UNIT II  PROBABILITY AND RANDOM VARIABLES

PART-I

Correlation, Correlation coefficient, Regression lines, Rank correlation Sample space and events, Probability, Axioms of Probability – conditional probability – total probability, Baye’s theorem

Prepared by
Dr. A.R. VIJAYALAKSHMI
Correlation:

The term correlation refers to the degree of relationship between two or more variables.

If a change in one variable effects a change in the other variable, the variables are said to be correlated.

There are basically three types of correlation, namely positive correlation, negative correlation and zero correlation.
Karl Pearson (1867-1936), a British Biometrician, developed the coefficient of correlation to express the degree of linear relationship between two variables.

Correlation coefficient between two random variables X and Y denoted by $r(X, Y)$, is given by

$$r(X, Y) = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}$$

Where

$$Cov(X, Y) = \frac{1}{n} \sum_{i} (X_i - \bar{X})(Y_i - \bar{Y})$$

(covariance between X and Y)

$$\sigma_X = \sqrt{\frac{1}{n} \sum_{i} (X_i - \bar{X})^2}$$

(standard deviation of X)

$$\sigma_Y = \sqrt{\frac{1}{n} \sum_{i} (Y_i - \bar{Y})^2}$$

(standard deviation of Y)
Karl Pearson correlation co-efficient

\[
r(X,Y) = \frac{1}{n} \sum_{i} (X_i - \bar{X})(Y_i - \bar{Y}) \quad \text{where} \quad (\bar{X}, \bar{Y}) = \left( \frac{1}{n} \sum_{i} X_i, \frac{1}{n} \sum_{i} Y_i \right)
\]

\[
= \frac{\sum xy}{\sqrt{\left( \sum x^2 \right) \left( \sum y^2 \right)}}
\]

**Note**

The following formula may also be used to compute correlation co-efficient between the two variables X and Y.

\[
(i) \ r(x, y) = \frac{N \sum XY - \sum X \sum Y}{\sqrt{N \sum X^2 - (\sum X)^2} \sqrt{N \sum Y^2 - (\sum Y)^2}}
\]

\[
(ii) \ r(x, y) = \frac{N \sum d_x d_y - \sum d_x \sum d_y}{\sqrt{N \sum d_x^2 - (\sum d_x)^2} \sqrt{N \sum d_y^2 - (\sum d_y)^2}}
\]

Where

\[d_x = x - A, \quad d_y = y - B\] are the deviations from arbitrary values A and B.
Note 2: Correlation is independent of change of scale & origin

\[ u = \frac{X - a}{h}, \quad v = \frac{Y - b}{k} \]

where \(a\) and \(b\) are arbitrary constants

\[ \bar{u} = \frac{\sum u}{n}, \quad \bar{v} = \frac{\sum v}{n} \]

\[ Cov(X, Y) = Cov(u, v) = \frac{1}{n} \sum uv - \bar{u} \bar{v} \]

\[ \sigma_u = \sqrt{\frac{1}{n} \sum u^2 - \bar{u}^2}, \quad \sigma_v = \sqrt{\frac{1}{n} \sum v^2 - \bar{v}^2} \]

\[ r(X, Y) = r(u, v) = \frac{Cov(u, v)}{\sigma_u \sigma_v} \]

**Limits for Correlation co-efficient**

Correlation co-efficient lies between -1 and +1. i.e. \(-1 < r(x, y) < 1\).

i) If \(r(x, y) = +1\) the variables \(x\) and \(y\) are said to be perfectly positively correlated.

ii) If \(r(x, y) = -1\) the variables \(x\) and \(y\) are said to be perfectly negatively correlated.

iii) If \(r(x, y) = 0\) the variables \(x\) and \(y\) are said to be uncorrelated.
Problem 1
Calculate the correlation co-efficient for the following heights (in inches) of fathers(X) and their sons(Y).

\[ X: 65 \quad 66 \quad 67 \quad 67 \quad 68 \quad 69 \quad 70 \quad 72 \]
\[ Y: 67 \quad 68 \quad 65 \quad 68 \quad 72 \quad 72 \quad 69 \quad 71 \]

**Solution:**

\[
\bar{X} = \frac{\sum X}{n} = \frac{544}{8} = 68
\]
\[
\bar{Y} = \frac{\sum Y}{n} = \frac{552}{8} = 69
\]

<table>
<thead>
<tr>
<th>X</th>
<th>Y</th>
<th>x = X - \bar{X}</th>
<th>y = Y - \bar{Y}</th>
<th>x^2</th>
<th>y^2</th>
<th>xy</th>
</tr>
</thead>
<tbody>
<tr>
<td>65</td>
<td>67</td>
<td>-3</td>
<td>-2</td>
<td>9</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>66</td>
<td>68</td>
<td>-2</td>
<td>-1</td>
<td>4</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>67</td>
<td>65</td>
<td>-1</td>
<td>-4</td>
<td>1</td>
<td>16</td>
<td>4</td>
</tr>
<tr>
<td>67</td>
<td>68</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>68</td>
<td>72</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>9</td>
<td>0</td>
</tr>
<tr>
<td>69</td>
<td>72</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>9</td>
<td>3</td>
</tr>
<tr>
<td>70</td>
<td>69</td>
<td>2</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>72</td>
<td>71</td>
<td>4</td>
<td>2</td>
<td>16</td>
<td>4</td>
<td>8</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>\sum x</th>
<th>\sum y</th>
<th>\sum x^2</th>
<th>\sum y^2</th>
<th>\sum xy</th>
</tr>
</thead>
<tbody>
<tr>
<td>544</td>
<td>552</td>
<td>0</td>
<td>0</td>
<td>36</td>
</tr>
</tbody>
</table>
\[
    r(X, Y) = \frac{\sum xy}{\sqrt{\sum x^2 \sum y^2}} = \frac{24}{\sqrt{36 \cdot 54}} = 0.603
\]

Since \( r(x, y) = 0.603 \), the variables X and Y are positively correlated. i.e. heights of fathers and their respective sons are said to be positively correlated.

