)dx \\ &= 0 + \int_{0}^{1} xdx + \int_{1}^{1.2} (2 - x)dx \\ &= \frac{1}{2} + 2(1.2 - 1) - \frac{1}{2}((1.2)^2 - 1) \\ &= \frac{1}{2} + 0.4 - 0.22 \\ &= 0.68. \end{aligned}$$

b)

$$\begin{aligned} P(0.5 < X < 1) &= \int_{0.5}^{1} x dx \\ &= \frac{x^2}{2}\bigg]_{0.5}^{1} \\ &= \frac{1}{2}(1 - 0.25) \\ &= 0.375. \end{aligned}$$

Problem 3.4. 10). Find a formula for the probability distribution of the random variable X representing the outcome when a single die is rolled once.
solution:-

$$P(X = x) = \begin{cases} \frac{1}{6}, & X = 1, 2, 3, 4, 5, 6, \\ 0, & elesewhere. \end{cases}$$

Problem 3.5. 11) A shipment of 7 television sets contains 2 defective sets. A hotel makes a random purchase of 3 of the sets. If x is the number of defective sets purchased by the hotel, find the probability distribution of X.
solution:-

$$f(x) = \frac{{}^{2}c_{x} \;\; {}^{5}c_{3-x}}{{}^{7}c_{3}}, \;\; x = 0, 1, 2.$$

When $x = 0, \;\; f(x) = \frac{2}{7}, \;\; x = 1, \;\; then f(x) = \frac{4}{7}$ similarly $x = 2, \;\; f(x) = \frac{1}{7}$.

Problem 3.6. 12). An investment firm offers its customers municipal bonds that mature after varying numbers of years, given that the cumulative distribution function of T, the number of years to maturity for a randomly selected bond, is

$$F(t) = \begin{cases} 0, & t < 1, \\ \frac{1}{4}, & 1 \leq t < 3, \\ \frac{1}{2}, & 3 \leq t < 5, \\ \frac{3}{4}, & 5 \leq t < 7, \\ 1, & t \geq 7. \end{cases}$$

find (a) $P(T = 5)$
(b) $P(T > 3)$
(c) $P(1.4 < T < 6)$

(d) $P(T \leq 5 \mid T \geq 2)$

solutions:- a) $P(T = 5) = F(5) - F(4) = \frac{3}{4} - \frac{1}{2} = \frac{1}{4}$

b) $P(T > 3) = 1 - P(T \leq 3) = 1 - F(3) = 1 - \frac{1}{2} = \frac{1}{2}$

c)$P(1.4 < T < 6) = F(6) - F(1.4) = \frac{3}{4} - \frac{1}{4} = \frac{1}{2}$

d)$P(T \leq 5 \mid T \geq 2) = \frac{P(2 \leq T \leq 5)}{P(T \geq 2)} = \frac{F(5)-F(2)}{1-F(2)} = \frac{\frac{3}{4}-\frac{1}{4}}{1-\frac{1}{4}} = \frac{1}{2} * \frac{4}{3} = \frac{2}{3}$.

Problem 3.7. 14). The waiting time, in hours, between successive speeders spotted by a radar unit is a continuous random variable with cumulative distribution function

$$F(x) = \begin{cases} 0, & x < 0, \\ 1 - e^{-8x}, & x \geq 0. \end{cases}$$

Find the probability of waiting less than 12 minutes between successive speeders

(a) using the cumulative distribution function of X;

(b) using the probability density function of X.

solutions:- a)$P(X < 0.2) = F(0.2) = 1 - e^{-8*0.2} = 1 - e^{-1.6} = 0.7981$.

b)$f(x) = F'(x) = 8 * e^{-8x}$

$P(X < 0.2) = 8 * \displaystyle\int_0^{0.2} e^{-8x}dx = -e^{-8x}\Big]_0^{0.2} = 0.7981$.

Problem 3.8. 21). Consider the density function

$$f(x) = \begin{cases} k\sqrt{x}, & 0 < x < 1, \\ 0, & \text{elsewhere} . \end{cases}$$

(a) Evaluate k.

(b) Find $F(x)$ and use it to evaluate $P(0.3 < X < 0.6)$

solutions:- a)$\int_{-\infty}^{\infty} f(x)dx = 1$

$\Rightarrow \int_0^1 k\sqrt{x}dx = 1$

$\Rightarrow k\frac{x^{\frac{3}{2}}}{\frac{3}{2}}\Big]_0^1 = 1$

$\Rightarrow \frac{2k}{3} = 1$

$\Rightarrow k = \frac{3}{2}$

b) $F(x) = \int_{-\infty}^{x} f(t)dt$

$= \int_{-\infty}^{0} 0dt + \int_0^x \frac{3}{2}\sqrt{t}dt = x^{\frac{3}{2}}$, **where**

$$F(x) = \begin{cases} 0, \ x < 0, \\ x^{\frac{3}{2}}, \ 0 \le x < 1, \\ 1, \ x \ge 1. \end{cases}$$

$P(0.3 < X < 0.6) = F(0.6) - F(0.3) = (0.6)^{\frac{3}{2}} - (0.3)^{\frac{3}{2}} = 0.3004.$

Problem 3.9. 29).3.29 An important factor in solid missile fuel is the particle size distribution. Significant problems occur if the particle sizes are too large. From production data in the past, it has been determined that the particle size (in micrometers) distribution is characterized by

$$f(x) = \begin{cases} 3x^{-4}, & x > 1, \\ 0, & \text{elsewhere} . \end{cases}$$

(a) Verify that this is a valid density function.
(b) Evaluate $F(x)$.
(c) What is the probability that a random particle from the manufactured fuel exceeds4 micrometers?
solution:- a) $f(x) \ge 0$ and $\int_{-\infty}^{\infty} f(x)dx$
$= \int_1^{\infty} 3x^{-4}dx = \left.\frac{3x^{-3}}{-3}\right]_1^{\infty}$
$= -(\infty^{-3} - 1)$
$= 1$
Hence this is a valid valid density function for $x \ge 1$.
b)$F(x) = \int_{-\infty}^{x} f(t)dt$
$= \int_{-\infty}^{1} 0dt + \int_1^x 3t^{-4}dt$
$= 0 + \left.\frac{3t^{-3}}{-3}\right]_1^x$
$= 1 - x^{-3}$, **where**

$$F(x) = \begin{cases} 0, \ x < 1 \\ 1 - x^{-3}, \ x \ge 1. \end{cases}$$

c) $P(X > 4) = 1 - F(4) = 1 - (1 - 4^{-3}) = 4^{-3} = 0.0156.$

Problem 3.10. 30). Measurements of scientific systems are always subject to variation, some more than others. There are many structures for measurement error, and statisticians spend a great deal of time modeling these errors. Suppose the measurement error X of a certain physical quantity is decided by the density function

$$f(x) = \begin{cases} k\left(3 - x^2\right), & -1 \leq x \leq 1, \\ 0, & \text{elsewhere} . \end{cases}$$

(a) Determine k that renders $f(x)$ a valid density function.
(b) Find the probability that a random error in measurement is less than 1/2.
(c) For this particular measurement, it is undesirable if the magnitude of the error (i.e., $|x|$) exceeds 0.8. What is the probability that this occurs?

solutions:- a)$f(x) \geq 0$

$\int_{-\infty}^{\infty} f(x)dx = 1$
$\Rightarrow k\int_{-1}^{1}(3-x^2)dx = 1$
$\Rightarrow 2k\int_{0}^{1}(3-x^2)dx = 1$
$\Rightarrow 2k\left[3x - \frac{x^3}{3}\right]_0^1 = 1$
$\Rightarrow \frac{16}{3}k = 1$
$\Rightarrow k = \frac{3}{16}$.

b) for$-1 \leq x \leq 1$

$F(x) = \int_{-1}^{x} \frac{3}{16}(3-t^2)dt = \frac{3}{16}\left[3t - \frac{t^3}{3}\right]_{-1}^{x}$
$= \frac{3}{16}\left[3(x+1) - \frac{1}{3}(x^3+1)\right]$
$= \frac{3}{16}\left(3x + 3 - \frac{x^3}{3} - \frac{1}{3}\right)$
$= \frac{9}{16}x - \frac{x^3}{16} + \frac{1}{2}$
$P(X < \frac{1}{2}) = F(\frac{1}{2}) = \frac{1}{2} + \frac{9}{16}\frac{1}{2} - \frac{1}{16}\frac{1}{8}$
$= 0.773.$

c)$P(\mid X \mid > 0.8) = P(X < -0.8) + P(X > 0.8)$
$= F(-0.8) + 1 - F(0.8)$
$= \frac{1}{2} + \frac{9}{16} * (-0.8) - \frac{1}{16}(-0.8)^3 + 1 - \frac{1}{2} + \frac{9}{16} * 0.8 - \frac{1}{16}(0.8)^3$
$= 0.164.$

Problem 3.11. 35).Suppose it is known from large amounts of historical data that X, the number of cars that arrive at a specific intersection during a 20 -second time period, is characterized by the following discrete probability function:

$$f(x) = e^{-6}\frac{6^x}{x!}, \text{ for } x = 0, 1, 2, \ldots$$

(a) Find the probability that in a specific 20 -second time period, more than 8 cars arrive at the intersection.