2. Calculate the correlation co-efficient from the following data

\[ \sum X = 125, \sum Y = 100 \]
\[ \sum X^2 = 650, \sum Y^2 = 436, \sum XY = 520 \]

Solution: We know,
\[
    r(X, Y) = \frac{N \sum XY - \sum X \sum Y}{\sqrt{N \sum X^2 - (\sum X)^2} \sqrt{N \sum Y^2 - (\sum Y)^2}}
\]

\[
    = \frac{25(520) - (125)(100)}{\sqrt{25(650) - (125)^2} \sqrt{25(436) - (100)^2}} = -0.667
\]
Spearman’s formula for the rank correlation

When the numerical values of X & Y are not available, but the position of the values are arranged in order of merit (ranks), then also we can find the correlation coefficient between the characteristics of group of individuals using the formula

$$r_s = 1 - \frac{6 \sum_{i=1}^{n} d_i^2}{n(n^2 - 1)}$$

$d_i = x_i - y_i$, $x_i$'s, $y_i$'s

Indicates ranks
3. The table lists 7 schools and provides data about the percentage of pupils who have free school meals and their GCSE results. Calculate Spearman’s rank correlation coefficient and explain what this means in the context of this question.

<table>
<thead>
<tr>
<th>School</th>
<th>% of pupils claiming free school meals, x</th>
<th>% of pupils gaining 5 or more GCSEs at grades A*-C, y</th>
</tr>
</thead>
<tbody>
<tr>
<td>Appledore</td>
<td>14.4</td>
<td>54</td>
</tr>
<tr>
<td>Butterscotch</td>
<td>7.2</td>
<td>64</td>
</tr>
<tr>
<td>Copperdale</td>
<td>27.5</td>
<td>44</td>
</tr>
<tr>
<td>Damson View</td>
<td>33.8</td>
<td>32</td>
</tr>
<tr>
<td>Eagle High</td>
<td>38.0</td>
<td>37</td>
</tr>
<tr>
<td>Forrest Green</td>
<td>15.9</td>
<td>68</td>
</tr>
<tr>
<td>Greengage</td>
<td>4.9</td>
<td>62</td>
</tr>
</tbody>
</table>
**Solution:** First rank the data:

<table>
<thead>
<tr>
<th>School</th>
<th>Rank x</th>
<th>Rank y</th>
<th>d</th>
<th>d^2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Appledore</td>
<td>3</td>
<td>4</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>Butterscotch</td>
<td>2</td>
<td>6</td>
<td>-4</td>
<td>16</td>
</tr>
<tr>
<td>Copperdale</td>
<td>5</td>
<td>3</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Damson View</td>
<td>6</td>
<td>1</td>
<td>5</td>
<td>25</td>
</tr>
<tr>
<td>Eagle High</td>
<td>7</td>
<td>2</td>
<td>5</td>
<td>25</td>
</tr>
<tr>
<td>Forrest Green</td>
<td>4</td>
<td>7</td>
<td>-3</td>
<td>9</td>
</tr>
<tr>
<td>Greengage</td>
<td>1</td>
<td>5</td>
<td>-4</td>
<td>16</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td></td>
<td></td>
<td><strong>96</strong></td>
</tr>
</tbody>
</table>
Using the formula

\[ r_s = 1 - \frac{6 \sum d^2}{n(n^2 - 1)} = 1 - \frac{6 \times 96}{7(7^2 - 1)} = 1 - \frac{576}{336} = -0.714 \]

This value represents negative rank correlation, so the schools with the highest proportion of pupils receiving free school meals tend to have the least successful GCSE results (and vice versa).

**Note:** When an item is repeated \( m \) times in a series, then common rank is the average of the ranks which these items would have assumed if they are slightly different from each other and the next item will get the rank next to the ranks already assumed. As a result of this, correction factor

\[ = \frac{m(m^2 - 1)}{12} \]

to be added to \( \sum d^2 \)

where \( m \) is the no. of item is repeated. This c.f should be added for each repeated value.
4. Obtain the rank correlation coeff. For the following data

\[ X: 68 \ 64 \ 65 \ 50 \ 64 \ 80 \ 75 \ 40 \ 55 \ 64 \]
\[ Y: 62 \ 58 \ 68 \ 45 \ 81 \ 60 \ 68 \ 48 \ 50 \ 70 \]

**Solution:**

<table>
<thead>
<tr>
<th>X</th>
<th>Y</th>
<th>rank in $x_i$</th>
<th>rank in $y_i$</th>
<th>$d = x_i - y_i$</th>
<th>$d^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>68</td>
<td>62</td>
<td>4</td>
<td>5</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>64</td>
<td>58</td>
<td>6</td>
<td>7</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>75</td>
<td>68</td>
<td>2.5</td>
<td>3.5</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>50</td>
<td>45</td>
<td>9</td>
<td>10</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>64</td>
<td>81</td>
<td>6</td>
<td>1</td>
<td>5</td>
<td>25</td>
</tr>
<tr>
<td>80</td>
<td>60</td>
<td>1</td>
<td>6</td>
<td>-5</td>
<td>25</td>
</tr>
<tr>
<td>75</td>
<td>68</td>
<td>2.5</td>
<td>3.5</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>40</td>
<td>48</td>
<td>10</td>
<td>9</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>55</td>
<td>50</td>
<td>8</td>
<td>8</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>64</td>
<td>70</td>
<td>6</td>
<td>2</td>
<td>4</td>
<td>16</td>
</tr>
</tbody>
</table>

\[ 72 \]
In X series 75 is repeated twice which are in the position 2 & 3 in ranks.

∴ common rank is 2.5 (which is the average of 2 & 3) is to be given for each 75. Also in X series 64 is repeated 3 times which are in the position 5, 6 & 7 in ranks.

∴ common rank is \( \frac{5 + 6 + 7}{3} = 6 \), to be given for each 64.

Similarly in Y series 68 is repeated 2 times which are in the position 3 & 4 in ranks.

∴ common rank is 3.5 (which is the average of 3 & 4) is to be given for each 68.
Correction factor

In X series 75 is repeated twice :\[ c.f \quad \text{is} \quad \frac{2(2^2 - 1)}{12} = \frac{1}{2} \]

64 is repeated thrice :\[ c.f \quad \text{is} \quad \frac{3(3^2 - 1)}{12} = \frac{24}{12} = 2 \]

In Y series 68 is repeated twice :\[ c.f \quad \text{is} \quad \frac{2(2^2 - 1)}{12} = \frac{1}{2} \]

\[ r = 1 - \frac{6(\sum d^2 + \frac{1}{2} + 2 + \frac{1}{2})}{10(10^2 - 1)} = 1 - \frac{6(72 + 5)}{10 \times 99} = 1 - \frac{450}{990} = 0.5454 \]
Regression Lines

For the pair of values of \((X, Y)\), where \(X\) is an independent variable and \(Y\) is the dependent variable, the line of regression of \(Y\) on \(X\) is given by

\[ Y - \bar{Y} = b_{yx} (X - \bar{X}) \]

where \(b_{yx}\) is the regression co-efficient of \(Y\) on \(X\) and given by

\[ b_{yx} = r \frac{\sigma_y}{\sigma_x} \]

where \(r\) is the correlation co-efficient between \(X\) and \(Y\) and \(\sigma_x\) and \(\sigma_y\) are the standard deviations of \(X\) and \(Y\) respectively

\[ \therefore b_{yx} = \frac{\sum xy}{\sum x^2} \quad \text{Where} \quad x = X - \bar{X}, \quad y = Y - \bar{Y} \]
Similarly when $Y$ is treated as an independent variable and $X$ as dependent variable, the line of regression of $X$ on $Y$ is given by

$$X - \bar{X} = b_{xy} (Y - \bar{Y})$$

where $b_{xy}$ is the regression co-efficient of $X$ on $Y$ and given by

$$b_{xy} = r \frac{\sigma_x}{\sigma_y} = \frac{\sum xy}{\sum y^2}$$

Where $x = X - \bar{X}$, $y = Y - \bar{Y}$
Problems

5. Out of 2 reg lines given by \( x + 2y - 5 = 0 \) & \( 2x + 3y - 8 = 0 \). Which one is the reg. of \( x \) on \( y \)?

Solution:

Suppose \( x + 2y - 5 = 0 \) is the equn. of the reg. line of \( x \) on \( y \) & \( 2x + 3y - 8 = 0 \) is the equn. of the reg. line of \( y \) on \( x \), then the 2 equns can be written as \( x = -2y + 5 \) & \( y = -\frac{2}{3}x + \frac{8}{3} \)

Hence \( b_{yx} = -\frac{2}{3} \) & \( b_{xy} = -2 \)

Now \( r^2 = \frac{4}{3} > 1 \)
This is impossible. Hence our assumption is wrong

\[ 2x + 3y - 8 = 0 \text{ is the equn. Of reg. line of } x \text{ on } y \]

6. If \( \sigma_x = \sigma_y = \sigma \text{ and the angle bet. 2 reg. lines} \)

is \( \tan^{-1} 3 \). Obtain the values the correlation coefficient.

Solution: The acute angle bet. 2 reg. lines is

\[
\theta = \tan^{-1} \left\{ \frac{1 - r^2}{r} \left( \frac{\sigma_x \sigma_y}{\sigma_x^2 + \sigma_y^2} \right) \right\} = \tan^{-1} \left\{ \frac{1 - r^2}{r} \left( \frac{\sigma^2}{2 \sigma^2} \right) \right\}
\]

\[
= \tan^{-1} \left\{ \frac{1 - r^2}{2r} \right\} = \tan^{-1} 3 \text{ (given)}
\]

\[
\therefore \frac{1 - r^2}{2r} = 3 \Rightarrow r^2 + 6r - 1 = 0
\]
Let the marks in Economics be denoted by X and statistics by Y.

7. Marks obtained by 10 students in Economics and Statistics are given below.

Marks In Eco.: 25 28 35 32 31 36 29 38 34 32
Marks In Stat: 43 46 49 41 36 32 31 30 33 39

Find (i) the regression equation of Y on X

(ii) estimate the marks in statistics when the marks in Economics is 30.

Solution:
Let the marks in Economics be denoted by X and statistics by Y.

\[ r = \frac{-6 \pm \sqrt{36 + 4}}{2} = \frac{-6 \pm \sqrt{40}}{2} = -3 \pm \sqrt{10} \]

\[ = -3 \pm 3.16 = 0.16 \text{ or } -6.16 = 0.16 \]

(-6.16 is not possible ) \[ \therefore -1 \leq r \leq 1 \]
\[
\bar{X} = \frac{\sum X}{n} = \frac{320}{10} = 32 \\
\bar{Y} = \frac{\sum Y}{n} = \frac{380}{10}
\]

\[
b_{yx} = \frac{\sum xy}{\sum x^2} = \frac{-93}{140} = -0.664
\]
(i) Regression equation of $Y$ on $X$ is

$$ Y - \bar{Y} = b_{yx} ( X - \bar{X} ) $$

$$ Y - 38 = -0.664(X - 32) $$

$$ Y = 59.25 - 0.664X $$

(ii) To estimate the marks in statistics ($Y$) for a given marks in the Economics ($X$), put $X = 30$, in the above equation we get,

$$ Y = 59.25 - 0.664(30) $$

$$ = 59.25 - 19.92 = 39.33 \ or \ 39 $$
8. If \( y = 2x - 3 \) & \( y = 5x + 7 \) are the two regression lines, find the mean values of \( x \) and \( y \). Find the correlation coefficient between \( x \) & \( y \). Find an estimate of \( x \) when \( y = 1 \).