(b) Find the probability that only 2 cars arrive.

solutions:- a)$P(X > 8) = 1 - P(X \leq 8)$

$= 1 - \sum_{x=0}^{8} e^{-6} \frac{6^x}{x!}$

$= 1 - \left(e^{-6} \frac{6^0}{0!} + e^{-6} \frac{e^1}{1!} + \cdots + e^{-6} \frac{6^8}{8!}\right)$

$= 0.1528.$

b) $P(X = 2) = f(2) = e^{-6} \frac{6^2}{2!} = 0.0446.$

LECTURE-8

4 Joint Probability Distributions

Definition 4.1. The function $f(x, y)$ is a joint probability distribution or probability mass function of the discrete random variable X and Y if
1. $f(x, y) \geq 0$ for all (x, y),
2. $\sum_x \sum_y f(x, y) = 1$,
3. $P(X = x, Y = y) = f(x, y)$.
For any region A in the xy-plane, $P[(X, Y) \in A] = \sum\sum_A f(x, y)$.

Definition 4.2. The function $f(x, y)$ is a joint density function of the continuous random variables X and Y if
1. $f(x, y) \geq 0$, for all (x, y),
2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)dxdy = 1$,
3. $P[(X, Y) \in A] = \int\int_A f(x, y)dxdy$. For any region A in the xy plane.

Definition 4.3. The marginal distributions of X alone and of Y alone are

$$g(x) = \sum_y f(x, y) \quad \text{and} \quad h(y) = \sum_x f(x, y)$$

for the discrete case, and

$$g(x) = \int_{-\infty}^{\infty} f(x, y)dy \quad \text{and} \quad h(y) = \int_{-\infty}^{\infty} f(x, y)dx$$

for the continuous case.

Definition 4.4. Let X and Y be two random variables, discrete or continuous. The conditional distribution of the random variable Y given that $X = x$ is

$$f(y \mid x) = \frac{f(x, y)}{g(x)}, \text{ provided } g(x) > 0$$

Similarly, the conditional distribution of X given that $Y = y$ is

$$f(x \mid y) = \frac{f(x, y)}{h(y)}, \text{ provided } h(y) > 0.$$

Definition 4.5. Let X and Y be two random variables, discrete or continuous, with joint probability distribution $f(x, y)$ and marginal distributions $g(x)$ and $h(y)$, respectively. The random variables X and Y are said to be **statistically independent** if and only if

$$f(x, y) = g(x)h(y)$$

for all (x, y) within their range.

Problem 4.6. 38. If the joint probability distribution is given by

$$f(x, y) = \frac{x+y}{30}, \qquad \text{for } x = 0, 1, 2, 3 \text{ and } y = 0, 1, 2 \text{ then}$$

find
(a) $P(X \leq 2, Y = 1)$.
(b) $P(X > 2, Y \leq 1)$.
(c) $P(X > Y)$.
(d) $P(X + Y = 4)$.
Solution:- a) $P(X \leq 2, Y = 1) = f(0, 1) + f(1, 1) + f(2, 1)$
$= \frac{1}{30} + \frac{2}{30} + \frac{3}{30} = \frac{1}{5}$.
b)$P(X > 2, Y \leq 1) = f(3, 0) + f(3, 1)$
$= \frac{3}{30} + \frac{4}{30} = \frac{7}{30}$.
c) $P(X > Y) = f(1, 0) + f(2, 0) + f(2, 1) + f(3, 0) + f(3, 1) + f(3, 2)$
$= \frac{1}{30} + \frac{2}{30} + \frac{3}{30} + \frac{3}{30} + \frac{4}{30} + \frac{5}{30}$
$= \frac{9}{15}$
d)$P(X + Y = 4) = f(2, 2) + f(3, 1) = \frac{4}{30} + \frac{4}{30} = \frac{4}{15}$.

Problem 4.7. 42. Let X and Y denote the lengths of life, in years, of two components in an electronic system. If the joint density function of these variables is

$$f(x, y) = \begin{cases} e^{-(x+y)}, & x > 0, y > 0, \\ 0, & \text{elsewhere .} \end{cases}$$

find $P(0 < X < 1 \mid Y = 2)$
Solution:- $f(x \mid y) = \frac{f(x,y)}{h(y)}$, provided $h(y) > 0$
$Now \;\; h(y) = \int_{-\infty}^{\infty} f(x, y)dx$
$= \int_0^{\infty} e^{-(x+y)}dx$
$= e^{-y} \int_0^{\infty} e^{-x}dx = e^{-y}\big[-e^{-x}\big]_0^{\infty}$

$= -e^{-y}(e^{-\infty} - e^{0})$
$= e^{-y} > 0, \ y > 0.$
$f(x \mid y) = \frac{e^{-(x+y)}}{e^{-y}} = e^{-x}$
$P(0 < X < 1 \mid Y = 2) = \int_0^1 e^{-x} dx \ , x > 0$
$= -e^{-x}\Big]_0^1$
$= -(e^{-1} - 1)$
$= 1 - \frac{1}{e}.$

LECTURE-9

Problem 4.8. 49). Let X denote the number of times a certain numerical control machine will malfunction: $1, 2$, or 3 times on any given day. Let Y denote the number of times a technician is called on an emergency call. Their joint probability distribution is given as

$f(x,y)$		x = 1	2	3
	1	0.05	0.05	0.10
	3	0.05	0.10	0.35
y	5	0.00	0.20	0.10

(a) Evaluate the marginal distribution of X.
(b) Evaluate the marginal distribution of Y.
(c) Find $P(Y = 3 \mid X = 2)$

Solution:- a).
We have $g(x) = \sum_y f(x,y)$,
where $g(1) = \sum_y f(1,y) = f(1,1) + f(1,3) + f(1,5)$
$= 0.05 + 0.05 + 0.00 = 0.1$
$g(2) = \sum_y f(2,y) = f(2,1) + f(2,3) + f(2,5)$
$= 0.05 + 0.10 + 0.20 = 0.35$
$g(3) = \sum_y f(3,y) = f(3,1) + f(3,3) + f(3,5)$
$= 0.10 + 0.35 + 0.10 = 0.55$
Here these are the marginal distributions of X.
b)
We have $h(y) = \sum_y f(x,y)$
$h(1) = f(1,1) + f(2,1) + f(3,1) = 0.05 + 0.05 + 0.10 = 0.20,$

$h(3) = f(1,3) + f(2,3) + f(3,3) = 0.05 + 0.10 + 0.35 = 0.50,$
$h(5) = f(1,5) + f(2,5) + f(3,5) = 0.00 + 0.20 + 0.10 = 0.30,$
Here these are the marginnal distribution of Y
c)we have $f(y \mid x) = \frac{f(x,y)}{g(x)}$, provided $g(x) > 0$
Hence $P(Y = 3 \mid X = 2) = \frac{f(2,3)}{g(2)} = \frac{0.10}{0.35} = \frac{10}{35} = \frac{2}{7.}$

Problem 4.9. 56). The joint density function of the random variables X and Y is

$$f(x,y) = \begin{cases} 6x, & 0 < x < 1, 0 < y < 1 - x, \\ 0, & \text{elsewhere} \end{cases}$$