**Solution:**

**Given** \( y = 2x - 3 \rightarrow (1) \) \( y = 5x + 7 \rightarrow (2) \)

Since both the lines of regression passes through the mean value \( \bar{x} \) and \( \bar{y} \) the point, \((\bar{x}, \bar{y})\) must satisfy the two given regression lines.

\[ \bar{y} = 2\bar{x} - 3 \rightarrow (3) & \bar{y} = 5\bar{x} + 7 \rightarrow (4) \]

Subtracting the equations (3) and (4), we have

\[ 3\bar{x} = -10 \quad \Rightarrow \quad \bar{x} = \frac{-10}{3} \]

\[ \bar{y} = 2\left(\frac{-10}{3}\right) - 3 = \frac{-29}{3} \]

Therefore mean values are \( \bar{x} = \frac{-10}{3} \) and \( \bar{y} = \frac{-29}{3} \).
Let us suppose that equation (1) is the line of regression of $y$ on $x$ and equation (2) is the equation of the line of regression of $x$ on $y$, we have

(1) $\Rightarrow y = 2x - 3$

(2) $\Rightarrow 5x = y - 7$

$b_{yx} = 2$

\[ i.e \quad x = \frac{1}{5}y - \frac{7}{5} \quad \therefore b_{xy} = \frac{1}{5} \]

\[ r = \sqrt{b_{xy} \times b_{yx}} = \sqrt{\frac{1}{5} \times 2} = \pm 0.63 \]

Since both the regression co-efficients are positive, $r$ must be positive.

Correlation co-efficient $r = 0.63$. Substituting $y=1$

in (2), we have $5x = 1 - 7 = -6$. $x = -\frac{6}{5}$. 
Properties of regression coefficients

i) The geometric mean between the regression coefficient is the correlation coefficient i.e
\[ r = \pm \sqrt{b_{xy} \times b_{yx}} \]

Note: sign of \( b_{xy}, b_{yx} \) & \( r \) are same always
Both the lines of reg. intersect at \((\bar{x}, \bar{y})\)

ii) If one of the reg. coe. is greater than unity, the other must be less than unity

iii) Arithmetic mean of the regression coefficients is greater than or equal to the correlation coefficient

iv) Regression coef. Are independent of change of origin but dependent on change of scale.
v) Angle between 2 regression lines

\[
\theta = \tan^{-1} \left\{ \frac{r^2 - 1}{r} \left( \frac{\sigma_x \sigma_y}{\sigma_x^2 + \sigma_y^2} \right) \right\} \text{ (obtuse)}
\]

\[
\theta = \tan^{-1} \left\{ \frac{1 - r^2}{r} \left( \frac{\sigma_x \sigma_y}{\sigma_x^2 + \sigma_y^2} \right) \right\} \text{ (acute)}
\]

Note: 1. If \( r = 0 \) then 2 variables are uncorrelated and lines of regression are perpendicular to each other.

2. If \( r = \pm 1 \) then the 2 lines are parallel. But the lines intersect at \((\bar{x}, \bar{y})\) implies the lines must coincide.
Sample space and events

**Random experiment**: An experiment in which all possible outcomes are known in advance and the exact outcome in any specific trial of the experiment is unpredictable is known as random experiment.

**Trial**: Any particular performance of a random experiment is called a trial and outcomes are known as events.

**Sample space of an experiment**: The set of all elementary results (outcomes) in a random experiment is called sample space. It is denoted by S. The elements of a sample space are called sample points.
Example 1 The experiment consists of tossing two coins, all the possible outcomes of this experiment form the following set:

\[ S = \{(H, H), (H, T), (T, H), (T, T)\} \]

**Discrete and continuous sample space**

A sample space is called **discrete** if it contains only finite or countably infinite points which can be arranged in a simple sequence.

Ex1. Two tosses of a coin \[ S = \{(H, H), (H, T), (T, H), (T, T)\} \]

Ex2. Single toss of a die \[ S = \{1, 2, 3, \ldots, 6 \} \]

**Continuous sample space**: If the elements of a sample space contains uncountable no of sample points.

Ex1. All points on a line.
The subset of the set $S$ is called event. The empty set $\emptyset$ of $S$ is called impossible event, and $S$ is called certain event or sure event.

**Favourable events:** The no. of cases Favourable to an event in a trial is the no. of outcomes which entail the happening of the event.

Ex. In drawing a card from a pack of cards, the no. of cases favourable to the event of getting an ace is 4, for drawing a spade is 13.

**Mutually exclusive & exhaustive system of events:**

The events $E_1, E_2, \ldots, E_n$ are said to form a mutually exclusive & exhaustive system of events if $E_i \cap E_j = \emptyset$ for $i \neq j$ and $E_1 \cup E_2 \ldots \cup E_n = S$. 
**Independent events:** Two events are said to be independent if the occurrence of one does not depend upon the occurrence of the other.

Ex. In tossing a coin, the event of getting a head in the first toss is independent of getting a head in the second, third & subsequent throw.

**Complementary events:**

Let $E$ be an event of a random experiment and $S$ be its sample space. The set containing all the other outcomes which are not in $E$ but in the sample space is called the complimentary event of $E$.

It is denoted by $\overline{E}$. Thus, $\overline{E} = S - E$. Note that $E$ and $\overline{E}$ are mutually exclusive events.
In throwing a die, let $E = \{2, 4, 6\}$ be an event of getting a multiple of 2. Then the complement ary of the event $E$ is given by $E = \{1, 3, 5\}$.

Ex. When a die is thrown, occurrence of an even no. $(2, 4, 6)$ and odd no. $(1, 3, 5)$ are complementary events.