(a) Show that X and Y are not independent.
(b) Find $P(X > 0.3 \mid Y = 0.5)$.
Solutions:- a) we have the X and Y are independent random variable if $f(x,y) = g(x)h(y)$ otherwise not independent

Now $g(x) = \int_{-\infty}^{\infty} f(x,y)dy = \int_0^{1-x} 6xdy = 6x\big[y\big]_0^{1-x} = \begin{cases} 6x(1-x), \ 0 < x < 1, \\ 0, \ elsewhere. \end{cases}$

$h(y) = \int_{-\infty}^{\infty} f(x,y)dx = \int_0^1 6xdx = 6\frac{x^2}{2}\big]_0^1 = \begin{cases} 3, \ 0 < y < 1 - x, \\ 0, \ elsewhere. \end{cases}$

as $g(x)h(y) = 6x(1-x) * 3 = 18x(1-x) \neq f(x,y)$
Hence X and Y are not independent.
b)we have $f(x \mid y) = \frac{f(x,y)}{h(y)}, \ provided \ h(y) > 0$
$= \frac{6x}{3} = 2x,$
$P(X > 0.3 \mid Y = 0.5) = \int_{0.3}^1 2xdx = 2\frac{x^2}{2}\big]_{0.3}^1 = 1 - 0.09 = 0.91$

LECTURE - 10

CHEPTER-4

4.1 Mean of random variable

If two coins are tossed 16 times and X is the number of heads that occur per toss, then the values of X are 0, 1, and 2. Suppose that the experiment yields no heads, one head, and two heads a total of 4, 7, and 5 times, respectively. The average number of heads per toss of the two coins is then

$$\frac{(0)(4)+(1)(7)+(2)(5)}{16} = 1.06$$

This can be written as

$$(0)(\frac{4}{16}) + (1)((\frac{7}{16}) + (2)((\frac{5}{16}) = 1.06$$

Here $\frac{4}{16}, \frac{7}{16}, and \frac{5}{16}$ are probabilities of getting 0 head, one head, two heads in tossing of two coin respectively . This average value is the mean of the random variable X or the mean ofthe probability distribution of X and write it as μ_x or simply as μ.

It is also common among statisticians to refer to this mean as the mathematical expectation, or the expected value of the random variable X, and denote it as E(X).

Definition 4.1

Mathematical Expectation E(X)

Let X be a random variable with probability distribution f(x). The mean, or expected value of X is

$$\mu = \text{E(X)} = \sum_x xf(x)$$

if X is discrete

$$\int_{-\infty}^{\infty} xf(x)dx$$

if X is continuous.

Example-4.1

A lot containing 7 components is sampled by a quality inspector; the lot contains 4 good components and 3 defective components. A sample of 3 is taken by the inspector. Find the expected value of the number of good components in this sample.

Solution : Let X represent the number of good components in the sample. The probability distribution of X is

$$f(x) = \frac{\binom{4}{x}\binom{3}{3-x}}{\binom{7}{3}}, \text{ x=0,1,2,3}$$

So $f(0) = \frac{1}{35}, f(1) = \frac{12}{35}, f(2) = \frac{18}{35} and\ f(3) = \frac{4}{35}$

Therefore $E(X) = (0)(\frac{1}{35}) + (1)(\frac{12}{35}) + (2)(\frac{18}{35}) + (3)(\frac{4}{35}) = \frac{12}{7}$

Exercise-4.4
A coin is biased such that a head is three times as likely to occur as a tail. Find the expected number of tails when this coin is tossed twice.

Solution: Let X denotes the number of tails. So X takes the values 0,1,2. Here a head is three times as likely to occur as a tail.
So p(H)=$\frac{3}{4}$ and p(T)=$\frac{1}{4}$.

Now f(0)=$\frac{9}{16}$, f(1)=$\frac{6}{16}$ and f(2)=$\frac{1}{16}$.

Therefore E(X)=(0)($\frac{9}{16}$)+(1)($\frac{6}{16}$)+(2)($\frac{1}{16}$)=$\frac{1}{2}$

Exercise-4.7
By investing in a particular stock, a person can make a profit in one year of \$4000 with probability 0.3 or take a loss of \$1000 with probability 0.7. What is this person's expected gain?
solution:
Let the profit variable is X

The person's expected gain
$E(X) = \sum_x xf(x) = (4000)(0.3) + (1000)(0.7) = \$1900$

Theorem-4.1
Let X be a random variable with probability distribution f(x). The expected value of the random variable g(X) is

$$\mu_g(X) = \text{E[g(X)]} = \sum_x g(x)f(x)$$

if X is discrete, and

$$\mu_g(X) = \text{E[g(X)]} = \int_{-\infty}^{\infty} g(x)f(x)dx$$

if X is continuous.

Example-4.4
Suppose that the number of cars X that pass through a car wash between

4:00 P.M. and 5:00 P.M. on any sunny Friday has the following probability distribution:

X	4	5	6	7	8	9
P(X=x)	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{6}$	$\frac{1}{6}$

Let g(X)=2X−1 represent the amount of money, in dollars, paid to the attendant by the manager. Find the attendant's expected earnings for this particular time period.
Solution : The attendant can expect to receive
E(g(X))=E(2X-1)=$\sum_{x=4}^{9}(2x-1)f(x)$

$$=(7)(\tfrac{1}{12} + (9)(\tfrac{1}{12}) + (11)(\tfrac{1}{4}) + (13)(\tfrac{1}{4}) + (15)(\tfrac{1}{6}) + (17)(\tfrac{1}{6})$$

$$= \$12.67$$

Exercise-4.12
If a dealer's profit, in units of \$5000, on a newautomobile can be looked upon as a random variable X having the density function

$$\text{f(x)=}\begin{cases} 2(1-x) & 0 \le x \le 1 \\ 0 & elsewhere \end{cases}$$

Find the average profit per automobile.
Solution : E(X)=$\int_0^1 xf(x)dx = x^2 - \frac{2x^3}{3}\Big|_0^1 = \frac{1}{3}$
The average profit per automobile($\frac{1}{3}$)(5000)=\$ $\frac{5000}{3}$

Definition 4.2
Let X and Y be random variables with joint probability distribution f(x, y). The mean, or expected value, of the random variable g(X, Y) is

$$\mu_{g(X,Y)} = E[g(X,Y)] = \sum_x \sum_y g(x,y)f(x,y)$$

if X and Y are discrete and

$$\mu_{g(X,Y)} = E[g(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)f(x,y)dxdy$$

if X and Y are continuous

Exercise-4.10
Two tire-quality experts examine stacks of tires and assign a quality rating to each tire on a 3-point scale. Let X denote the rating given by expert A and Y denote the rating given by B. The following table gives the joint distribution for X and Y .

	f(x,y)	y: 1	2	3
	1	0.10	0.05	0.02
x	2	0.10	0.35	0.05
	3	0.10	0.35	0.05

Solution :

	f(x,y)	y: 1	2	3	row total g(x)
	1	0.10	0.05	0.02	0.17
x	2	0.10	0.35	0.05	0.50
	3	0.03	0.10	0.20	0.33
column total	h(y)	0.23	0.50	0.27	1

$\mu_X = \sum_x xg(x) = (1)(0.17) + (2)(0.50) + (3)(0.33) = 2.16$
$\mu_Y = \sum_y yh(y) = (1)(0.23) + (2)(0.50) + (3)(0.27) = 2.04$

Exercise-4.20
A continuous random variable X has the density function

$$f(x)=\begin{cases} e^{-x} & x > 1 \\ 0 & elsewhere \end{cases}$$

Find the expected value of g(X) $= e^{\frac{2X}{3}}$
Solution :
E(g(X))=$\int_1^\infty g(x)f(x)dx$

$$= \int_1^\infty (e^{\frac{2x}{3}})(e^{-x})dx = \int_1^\infty e^{\frac{-x}{3}}dx = -3(e^{\frac{-x}{3}})\Big|_1^\infty$$

$$= 3e^{\frac{-1}{3}}$$

Exercise-4.23
Suppose that X and Y have the following joint probability function:

	f(x,y)	x: 2	4
	1	0.10	0.15
y	3	0.20	0.30
	5	0.10	0.15

(a) Find the expected value of g(X, Y) = XY^2.
(b) Find μ_X and μ_Y.
Solution :

	f(x,y)	x: 2	x: 4	Row total h(y)
y	1	0.10	0.15	0.25
	3	0.20	0.30	050
	5	0.10	0.15	0.25
Column total	g(x)	0.40	0.60	1

(a) E(g(X,Y))=$\sum_x \sum_y g(x,y)f(x,y) = \sum_x \sum_y xy^2 f(x,y)$

=(2)(1)(0.10)+(4)(1)(0.15)+(2)(9)(.20)+(4)(9)(0.30)+(2)(25)(0.10)

+(4)(25)(0.15)

=0.20+0.60+3.60+10.80+5.00+15.00=35.20

(b)

$(\mu_X = \sum_x xg(x) = (2)(0.40) + (4)(0.60) = 3.20$

$\mu_Y = \sum_y yh(y) = (1)(0.25) + (3)(0.50) + (5)(0.25) = 3$

*****Completed*****

LECTURE - 11

CHEPTER-4

4.2 Variance and Covariance of random variables

The most important measure of variability of a random variable X is the variance of the random variable X or the variance of the probability distribution of X and is denoted by Var(X) or the symbol σ_X^2, or simply by σ^2

Definition 4.3

Let X be a random variable with probability distribution f(x) and mean μ. The variance of X is

$\sigma^2 = E[(X-\mu)^2] = \sum_x (x-\mu)^2 f(x)$, if X is discrete, and

$\sigma^2 = E[(X-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx$, if X is continuous.

The positive square root of the variance σ is called the standard deviation of X.

Theorem-4.2:

The variance of a random variable X is $\sigma^2 = E(X^2) - \mu^2$.

Theorem-4.3:

Let X be a random variable with probability distribution f(x). The variance of the random variable g(X) is

$\sigma^2_{g(X)} = E[g(X) - \mu_{g(X)}]^2 = \sum_x [g(x) - \mu_{g(X)}]^2 f(x)$

if X is discrete, and

$\sigma^2_{g(X)} = E\{[g(X) - \mu_{g(X)}]^2\} = \int_{-\infty}^{\infty} [g(x) - \mu_{g(X)}]^2 f(x)dx$

if X is continuous.

Definition 4.4

Let X and Y be random variables with joint probability distribution f(x, y). The covariance of X and Y is

$\sigma_{XY} = E[(X-\mu_X)(Y-\mu_Y)] = \sum_x \sum_y (x-\mu_x)(y-\mu_y) f(x,y)$

if X and Y are discrete, and

$\sigma_{XY} = E[(X-\mu_X)(Y-\mu_Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-\mu_x)(y-\mu_y) f(x,y)dxdy$

if X and Y are continuous

Theorem-4.4:

The covariance of two random variables X and Y with means μ_X and μ_Y, respectively, is given by

$\sigma_{XY} = E(XY) - \mu_X \mu_Y$.

Example-4.9
Let the random variable X represents the number of defective parts for a machine when 3 parts are sampled from a production line and tested. The following is the probability distribution of X.

x	0	1	2	3
f(x)	0.51	0.38	0.10	0.01

Calculate σ^2.
Solution :
$\mu = (0)(0.51) + (1)(0.38) + (2)(0.10) + (3)(0.01) = 0.61.$

$E(X^2) = (0)(0.51) + (1)(0.38) + (4)(0.10) + (9)(0.01) = 0.87.$
Therefore,
$\sigma^2 = 0.87 - (0.61)^2 = 0.4979.$

Example-4.10
The weekly demand for a drinking-water product, in thousands of liters, from a local chain of efficiency stores is a continuous random variable X having the probability density

$$f(x)=\begin{cases} 2(x-1) & 1 < x < 2 \\ 0 & elsewhere \end{cases}$$

Find the mean and variance of X.
Solution :
Calculating E(X) and$E(X^2)$,
we have $\mu = E(X) = 2\int_1^2 x(x-1)dx = \frac{5}{3}$
and
$E(X^2) = 2\int_1^2 x^2(x-1)dx = \frac{17}{6}$
Therefore,
$\sigma^2 = \frac{17}{6} - (\frac{5}{3})^2 = \frac{1}{18}.$

Exercise-4.34
Let X be a random variable with the following probability distribution:

x	-2	3	5
f(x)	0.3	0.2	0.5

Find the standard deviation of X.
Solution :

Calculating E(X) and $E(X^2)$,
we have
$\mu = (-2)(0.3) + (3)(0.2) + (5)(0.5) = 2.5$

$E(X^2) = (4)(0.3) + (9)(0.2) + (25)(0.5) = 15.5$
$Therefore,$
$\sigma^2 = 15.5 - 2.5^2 = 9.25.$
The standard deviation $\sigma = \sqrt{9.25} = 3.04138.$

Exercise-4.35
The random variable X, representing the number of errors per 100 lines of software code, has the following probability distribution:

x	2	3	4	5	6
f(x)	0.01	0.25	0.4	0.3	0.04

Solution :
E(X)=$\sum_x xf(x) = 2(0.01) + 3(0.25) + 4(0.4) + 5(0.3) + 6(0.04) = 4.11$

$E(X^2) = \sum_x x^2 f(x) = 4(0.01)+9(0.25)+16(0.4)+25(0.3)+36(0.04) = 17.63$

$\sigma^2 = E(X^2) - (E(X))^2 = 17.63 - (4.11)^2 = 0.7379$

Exercise-4.50
For a laboratory assignment, if the equipment is working, the density function of the observed outcome X is

$$\text{f(x)=}\begin{cases} 2(1-x) & 0 < x < 1 \\ 0 & elsewhere \end{cases}$$

Find the variance and standard deviation of X.
Solution :
Calculating E(X) and $E(X^2)$,
we have $\mu = E(X) = 2\int_0^1 x(1-x)dx = \frac{1}{3}$
and
$E(X^2) = 2\int_0^1 x^2(1-x)dx = \frac{1}{6}$
$Therefore,$
$\sigma^2 = \frac{1}{6} - (\frac{1}{3})^2 = \frac{1}{18}$ and $\sigma = \sqrt{\frac{1}{18}} = 0.2357$

*****Completed*****

LECTURE - 12 and 13

CHEPTER-4

4.3 Means and Variances of Linear Combinations of Random Variables

Theorem-4.5
If a and b are constants, then E(aX + b) = aE(X) + b.
Substituting a=0 ,we get E(b)=b and b=0 we get E(aX)=aE(X)

Example-4.17
Suppose that the number of cars X that pass through a car wash between 4:00 P.M. and 5:00 P.M. on any sunny Friday has the following probability distribution:

X	4	5	6	7	8	9
P(X=x)	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{6}$	$\frac{1}{6}$

Let g(X)=2X−1 represent the amount of money, in dollars, paid to the attendant by the manager. Find the attendant's expected earnings for this particular time period.
Solution :
we can write E(2X-1) = 2E(X)-1.
Now
$\mu = E(X) = \sum_{x=4}^{9} xf(x)$

$$=(4)(\tfrac{1}{12}) + (5)(\tfrac{1}{12}) + (6)(\tfrac{1}{4}) + (7)(\tfrac{1}{4}) + (8)(\tfrac{1}{6}) + (9)(\tfrac{1}{6}) = \tfrac{41}{6}.$$

Therefore,
$\mu_{2X-1} = (2)(\frac{41}{6}) - 1 = \$12.67$

Theorem-4.6
The expected value of the sum or difference of two or more functions of a random variable X is the sum or difference of the expected values of the functions. That is,
E[g(X) ± h(X)] = E[g(X)] ± E[h(X)].