**Probability (Mathematical or classical)**

If random experiment or a trial results in ‘n’ exhaustive, mutually exclusive and equally likely outcomes (or cases), out of which m are favourable to the occurrence of an event E, then the probability ‘p’ of occurrence (or happening) of E, usually denoted by $P(E)$, is given by
\[ p = P(E) = \frac{\text{No. of favourable cases}}{\text{Total no. of exhaustive cases}} = \frac{m}{n} \]

Note:

(i) The above classical definition of probability is not applicable if the number of possible outcomes is infinite and the outcomes are not equally likely.

(ii) The probability of an event \( A \) lies between 0 and 1, both inclusive; That is \( 0 \leq P(A) \leq 1 \).

(iii) The probability of the sure event is 1. That is \( P(S) = 1 \)
(iv) The probability of an impossible event is 0. That is $P(\emptyset) = 0$

(v) The probability that the event $A$ will not occur is given by $P(\text{not } A) = P(\overline{A})$ or $P(A') = 1 - P(A)$

(vi) $P(A) + P(\overline{A}) = 1$
Problems

9. How many elements will be there in \( S \)

i) When a coin is tossed 3 times? (Ans : \( 2^3 = 8 \))

ii) When a die is rolled 2 times? (Ans : \( 6^2 = 36 \))

10. Two dice are thrown together (or single thro' of 2 dice). Find the following probabilities.

Solution: When two dice are thrown, the sample space is

\[
S = \{(1,1), (1,2), (1,3), (1,4), (1,5), (1,6), \\
(2,1), (2,2), (2,3), (2,4), (2,5), (2,6), \\
(3,1), (3,2), (3,3), (3,4), (3,5), (3,6), \\
(4,1), (4,2), (4,3), (4,4), (4,5), (4,6), \\
(5,1), (5,2), (5,3), (5,4), (5,5), (5,6), \\
(6,1), (6,2), (6,3), (6,4), (6,5), (6,6)\}
\]
i) Probability of getting not the same no. on the 2 dice.

Let E be the event of getting the same no. on the 2 dice.

\[ n(E) = 30 \quad \& \quad P(E) = \frac{30}{36} \]

ii) Probability of a total of 8.

Let E be the event of getting a total of 8.

\[ n(E) = 5 \quad \& \quad P(E) = \frac{5}{36} \]

iii) Probability of a total of 9 or 11.

\[ E = \{(4,5), (5,4), (3,6), (6,3), (5,6), (6,5)\} \]

\[ n(E) = 6 \quad \& \quad P(E) = \frac{6}{36} \]
iv) Probability of both odd digits

\[ E = \{(1,1), (1,3), (1,5), (3,1), (3,3), (3,5), (5,1), (5,3), (5,5)\} \]

\[ \therefore n(E) = 9 \text{ and } P(E) = \frac{9}{36} \]

v) An even number as sum

\[ E = \begin{cases} 
(1,1), (1,3), (1,5), (2,2), (2,4), (2,6), (3,1), (3,3), (3,5), \\
(4,2), (4,4), (4,6), (5,1), (5,3), (5,5), (6,2), (6,4), (6,6) 
\end{cases} \]

\[ \therefore n(E) = 18 \text{ and } P(E) = \frac{18}{36} \]

vi) Multiples of 3 as sum
\[ E = \{(1,2), (2,3), (1,5), (5,1), (2,4), (4,2), (4,5), (5,4), (3,3), (3,6), (6,3), (6,6)\} \]

\[
\therefore n(E) = 12 \quad \text{and} \quad P(E) = \frac{12}{36} = \frac{1}{3}
\]

vii) a doublet

\[
\therefore n(E) = 6 \quad \text{and} \quad P(E) = \frac{1}{6}
\]

11. What is the probability that an ordinary year has 53 Sundays?

**Solution:** An ordinary year has 365 days. (i.e) 52 weeks + 1 day. This day can be any one of 7 days. \( P(\text{Sunday}) = \frac{1}{7} \)
12. A person is known to hit the target in 3 out of 4 shots. whereas another person is known to hit the target in 2 out of 3 shots. Find the probability of the target being hit at all when the both person try.

**Solution:** The probability that first person hit the target is

\[ P(A) = \frac{3}{4}, \quad P(\overline{A}) = \frac{1}{4} \]

The probability that second person hit the target is

\[ P(B) = \frac{2}{3}, \quad P(\overline{B}) = \frac{1}{3}. \]

The two events are independent.

\[ P(\text{Not hitting the target}) = P(\overline{A} \cap \overline{B}) = P(\overline{A})P(\overline{B}) = \frac{1}{4} \times \frac{1}{3} = \frac{1}{12} \]
Problems on independent events:

1. A&B appear in a interview for 2 vacancies in the same post. The prob, of A’s selection is $\frac{1}{6}$ & that of B’s selection is $\frac{1}{4}$. Find the prob. that i) both of them are selected  ii) only one of them is selected  iii) none of them is selected iv) at least one of them is selected

Solution: Given $P(A) = \frac{1}{6}, P(B) = \frac{1}{4}$,

$$\therefore P(\bar{A}) = 1 - \frac{1}{6} = \frac{5}{6}, P(\bar{B}) = 1 - \frac{1}{4} = \frac{3}{4}$$

$P(both\ selected) = P(A \cap B) = P(A)P(B) = \frac{1}{24}$

$P(only\ one\ is\ selected) = P(A \cap \bar{B}) \ or \ P(\bar{A} \cap B) = P(A \cap \bar{B}) \cup P(\bar{A} \cap B)$
\[
= P(A)P(\overline{B}) + P(A)P(\overline{B}) = \frac{1}{6} \times \frac{3}{4} + \frac{1}{4} \times \frac{5}{6} = \frac{3}{8} + \frac{5}{24} = \frac{8}{24}
\]

\[
iii) P(\text{none is selected}) = P(\text{not } A \& \text{ not } B) = P(\overline{A} \cap \overline{B}) = \frac{5}{6} \times \frac{3}{4} = \frac{5}{8}
\]

\[
iv) \text{at least one of them is selected}
\]

\[
P(\text{at least one of them is selected}) = 1 - P(\text{none is selected}) = 1 - \frac{5}{8} = \frac{3}{8}
\]