Exercise-4.57
Let X be a random variable with the following probability distribution:

x	-3	6	9
f(x)	$\frac{1}{6}$	$\frac{1}{2}$	$\frac{1}{3}$

Find E(X) and $E(X^2)$ and then, using these values, evaluate$E[(2X+1)^2]$.
Solution :
$E(X) = (-3)(\frac{1}{6}) + (6)(\frac{1}{2}) + (9)(\frac{1}{3}) = 5.5$
$E(X^2) = (9)(\frac{1}{6}) + (36)(\frac{1}{2}) + (81)(\frac{1}{3}) = 46.5$
$E[(2X+1)^2] = E(4(X^2) + 4(X) + 1) = 4E(X^2) + 4E(X) + 1$
$= 4(46.5) + 4(5.5) + 1 = 209$

Theorem-4.7
The expected value of the sum or difference of two or more functions of the random variables X and Y is the sum or difference of the expected values of the functions. That is,
E[g(X, Y) ± h(X, Y)] = E[g(X, Y)] ± E[h(X, Y)].

Theorem-4.8
Let X and Y be two independent random variables.Then E(XY) = E(X)E(Y).
Let X and Y be two independent random variables. Then $\sigma_{XY} = 0$.

Theorem-4.9
If X and Y are random variables with joint probability distribution f(x, y) and a, b, and c are constants, then $\sigma^2_{aX+bY+c} = a^2\sigma^2_X + b^2\sigma^2_Y + 2ab\sigma_{XY}$
Setting b = 0, we see that $\sigma^2_{aX+c} = a^2\sigma^2_X = a^2\sigma^2$
Setting a = 1 and b = 0, we see that $\sigma^2_{X+c} = \sigma^2_X = \sigma^2$.
Setting b = 0 and c = 0, we see that $\sigma^2_{aX} = a^2\sigma^2_X = a^2\sigma^2$.
If X and Y are independent random variables, then $\sigma^2_{aX+bY} = a^2\sigma^2_X + b^2\sigma^2_Y$ and $\sigma^2_{aX-bY} = a^2\sigma^2_X + b^2\sigma^2_Y$

Example-4.22
If X and Y are random variables with variances $\sigma^2_X = 2 \text{and} \sigma^2_Y = 4$ and covariance $\sigma_{XY} = -2$, find the variance of the random variable Z = 3X-4Y+ 8.
Solution :
$\sigma^2_Z = \sigma^2_{3X-4Y+8} = \sigma^2_{3X-4Y}$

$$= 9\sigma^2_X + 16\sigma^2_Y - 24\sigma_{XY} = (9)(2) + (16)(4) - (24)(-2) = 130$$

Example-4.23
Let X and Y denote the amounts of two different types of impurities in a batch of a certain chemical product. Suppose that X and Y are independent random variables with variances $\sigma^2_X = 2$ and $\sigma^2_Y = 3$. Find the variance of the random variable Z = 3X-2Y+ 5.

Solution :
$\sigma_Z^2 = \sigma_{3X-2Y+5}^2 = \sigma_{3X-2Y}^2 = 9\sigma_X^2 + 4\sigma_Y = (9)(2) + (4)(3) = 30$

Exercise-4.58
The total time, measured in units of 100 hours, that a teenager runs her hair dryer over a period of one year is a continuous random variable X that has the density function

$$\text{f(x)}=\begin{cases} x & 0 < x < 1 \\ 2-x & 1 < x < 2 \\ 0 & elsewhere \end{cases}$$

Evaluate the mean of the random variable Y = $60X^2$ + 39X, where Y is equal to the number of kilowatt hours expended annually.
Solution:
E(Y)=E(60X^2 + 39X)=60E(X^2)+39E(X)
$E(X) = \int_{-\infty}^{\infty} xf(x)dx = \int_0^1 (x)(x)dx + \int_1^2 (x)(2-x)dx$

$$=\frac{x^3}{3}\Big|_0^1 + (2\frac{x^2}{2} - \frac{x^3}{3})\Big|_1^2 = 1$$

$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)dx = \int_0^1 (x^2)(x)dx + \int_1^2 (x^2)(2-x)dx$

$$=\frac{x^4}{4}\Big|_0^1 + (2\frac{x^3}{3} - \frac{x^4}{4})\Big|_1^2 = \frac{7}{6}$$

$E(Y) = (60)(\frac{7}{6}) + (39)(1) = 109$
Total time=(109)(100)=10900 hours.

Exercise-4.60
Suppose that X and Y are independent random variables having the joint probability distribution

		x	
	f(x,y)	2	4
	1	0.10	0.15
y	3	0.20	0.30
	5	0.10	0.15

Find (a) E(2X-3Y); (b) E(XY).
Solution:

	f(x,y)	x: 2	4	Row total h(y)
	1	0.10	0.15	0.25
y	3	0.20	0.30	0.50
	5	0.10	0.15	0.25
	Column total g(x)	0.40	0.60	1

E(2X-3Y)=2E(X)-3E(Y)=$2\sum_x xg(x) - 3\sum_y yh(y)$

$=2((2)(0.40) + (4)(0.60)) - 3((1)(0.25) + (3)(0.5) + (5)(0.25))$

$= 6.40 + 9 = 6.40 - 9 = -2.60$

$E(XY) = E(X)E(Y) = (\sum_x xg(x))(\sum_y yh(y)) = (3.20)(3) = 9.60$

4.4 Chebyshev's Theorem
The probability that any random variable X will assume a value within k standard deviations of the mean is at least $1 - 1/k^2$. That is,
$P(\mu - k\sigma < X < \mu + k\sigma) \geq 1 - 1/k^2$.

Example-4.23
A random variable X has a mean $\mu = 8$, a variance $\sigma^2 = 9$, and an unknown probability distribution. Find (a)P$(-4 < X < 20)$ (b) P($\mid (X - 8) \mid \geq 6$).
Solution:
(a)$P(-4 < X < 20)$

$=P[8 - (4)(3) < X < 8 + (4)(3)] \geq \frac{15}{16}$.

(b) $P(\mid (X - 8) \mid \geq 6)$

$=1 - P(\mid (X - 8) \mid < 6) = 1 - P(-6 < X - 8 < 6)$

$=1 - P[8 - (2)(3) < X < 8 + (2)(3)] \leq \frac{1}{4}$

Exercise-4.75
An electrical firm manufactures a 100-watt light bulb, which, according to specifications written on the package, has a mean life of 900 hours with a standard deviation of 50 hours. At most, what percentage of the bulbs fail to last even 700 hours? Assume that the distribution is symmetric about the mean.
Solution:

X is the random variable define life of the 100-watt bulb
Here μ=900 hours and $\sigma = 50$
To find the probability of$P(X \leq 700)$.
Given that the distribution is symmetric about the mean. According to Chebyshev's theorem
$P(\mu - k\sigma < X < \mu + k\sigma) \geq 1 - 1/k^2$.

$P(X \leq 700) = (0.5)(P(\mid X - 900 \mid \geq 200)) = (0.5)(1 - P(\mid X - 900 \mid \leq 200))$

$=(0.5)(1 - P(900 - (4)(50) < X < 900 + (4)(50))) \leq (0.5)(\frac{1}{4^2}) = 0.03215$

Therefore the percentage of the bulbs fail to last even 700 hours is 3.215%.

Exercise-4.77
A random variable X has a mean$\mu = 10$ and a variance$\sigma^2 = 4$. Using Chebyshev's theorem, find
$(a)P(\mid X - 10 \mid \geq 3)$;
$(b)P((\mid X - 10 \mid < 3)$;
$(c)P(5 < X < 15)$;
(d) the value of the constant c such that $P(\mid X - 10 \mid \geq c) \leq 0.04$.
Solution:
(a)$P(\mid X - 10 \mid \geq 3$

$= 1 - P(\mid (X - 10) \mid < 3) = 1 - P(-3 < X - 10 < 3)$

$=1 - P[10 - (2)(\frac{3}{2}) < X < 10 + (2)(\frac{3}{2})] \leq \frac{4}{9}$

(b) $P((\mid X - 10 \mid < 3)$

$=P(-3 < X - 10 < 3) = P[10 - (2)(\frac{3}{2}) < X < 10 + (2)(\frac{3}{2})] \geq 1 - \frac{1}{(\frac{3}{2})^2} = \frac{5}{9}$

(c)$P(5 < X < 15)$

$=P[10 - (2)(\frac{5}{2}) < X < 10 + (2)(\frac{5}{2})] \geq 1 - \frac{1}{(\frac{5}{2})^2} = \frac{21}{25}$

(d)$P(\mid X - 10 \mid \geq c)$

$=1 - P(\mid X - 10 \mid \leq c) = 1 - P(10 - c < X < 10 + c)$

$= 1 - P(10 - 2(\frac{c}{2}) < X < 10 + 2(\frac{c}{2})) \leq \frac{1}{(\frac{c}{2})^2} = \frac{4}{c^2}$

Given that $P(\mid X - 10 \mid \geq c) \leq 0.04$.
Therefore $\frac{4}{c^2} = 0.04 \implies c^2 = 100 \implies c = 10$

Example-4.78
Compute$P(\mu - 2\sigma < X < \mu + 2\sigma)$, where X has the density function

$$\text{f(x)=}\begin{cases} 6x(1-x) & 0 < x < 1 \\ 0 & elsewhere \end{cases}$$

Solution:
E(X)=$\int_0^1 (x)(6x(1-x))dx = \int_0^1 (6x^2 - 6x^3)dx = \frac{1}{2}$

$E(X^2) = \int_0^1 (x^2)(6x(1-x))dx = \int_0^1 (6x^3 - 6x^4)dx = \frac{3}{10}$

$\sigma^2 = E(X^2) - (E(X))^2 = \frac{3}{10} - (\frac{1}{2})^2 = \frac{1}{20} \implies \sigma = \sqrt{(\frac{1}{20})} = 0.223$

$P(\mu - 2\sigma < X < \mu + 2\sigma) = P(0.053 < X < 0.9472) = \int_{0.053}^{0.947} 6x(1-x)dx = \int_{0.053}^{0.947} (6x - 6x^2)dx = 6(\frac{x^2}{2} - \frac{x^3}{3})\Big|_{0.053}^{0.9472} = 0.9838$

By Chebyshev's theorem,
$P(\mu - 2\sigma < X < \mu + 2\sigma) \geq 1 - \frac{1}{2^2} = 0.75$
$P(\mu - 2\sigma < X < \mu + 2\sigma) = 0.9838 \geq 0.75$
Hence Chebyshev's theorem is verified.

*****Completed*****

Lecture-14

Binomial Distribution

Bernoulli trails:

A Series of trails that satisfies the following assumptions is known as Bernoulli trail.

1. There are only two possible outcomes for each trail (success and failure)
2. The probability of success is same for each trail
3. The outcomes from different trails are independent

Example-1:

Tossing a Coin 100 times is a Bernoulli trail.

There are only 2 outcomes, Head and Tail. Here getting Head can be considered as success and Tail as failure. Probability of getting Head or success remains same though out the process and the events are independent.

Binomial distribution

Consider a Bernoulli trail which results in success with probability p and a failure with probability $q = 1 - p$.

Let X: the number of success in n trails.

Then the probability distribution of the binomial random variable X is

$$P(X = x) = f(x) = b(x; n, p) = \begin{pmatrix} n \\ x \end{pmatrix} p^x q^{n-x}, \qquad x = 0, 1, 2, ...n$$

Example-2:

If Probability of hitting the target is $\frac{3}{4}$ and three shots are fired, then

(i)Find the probability of hitting the target 2 times.

(ii)Formulate the binomial Probability distribution function

Ans:

Total no. of trails n=3.

Let X = no. of times hitting the target (no. of success), $x = 0, 1, 2, 3$

P (success) = P (hitting the target) $=\frac{3}{4}$

P (failure) = $\frac{1}{4}$

Therefore

(i) $P(X = 2) = f(2) = b\left(2; 3, \frac{3}{4}\right) = \begin{pmatrix} 3 \\ 2 \end{pmatrix} \left(\frac{3}{4}\right)^2 \left(\frac{1}{4}\right)^{3-2}$

(ii) $P(X = x) = f(x) = b\left(x; 3, \frac{3}{4}\right) = \begin{pmatrix} 3 \\ x \end{pmatrix} \left(\frac{3}{4}\right)^x \left(\frac{1}{4}\right)^{3-x}, x = 0, 1, 2, .3$

Note: The mean and variance of binomial distribution $f(x) = b(x; n, p)$ are $\mu = np$ and $\sigma^2 = npq$

(Q.11) The probability that a patient recovers from a delicate heart operation is 0.9. What is the probability that exactly 5 of the next 7 patients having this operation survive?

Ans:

Let X = no. of patients recovered from the heart operation i.e. $x = 0, 1, 2, ...7$

Here, $n = 7$, $p = 0.9$, $q = 0.1$

hence,

$$P(X = 5) = f(5) = b(5; 7, 0.9) = \begin{pmatrix} 7 \\ 5 \end{pmatrix} (0.9)^5 (0.1)^{7-5}$$

$$= 0.1240$$

Binomial distribution Table:

The cumulative distribution for the binomial distribution is pre-calculated and given in the form of a table.

Examples-3:(Use of binomial distribution Table)

We know

$$\begin{aligned} P(X \leq 4) &= F(4) \\ &= P(X=0)+P(X=1)+P(X=2)+P(X=3)+P(X=4) \end{aligned}$$

Suppose, $n = 5$, $p = 0.6$

Hence,

$$\begin{aligned} P(X \leq 4) &= F(4) \\ &= \begin{pmatrix} 5 \\ 0 \end{pmatrix} (0.6)^0 (0.4)^{5-0} + \begin{pmatrix} 5 \\ 1 \end{pmatrix} (0.6)^1 (0.4)^{5-1} + ... + \begin{pmatrix} 5 \\ 4 \end{pmatrix} (0.6)^4 (0.4)^{5-4} \\ &= \sum_{x=0}^{4} b(x; 5, 0.6) = 0.9222 \end{aligned}$$

$P(X \leq 4) = B(4; 5, 0.6) = 0.9222$ (from binomial distribution table)

In general $P(X \leq x) = B(x; n, p)$ and $b(x; n, p) = B(x; n, p) - B(x-1; n, p)$

(Q.15) It is known that 60% of mice inoculated with a serum are protected from a certain disease. If 5 mice are inoculated, find the probability that (a) none contracts the disease; (b) fewer than 2 contract the disease; (c) more than 3 contract the disease.

Ans:

Let X = no. of mice from the disease after inoculated, $x = 0, 1, 2, 3, 4, 5$

Here, $n = 5$, $p = 0.4$, $q = 0.6$

(i) $P(X = 0) = f(0) = \begin{pmatrix} 5 \\ 0 \end{pmatrix} (0.4)^0 (0.6)^5 = 0.0778$

(ii) $P(X < 2) = P(X \leq 1)$

$$\begin{aligned}
&= P(X = 0) + P(X = 1) \\
&= \begin{pmatrix} 5 \\ 0 \end{pmatrix} (0.4)^0 (0.6)^5 + \begin{pmatrix} 5 \\ 1 \end{pmatrix} (0.4)^1 (0.6)^{5-1} = 0.3370
\end{aligned}$$

(iii) $P(X > 3) = P(X = 4) + P(X = 5)$

$$= \begin{pmatrix} 5 \\ 4 \end{pmatrix} (0.4)^4 (0.6)^{5-4} + \begin{pmatrix} 5 \\ 5 \end{pmatrix} (0.4)^5 (0.6)^{5-5} = 0.087$$

(Q.16) Suppose that airplane engines operate independently and fail with probability equal to 0.4. Assuming that a plane makes a safe fight if at least one-half of its engines run, determine whether a 4-engine plane or a 2 engine plane has the higher probability for a successful fight.

Ans:

Case-1

The plane has 4 engines; $n = 4$

The plane will make a safe flight if 2 or more engines are working.

$$\begin{aligned}
P(X \geq 2) &= 1 - P(X \leq 1) \\
&= 1 - [P(X = 0) + P(X = 1)] \\
&= 1 - \left[\begin{pmatrix} 4 \\ 0 \end{pmatrix} (0.6)^0 (0.4)^{4-0} + \begin{pmatrix} 4 \\ 1 \end{pmatrix} (0.6)^1 (0.4)^{4-1} \right] = 0.8208
\end{aligned}$$

Case-2

The plane has 2 engines; $n = 2$

The plane will make a safe flight if 1 or more engines are working.

$$\begin{aligned}
P(X \geq 1) &= 1 - P(X = 0) \\
&= 1 - \left[\begin{pmatrix} 2 \\ 0 \end{pmatrix} (0.6)^0 (0.4)^{2-0} \right] = 0.84
\end{aligned}$$

Conclusion: comparing the above two cases, a 2 engine flight has higher probability for a successful flight.