Problems with replacement (Independent event)

From a bag containing 4 white & 6 black balls, two balls are drawn at random, if the balls are drawn one after the other with replacement, find the probability that:

i) both are white  
ii) none is white  
iii) at least one is white  
iv) the 1st ball is white and the 2nd ball is black
Solution: Total no. of balls=10

i) both are white

\[ P(\text{1}\text{st ball is white}) = \frac{4C_1}{10C_1} = \frac{4}{10} \quad \text{and} \quad P(\text{2}\text{nd ball is white}) = \frac{4C_1}{10C_1} = \frac{4}{10} \]

\[ \therefore P(\text{both are white}) = P(A \cap B) = P(A)P(B)(\because A \& B \text{ Independent}) \]

\[ = \frac{4}{10} \times \frac{4}{10} = \frac{4}{25} \]

ii) none is white

Out of 10 balls, none white balls=6

\[ P(\text{no white in 1}\text{st draw}) = \frac{6C_1}{10C_1} = \frac{6}{10} \]

\[ P(\text{no white in 2}\text{nd draw}) = \frac{6C_1}{10C_1} = \frac{6}{10} \]

Required probability = \[ P(\text{none is white}) = \frac{6}{10} \times \frac{6}{10} = \frac{9}{25} \]
iii) at least one is white

\[ P( \text{at least one is white}) = 1 - P( \text{none is white}) = 1 - \frac{9}{5} = \frac{16}{25} \]

iv) the 1\textsuperscript{st} ball is white and the 2\textsuperscript{nd} ball is black

\[ P( \text{1}\textsuperscript{st} \text{ball is white}) = \frac{\binom{4}{1}}{\binom{10}{1}} = \frac{4}{10} \]

\[ P( \text{2}\textsuperscript{nd} \text{ball is black}) = \frac{6}{10} \]

\[ \therefore P( \text{1}\textsuperscript{st} \text{ball is white} \& \text{2}\textsuperscript{nd} \text{ball is black} ) = \frac{4}{10} \times \frac{6}{10} = \frac{6}{25} \]
**Dependent events:**

Two events A&B are said to be dependent events when B can occur only when A known to have occurred (or vice versa). Probability attached to such events are called conditional probability.

**Conditional Probability :**

The prob. of an event A given that B has already occurred is called the conditional prob. of A given B is

\[ P(A \mid B) = \frac{P(AB)}{P(B)}, P(B) \neq 0 \]

Also conditional prob. of

B given A is \[ P(B \mid A) = \frac{P(AB)}{P(A)}, P(A) \neq 0 \]
Without replacement (dependent event)
1. A bag contains 25 tickets, a ticket is drawn and without replacement another ticket is drawn. Find the prob. that both tickets will show even no.s.

Solution: \[ S = \{1, 2, 3, 4, 5, ..., 25\}, \quad n(s) = 25 \]

\[ E = \{2, 4, 6, 8, 10, ..., 24\}, \quad n(E) = 12 \]

\[ P(\text{1st even number}) = \frac{12}{25} \]

This ticket is not replaced. \(\therefore\) Total is 24. Out of which 11 are even number.

\[ P(\text{2nd even number}) = \frac{11}{24} \]

\[ \therefore P(\text{both even number}) = P(1\text{st even number} \& 2\text{nd even number}) \]

\[ = \frac{12}{25} \times \frac{11}{24} = \frac{11}{50} \]
AXIOMS OF PROBABILITY

Let S be a sample space. A probability is an assignment of a non-negative number $P(A)$ to each event $A \subseteq S$ that satisfies the following 3 axioms

1. $P(\emptyset) = 0$ & $P(S) = 1$  
2. $0 \leq P(A) \leq 1$.

3. If $P(A \cap B) = \emptyset$ then $P(A \cup B) = P(A) + P(B)$
Laws of probability:

Addition theorem on probability

Let A and B be subsets of a finite non-empty set S

Then \( n(A \cup B) = n(A) + n(B) - n(A \cap B) \)

Divide both sides by \( n(S) \), we get

\[
\frac{n(A \cup B)}{n(S)} = \frac{n(A)}{n(S)} + \frac{n(B)}{n(S)} - \frac{n(A \cap B)}{n(S)} \rightarrow (1)
\]

If the subsets A and B correspond to two events A and B of a random experiment and if the set S corresponds to the sample space S of the experiment, then (1) becomes
\[ P(A \cup B) = P(A) + P(B) - P(A \cap B) \]

This result is known as the addition theorem on probability

**Multiplication Theorem on Probability**

Let \( A \& B \) be two events associated with a sample space \( S \).

Clearly, the set \( A \cap B \) denotes the event that both \( A \& B \)

Or the simultaneous occurrence of the events \( A \& B \)

The event \( A \cap B \) is also written as \( AB \). The probability of event \( AB \) is obtained by using the conditional probability as obtained below:
We know that the conditional probability of event $A$ given that $B$ has occurred is

$$P(A/B) = \frac{P(AB)}{P(B)}, P(B) \neq 0$$

From this result, we can write

$$P(AB) = P(B)P(A/B) \rightarrow (1)$$

Also, we know that

$$P(B/A) = \frac{P(BA)}{P(A)}, P(A) \neq 0$$

$$P(B/A) = \frac{P(AB)}{P(A)} (\therefore P(AB) = P(BA))$$

Thus $P(AB) = P(A)P(B/A) \rightarrow (2)$

Combining (1) and (2), we find that

$$P(AB) = P(A)P(B/A) = P(B)P(A/B), \text{ provided}$$

$$P(A) \neq 0 \& P(B) \neq 0$$

The above result is known as the multiplication rule of probability
**Multiplication rule of probability for more than two events**

If $A, B & C$ are three events of sample space, we have

$$P(ABC) = P(A)P(A/B)P(C/AB)$$

**Definition:**

Two events $A$ and $B$ are said to be independent if

if $P(B/A) = P(B)$, $P(A) \neq 0$ and

$$P(A/B) = P(A), P(B) \neq 0$$

Now, by the multiplication rule of probability, we have

$$P(AB) = P(B)P(A/B) \rightarrow (1)$$

If $A$ and $B$ are independent, then (1) becomes

$$P(AB) = P(A)P(B)$$
If $A$ and $B$ are independent, then (1) becomes

$$P(AB) = P(A)P(B)$$

**Prove that if $E$ and $F$ are independent events, then so are the events $i)\overline{E} \& F$ $ii)\ E \& \overline{F}$ $iii)\ \overline{E} \& \overline{F}$**

**Solution:**

ii). Since $E$ and $F$ are independent, we have

$$P(EF) = P(E)P(F) \rightarrow (1)$$

From the venn, it is clear that $E \cap F$ and $\overline{E} \cap F$ are mutually exclusive events

and also $E = (E \cap F) \cup (E \cap \overline{F})$. Therefore

$$P(E) = P((E \cap F)) + P(E \cap \overline{F}) \text{ or } P(E \cap \overline{F}) = P(E) - P((E \cap F))$$
\[ P(E \cap \bar{F}) = P(E) - P(E)P(F) \text{ by (1)} \]
\[ = P(E)(1 - P(F)) \]
\[ = P(E)P(\bar{F}) \]

Hence, \( E \& \bar{F} \) are independent

Similarly we can prove i) using
\[ F = (E \cap F) \cup (\bar{E} \cap F) \]

iii) w.k.t \( (E \cup F) = \bar{E} \cap \bar{F} \)

\[ P(\bar{E} \cap \bar{F}) = P((\bar{E} \cup F)) \]
\[ = 1 - P(E \cup F) \]
\[ = 1 - [P(E) + P(F) - P(E \cap F)] \]
\[ = 1 - P(E) - P(F) + P(E \cap F) \]

\[ = 1 - P(E) - P(F) + P(E P(F)) \]
\[ = 1 - P(E) - P(F)(1 - P(E)) \]
\[ = P(\bar{E})P(\bar{F}) \]

Hence, \( \bar{E} \& \bar{F} \) are independent
Partition of a sample space

A set of events $E_1, E_2, E_3 \ldots E_n$ is said to represent a partition of the sample space $S$ if

i) $E_i \cap E_j = \emptyset$, $i \neq j$, $i, j = 1, 2, 3 \ldots n$

ii) $S = E_1 \cup E_2 \cup E_3 \ldots \cup E_n$

iii) $P(E_i) > 0$ for all $i = 1, 2, 3 \ldots n$

In other words, the events $E_1, E_2, E_3 \ldots E_n$ represent a partition of the samplespace $S$ if they are pairwisedisjoint, exhaustive and have non zero probabilities.
Theorem of total probability

Let \( \{ E_1, E_2, E_3, \ldots, E_n \} \) be a partition of the sample space \( S \), and suppose that each of the events \( E_1, E_2, E_3, \ldots, E_n \) has nonzero probability of occurrence. Let \( A \) be any event associated with \( S \), then

\[
P(A) = P(E_1)P(A/E_1) + P(E_2)P(A/E_2) + \ldots + P(E_n)P(A/E_n)
\]

\[
= \sum_{j=1}^{n} P(E_j)P(A/E_j)
\]

Proof:

Given that \( E_1, E_2, E_3, \ldots, E_n \) is a partition of the sample space \( S \) (Fig). Therefore,
\[ S = E_1 \cup E_2 \cup E_3 \ldots \cup E_n \] and \[ E_i \cap E_j = \emptyset, \ i \neq j, \ i, j = 1, 2, 3 \ldots n \]

For any event \( A \), \[ A = A \cap S \]

\[ = A \cap E_1 \cup E_2 \cup E_3 \ldots \cup E_n \]

\[ = (A \cap E_1) \cup (A \cap E_2) \cup (A \cap E_3) \cup \ldots \cup (A \cap E_n) \]

Also \( A \cap E_i \) and \( A \cap E_j \) are respectively by the subsets of \( E_i \) and \( E_j \). We know that

\( E_i \) and \( E_j \) are disjoint, for \( i \neq j \), therefore

\( A \cap E_i \) and \( A \cap E_j \) are also disjoint

for all \( i \neq j, \ i, j = 1, 2, 3 \ldots n \). Thus,

\[ P(A) = P [(A \cap E_1) \cup (A \cap E_2) \cup (A \cap E_3) \cup \ldots \cup (A \cap E_n)] \]

\[ = P(A \cap E_1) + P(A \cap E_2) + P(A \cap E_3) + \ldots + P(A \cap E_n) \]
By multiplication rule of prob. we have

\[ P(A \cap E_i) = P(E_i)P(A / E_i) \] as \( P(E_i) \neq 0 \) \( \forall i = 1,2,3,...n \)

\[ P(A) = P(E_1)P(A / E_1) + P(E_2)P(A / E_2) + ... + P(E_n)P(A / E_n) \]

\[ = \sum_{j=1}^{n} P(E_j)P(A / E_j) \]

2. A person has undertaken a construction job. The probabilities are 0.65 that there will be strike, 0.80 that the construction job will be completed on time if there is no strike, and 0.32 that the construction job will be completed on time if there is a strike. Determine the probability that the construction job will be completed on time.
Solution.