Lecture-15

Multinomial Distribution

Multinomial distribution:

Consider a trial which results k outcomes, E_1, E_2, ...E_k with probabilitites p_1, p_2, ..., p_k respectively such that

$$\sum_{i=1}^{k} p_i = 1$$

Let

X_1 = no. of times E_1 occurs in n independent trials

X_2 = no. of times E_2 occurs in n independent trials

....

X_k = no. of times E_k occurs in n independent trials

Now,

$$P(X_1 = x_1, X_2 = x_2, ...X_k = x_k) = f(x_1, x_2, ...x_k; p_1, p_2, ...p_k)$$
$$= \left(\begin{array}{c} n \\ x_1, x_2, ...x_k \end{array} \right) {p_1}^{x_1} {p_2}^{x_2} .....{p_k}^{x_k}$$

With $\sum_{i=1}^{n} x_i = n, \quad \sum_{i=1}^{n} p_i = 1$

(Q.19) As a student drives to school, he encounters a trafic signal. This trafic signal stays green for 35 seconds, yellow for 5 seconds, and red for 60 seconds. Assume that the student goes to school each weekday between 8:00 and 8:30 a.m. Let X_1 be the number of times he encounters a green light, X_2 be the number of times he encounters a yellow light, and X_3 be the number of times he encounters a red light. Find the joint distribution of X_1, X_2, and X_3

Ans:

Let

X_1 = no. of times he encounters a green light

X_2 = no. of times he encounters a yellow light

X_3 = no. of times he encounters a red light

Given $p_1 = 0.35, p_2 = 0.05, p_3 = 0.60$

Therefore,

$$
\begin{aligned}
P(X_1 = x_1, X_2 = x_2, X_3 = x_3) &= f(x_1, x_2, .x_3; n, 0.35, 0.05, 0.60) \\
&= \begin{pmatrix} n \\ x_1, x_2, ...x_k \end{pmatrix} (0.35)^{x_1} (0.05)^{x_2} (0.60)^{x_3}
\end{aligned}
$$

Where $x_1 + x_2 + x_3 = n$

(Q.22) According to a genetics theory, a certain cross of guinea pigs will result in red, black, and white ofspring in the ratio 8:4:4. Find the probability that among 8 ofspring, 5 will be red, 2 black, and 1 white.

Ans:

Let

X_1 = no. of red guinea pigs

X_2 = no. of black guinea pigs

X_3 = no. of white guinea pigs

It is given that the ratio of red, black, and white guinea pigs is 8:4:4

Hence,

P (guinea pig is red) = 8/16 =0.5

P (guinea pig is black) = 4/16 =0.25

P (guinea pig is white) = 4/16 =0.25

$$
\begin{aligned}
P(X_1 = 5, X_2 = 2, X_3 = 1) &= f(5, 2, .1; 8, 0.5, 0.25, 0.25) \\
&= \begin{pmatrix} 8 \\ 5, 2, 1 \end{pmatrix} (0.5)^5 (0.25)^2 (0.25)^1 \\
&= 21/256
\end{aligned}
$$

Lecture-16

Hypergeometric Distribution

General discussion:

Suppose total number of items in a a bag $= N$

Total number of defective items (out of N) $= k$

Total number of items selected $= n$

lets discuss the probability that x out of n $(x \leq n)$ items selected is defective.

Now, total number of ways n items can be selected out of N items $= \begin{pmatrix} N \\ n \end{pmatrix}$.

Our requirement is x out of n are defective i.e. remaining $(n - x)$ are non-defective.

Number of ways x defective items can be selected from k defective items $= \begin{pmatrix} k \\ x \end{pmatrix}$.

Number of ways $n - x$ defective items can be selected from $N - k$ defective items $= \begin{pmatrix} N-k \\ n-x \end{pmatrix}$.

Probability of selecting x defectives

$$= \frac{all\ favorable\ cases}{all\ possible\ cases}$$

$$= \frac{\begin{pmatrix} k \\ x \end{pmatrix} \begin{pmatrix} N-k \\ n-x \end{pmatrix}}{\begin{pmatrix} N \\ n \end{pmatrix}}$$

Definition:

Let $X =$ The number of successes in a random sample size n selected from N items of which k are labeled success and $(N - k)$ labeled failure.Then probability distribution of the above hypergeometric random variable is

$$f(x) = h(x; N, n, k) = \frac{\begin{pmatrix} k \\ x \end{pmatrix} \begin{pmatrix} N-k \\ n-x \end{pmatrix}}{\begin{pmatrix} N \\ n \end{pmatrix}}$$

Such that $\max\{0, n-(N-k)\} \leq x \leq \min(n, k)$

(Q.30) A random committee of size 3 is selected from 4 doctors and 2 nurses. Write a formula for the probability distribution of the random variable X representing the number of doctors on the committee. Find $P(2 \leq X \leq 3)$.

Ans:

There are 4 doctors and 2 nurses

3 persons will be selected out of 4+2=6 persons

Let X: number of doctors in the committee which consists of 3 persons

So $x = 1, 2, 3$ ($x \neq 0$ why?)

$$P(X = x) = f(x) = \frac{\begin{pmatrix} 4 \\ x \end{pmatrix} \begin{pmatrix} 2 \\ 3-x \end{pmatrix}}{\begin{pmatrix} 6 \\ 3 \end{pmatrix}}$$

Now,

$P(2 \leq X \leq 3) = P(X = 2) + P(X = 3)$

$$= \frac{\begin{pmatrix} 4 \\ 2 \end{pmatrix} \begin{pmatrix} 2 \\ 3-2 \end{pmatrix}}{\begin{pmatrix} 6 \\ 3 \end{pmatrix}} + \frac{\begin{pmatrix} 4 \\ 3 \end{pmatrix} \begin{pmatrix} 2 \\ 3-3 \end{pmatrix}}{\begin{pmatrix} 6 \\ 3 \end{pmatrix}} = \frac{4}{5}$$

(Q.32) From a lot of 10 missiles, 4 are selected at random and fired. If the lot contains 3 defective missiles that will not fire, what is the probability that (a) all 4 will fire? (b) at most 2 will not fire?

Ans:

Total number of missiles = 10

Total number of defective missiles = 3

Hence, total number of non-defective missiles = 7

4 missiles will be fired

Let X: number of non-defective missiles fired

(a) $P(X=0) = \dfrac{\binom{7}{4}\binom{3}{0}}{\binom{10}{4}} = \frac{1}{6}$

(b) At most 2 will not fire means 2 or more will fire

$$\begin{aligned} P(X \geq 2) &= P(X=2) + P(X=3) + P(X=4) \\ &= \frac{\binom{7}{2}\binom{3}{2}}{\binom{10}{4}} + \frac{\binom{7}{3}\binom{3}{1}}{\binom{10}{4}} + \frac{\binom{7}{4}\binom{3}{0}}{\binom{10}{4}} = \frac{29}{30} \end{aligned}$$

Multivariate Hyper-geometric Distribution:

If N items can be partitioned into k cells $A_1, A_2, ...A_k$ with $a_1, a_2, ...a_k$elements, respectively then probability distribution of the random variables $X_1, X_2, ...X_k$representing the number of elements selected from $A_1, A_2, ...A_k$in a random sample of size n is

$$f(x_1, x_2, ...x_k; a_1, a_2, ...a_k, N, n) = \frac{\binom{a_1}{x_1}\binom{a_2}{x_2}.......\binom{a_k}{x_k}}{\binom{N}{n}}$$

With $\sum_{i=1}^{n} x_i = n, \ \sum_{i=1}^{n} a_i = N$

(Q.43) A foreign student club lists as its members 2 Canadians, 3 Japanese, 5 Italians, and 2 Germans. If a committee of 4 is selected at random, ?nd the probability that (a) all nationalities are represented; (b) all nationalities except Italian are represented.

Ans:

Total number of members = 2+3+5+2 12

Total number of members selected = 4

(a) All nationalities represented means one from each country.

One person can be selected from 2 Canadians in $\begin{pmatrix} 2 \\ 1 \end{pmatrix}$ ways

One person can be selected from 3 Japanies in $\begin{pmatrix} 3 \\ 1 \end{pmatrix}$ ways

One person can be selected from 5 Italians in $\begin{pmatrix} 5 \\ 1 \end{pmatrix}$ ways

One person can be selected from 2 Germans in $\begin{pmatrix} 2 \\ 1 \end{pmatrix}$ ways

Four persons can be selected from 12 persons in $\begin{pmatrix} 12 \\ 4 \end{pmatrix}$ ways

$$P(all\ nationalities\ are\ represented)$$
$$= \frac{\begin{pmatrix} 2 \\ 1 \end{pmatrix}\begin{pmatrix} 3 \\ 1 \end{pmatrix}\begin{pmatrix} 5 \\ 1 \end{pmatrix}\begin{pmatrix} 2 \\ 1 \end{pmatrix}}{\begin{pmatrix} 12 \\ 4 \end{pmatrix}} = \tfrac{4}{33}$$

(b) All nationalities except Italians are represented, then 3 cases arise;

Case-I: 2 Canadians + 1 Japanies+ 0 Italian + 1 German

Case-2: 1 Canadians + 2 Japanies+ 0 Italian + 1 German

Case-3: 1 Canadians + 1 Japanies+ 0 Italian + 2 German

$$P(all\ nationalities\ are\ represented)$$
$$= \frac{\begin{pmatrix} 2 \\ 2 \end{pmatrix}\begin{pmatrix} 3 \\ 1 \end{pmatrix}\begin{pmatrix} 5 \\ 0 \end{pmatrix}\begin{pmatrix} 2 \\ 1 \end{pmatrix}}{\begin{pmatrix} 12 \\ 4 \end{pmatrix}} + \frac{\begin{pmatrix} 2 \\ 1 \end{pmatrix}\begin{pmatrix} 3 \\ 2 \end{pmatrix}\begin{pmatrix} 5 \\ 0 \end{pmatrix}\begin{pmatrix} 2 \\ 1 \end{pmatrix}}{\begin{pmatrix} 12 \\ 4 \end{pmatrix}} + \frac{\begin{pmatrix} 2 \\ 1 \end{pmatrix}\begin{pmatrix} 3 \\ 1 \end{pmatrix}\begin{pmatrix} 5 \\ 0 \end{pmatrix}\begin{pmatrix} 2 \\ 2 \end{pmatrix}}{\begin{pmatrix} 12 \\ 4 \end{pmatrix}} = \tfrac{8}{165}$$

(Q.44) An urn contains 3 green balls, 2 blue balls, and 4 red balls. In a random sample of 5 balls, find the probability that both blue balls and at least 1 red ball are selected.

Ans:

Total number of balls = 3+2+4 = 9

Total number of balls selected = 5

Number of blue balls = 2, green balls = 3, red balls = 4

$$P(2\, blue\ balls\ and\ atleast\ 1\ red\ ball)$$
$$= \frac{\binom{2}{2}\binom{4}{1}\binom{3}{2}}{\binom{9}{5}} + \frac{\binom{2}{2}\binom{4}{2}\binom{3}{1}}{\binom{9}{5}} + \frac{\binom{2}{2}\binom{4}{3}\binom{3}{0}}{\binom{9}{5}} = \frac{17}{63}$$

Negative Binomial Distribution:

Let repeated independent trials results in a success with probability p and a failure with probability $q = 1 - p$.

Where X: number of trials in which the kth success occurs,

Then,

$$P(X = x) = f(x) = b^*(x; k, p) = \binom{x-1}{k-1} p^k q^{x-k}$$

Where $x = k, k+1, k+2....$

Geometric Distribution:

A particular case of negative binomial distribution for $k = 1$ is known as geometric distribution.

Here, X= the number of trials on which the first success occurs.

$$P(X = x) = f(x) = g(x; p) = pq^{x-1}; \quad x = 1, 2, 3........$$

Where $q = 1 - p$

(Q.49) The probability that a person living in a certain city owns a dog is estimated to be 0.3. Find the probability that the tenth person randomly interviewed in that city is the fifth one to own a dog.

Ans:

Here, $p = 0.3, q = 1 - p = 0.7$

X= The number of persons interviewed in which kth person own a dog.

Given $x = 10, k = 5$

$$\begin{aligned} b^*(10; 5, 0.3) &= \begin{pmatrix} 10-1 \\ 5-1 \end{pmatrix} p^5 q^{10-5} \\ &= \begin{pmatrix} 9 \\ 4 \end{pmatrix} (0.3)^5 \, (0.7)^{10-5} = 0.0515 \end{aligned}$$

(Q.50) Find the probability that a person flipping a coin gets (a) the third head on the seventh flip; (b) the first head on the fourth flip.

Ans:

Here, $p = 0.5, q = 1 - p = 0.5$

X= The number of trials in which kth head occurs.

(a)Third head in 7^{th} flip means $x = 7, k = 3$

Hence,

$$b^*(7; 3, 0.5) = \begin{pmatrix} 7-1 \\ 3-1 \end{pmatrix} (0.5)^3 (0.5)^4 = 0.1172$$

(b)First head in the fourth flip means$x = 4, k = 1$

Using negative binomial or geometric distribution

$$g(x; p) = g(4, 0.5) = 0.5(1 - 0.5)^{4-1} = (0.5)^4$$

Lecture-17

Poisson Distribution

Poisson Distribution:

The probability distribution of the Poisson random variable X, representing the number of outcomes occurring a given time interval t is

$$P(x;\lambda t) = \frac{e^{-\lambda t}\,(\lambda t)^x}{x!}, \qquad x = 0, 1, 2, ...$$

For $t = 1$

$$P(x;\lambda) = \frac{e^{-\lambda}\lambda^x}{x!}, \qquad x = 0, 1, 2, ...$$

The Poission random variable X, has the mean $\mu = \lambda$ and variance $\sigma^2 = \lambda$.

Note: For the purpose of minimizing calculation, refer Poisson distribution table in the problems.

(Q.58) A certain area of the eastern United States is, on average, hit by 6 hurricanes a year. Find the probability that in a given year that area will be hit by (a) fewer than 4 hurricanes; (b) anywhere from 6 to 8 hurricanes.

Ans:

The average number of hurricane hits in a year is 6 i.e. $\lambda = 6$

Let X : The number of hurricane hits in a year

a.

P(fewer than 4 hurricanes) means

$$P(X \leq 3) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3)$$
$$= \tfrac{e^{-6}6^0}{0!} + \tfrac{e^{-6}6^1}{1!} + \tfrac{e^{-6}6^2}{2!} + \tfrac{e^{-6}6^3}{3!} = 0.1512$$

b.

P(anywhere from 6 to 8 hurricanes) means

$$\begin{aligned} P(6 \le X \le 8) &= P(X=6) + P(X=7) + P(X=8) \\ &= \tfrac{e^{-6}6^6}{6!} + \tfrac{e^{-6}6^7}{7!} + \tfrac{e^{-6}6^8}{8!} = 0.4015 \\ &\quad Or \\ &= F(8) - F(5) = 0.4015 \ (using\ table) \end{aligned}$$

(Q.60) The average number of field mice per acre in a 5-acre wheat field is estimated to be 12. Find the probability that fewer than 7 field mice are found (a) on a given acre; (b) on 2 of the next 3 acres inspected.

Ans:

The average number of field mice per acre is 12 i.e. $\lambda = 12$

(a) Let X : The number of mice per acre

Hence,

$$P(X < 7) = P(X \le 6) = 0.0458 \ \ (using\ table)$$

(b) Let Y : The number of acres of land inspected.

Here Y follows binomial distribution.

Given $n = 3, y = 2$

Hence, $P(Y=2) = \begin{pmatrix} 3 \\ 2 \end{pmatrix} p^2 q^{3-2},\ \ with\ p = 0.0458$ (from part a), $q = 1 - p$

(Q.69) The probability that a person will die when he or she contracts a virus infection is 0.001. Of the next 4000 people infected, what is the mean number who will die?

Ans:

Given $p = 0.001,\ \ n = 4000$

Therefore, $\mu = np = 4000 \times 0.001 = 4$