Let A be the event that the construction job will be completed on time, and B be the event that there will be a strike. We have to find \( P(A) \).

We have \( P(B) = 0.65 \), \( P(\text{no strike}) = P(\overline{B}) = 1 - P(B) \)

\[ P(A/B) = 0.32, \quad P(A/\overline{B}) = 0.80 \]

Since events \( B \) and \( \overline{B} \) form a partition of the sample space \( S \), therefore, by theorem on total probability, we have

\[
P(A) = P(B) P(A/B) + P(\overline{B}) P(A/\overline{B})
\]

\[
= 0.65 \times 0.32 + 0.35 \times 0.8
\]

\[
= 0.208 + 0.28 = 0.488
\]

Thus, the probability that the construction job will be completed in time is 0.488.
3. An Urn contains 5W & 3G balls and another urn contains 3W & 7G. 2 balls are chosen at random from the first urn and put it into the second urn. Then a ball is drawn from the second urn. What is the prob. That it is a white.

**Solution.** The 2 balls drawn from the first urn can be

i) both white (event $E_1$)  
ii) both green (event $E_2$)  
iii) one white & one green (event $E_3$)

$$P(E_1) = \frac{\binom{5}{2}}{\binom{8}{2}} = \frac{5}{14}$$

$$P(E_2) = \frac{\binom{3}{2}}{\binom{8}{2}} = \frac{3}{28}$$

$$P(E_3) = \frac{\binom{5}{1} \times \binom{3}{1}}{\binom{8}{2}} = \frac{15}{28}$$
After the balls are transferred from first urn to the second urn, the second urn will contain

i) 5W & 7G  
ii) 3W & 9G  
iii) 4W & 8G

Let A be the event of drawing a W ball from the second urn, Then

\[
P(A/E_1) = \frac{5C_1}{12C_1} = \frac{5}{12}, \quad P(A/E_2) = \frac{3C_1}{12C_1} = \frac{3}{12}, \quad P(A/E_3) = \frac{4C_1}{12C_1} = \frac{4}{12}
\]

Therefore the required probability

\[
P(A) = \sum_{j=1}^{3} P(E_j) P(A/E_j) = 0.354
\]
Bayes’ Theorem

If \( E_1, E_2, E_3, \ldots, E_n \) are \( n \) non empty events which constitute a partition of sample space \( S \), i.e. \( E_1, E_2, \ldots, E_n \) are pairwise disjoint & \( E_1 \cup E_2 \ldots \cup E_n = S \) and \( A \) is any event of nonzero probability, then

\[
P(E_i / A) = \frac{P(E_i)P(A / E_i)}{\sum_{j=1}^{n} P(E_j)P(A / E_j)} \quad \text{for any } i = 1, 2, 3, \ldots, n
\]

Proof: By formula of conditional probability, we know that

\[
P(E_i / A) = \frac{P(A \cap E_i)}{P(A)} = \frac{P(E_i)P(A / E_i)}{P(A)} \quad \text{(By multiplication rule of probability)}
\]

\[
= \frac{P(E_i)P(A / E_i)}{\sum_{j=1}^{n} P(E_j)P(A / E_j)} \quad \text{(by total prob.theo.)}
\]
Note: The events \( E_1, E_2, E_3, \ldots, E_n \) are called hypotheses. The probability \( P(E_i) \) is called the priori probability of the hypothesis \( E_i \).

The conditional probability \( P(E_i / A) \) is called a posteriori probability of the hypothesis hypothesis \( E_i \).

4. Consider two boxes. The first one contains 2 white balls and 3 black balls, and the second contains 7 white balls and 5 black balls.

The event \( A_1 \) means choosing the first box, and the event \( A_2 \) means choosing the second box. It is known that the probability of the event \( A_1 \) is \( P(A_1) = 0.4 \), and the probability of the event \( A_2 \) is \( P(A_2) = 0.6 \).
We randomly choose one box and a black ball.
What is the probability that this black ball is chosen from the second box?

**Solution:** Let $X$ be the event "a black ball has been extracted".

By Bayes' formula, we have:

$$P(A_1 / X) = \frac{P(A_1)P(X / A_1)}{P(A_1)P(X / A_1) + P(A_2)P(X / A_2)} = \frac{0.4 \cdot \frac{3}{5}}{0.4 \cdot \frac{3}{5} + 0.6 \cdot \frac{5}{12}} \approx 0.49$$

$$P(A_2 / X) = \frac{P(A_2)P(X / A_2)}{P(A_1)P(X / A_1) + P(A_2)P(X / A_2)} = \frac{0.6 \cdot \frac{5}{12}}{0.4 \cdot \frac{3}{5} + 0.6 \cdot \frac{5}{12}} \approx 0.51$$
2. In a bolt factory, machines A, B & C manufacture respective by 25%, 35%, 40% of the total. Of their output, 5, 4, 2% are defective bolts. A bolt is drawn from a day's production and found to be defective. What is the prob. That it was manufactured by machine C.

Solution:

Let A, B & C be the events that the parts made by respective machines.

\[ P(A) = 0.25 \quad P(B) = 0.35 \quad P(C) = 0.40 \]

A bolt is drawn and found to be defective. So let D be the event that a bolt drawn from a day's production is defective.

\[ P(D/A) = 0.05, \quad P(D/B) = 0.04, \quad P(D/C) = 0.02 \]
\[ P(D) = P(D/A)P(A) + P(D/B)P(B) + P(D/C)P(C) \]
\[ = 0.05(0.25) + (0.04)(0.35) + (0.02)(0.40) = 0.0345 \]

**If a bolt drawn is defective, what is the prob. That it was manufactured by C?**

\[ P(C/D) = \frac{P(D/C)P(C)}{P(D)} \]
\[ = \frac{(0.02)(0.40)}{0.0345} = 0.23 \]