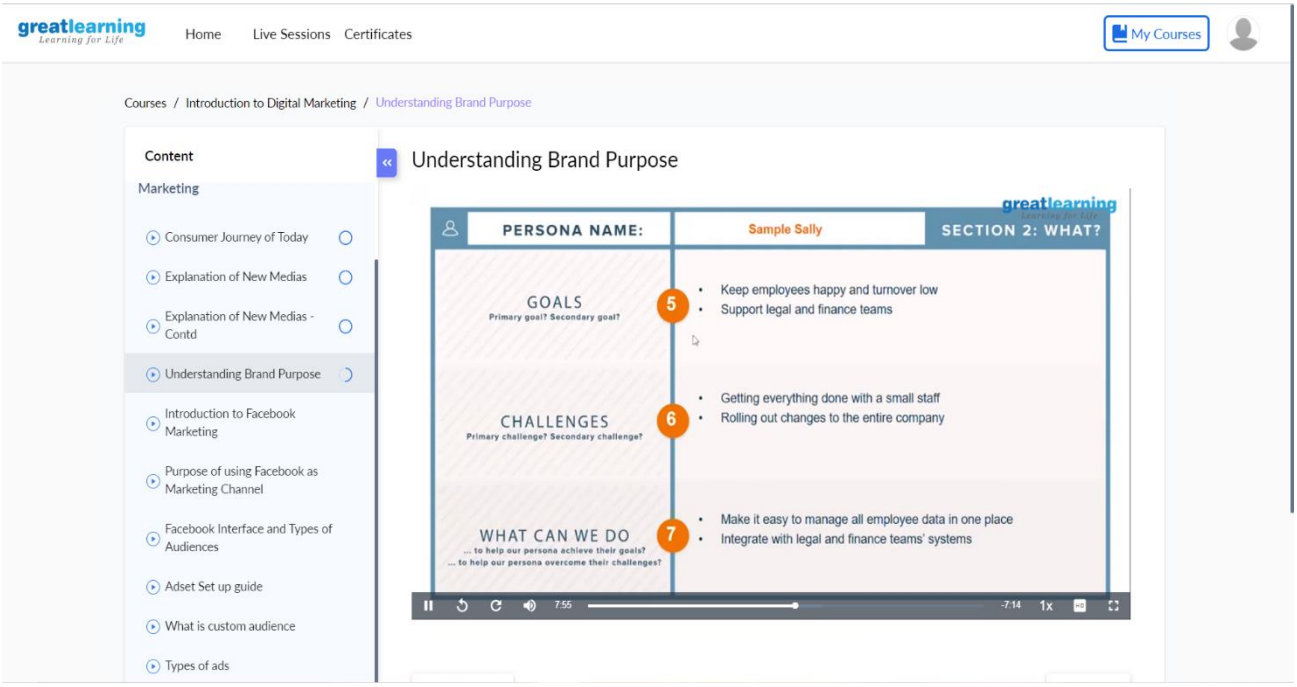
DAILY ASSESSMENT FORMAT

Date:	**16 JUNE 2020**	**Name:**	**PAVITHRAN S**
Course:	**PROBALITY AND STATISTICS**	**USN:**	**4AL17EC068**
Topic:	**PROBABLITY AND STATISTICS**	**Semester & Section:**	**6TH B**
Github Repository:	**Pavithran**		

FORENOON SESSION DETAILS

Image of session



Report – Report can be typed or hand written for up to two pages.

want to discuss some very fundamental terms/concepts related to probability and statistics that often come across any literature related to Machine Learning and AI.

R**andomt**

A random experiment is a physical situation whose outcome cannot be predicted until it is observed.

S**ample**

A sample space, is a set of all possible outcomes of a random experiment.

Example:
Random Experiment: Toss a fair coin once.
Sample Space: Ω ={Head, Tail}

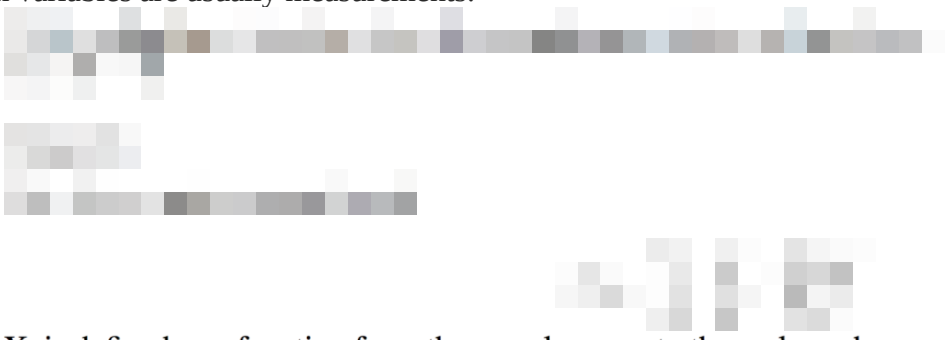
Random

A ***random variable***, is a variable whose possible values are numerical outcomes of a **random experiment**.

There are two types of random variables.

1. ***Discrete Random Variable*** is one which may take on only a countable number of distinct values such as 0,1,2,3,4,…….. Discrete random variables are usually (but not necessarily) counts.

2. ***Continuous Random Variable*** is one which takes an infinite number of possible values. Continuous random variables are usually measurements.



X, is defined as a function from the sample space to the real numbers.
$X : \Omega \rightarrow R$

Example
For above random experiment

$$X = \begin{cases} 1 & if \quad Head \\ 0 & if \quad Tail \end{cases}$$

Probability

Probability is the measure of the likelihood that an event will occur in a Random Experiment. Probability is quantified as a number between 0 and 1, where, loosely speaking, 0 indicates impossibility and 1 indicates certainty. The higher the probability of an event, the more likely it is that the event will occur.

Example

A simple example is the tossing of a fair (unbiased) coin. Since the coin is fair, the two outcomes ("heads" and "tails") are both equally probable; the probability of "heads" equals the probability of "tails"; and since

no other outcomes are possible, the probability of either “heads” or “tails” is 1/2 (which could also be written as 0.5 or 50%).

C**onditional**

Conditional Probability is a measure of the probability of an event given that (by assumption, presumption, assertion or evidence) another event has already occurred. If the event of interest is A and the event B is known or assumed to have occurred, “the conditional probability of A given B”, is usually written as P(A|B).

$$P(A \mid B) = \frac{P(A \bigcap B)}{P(B)}$$

Example

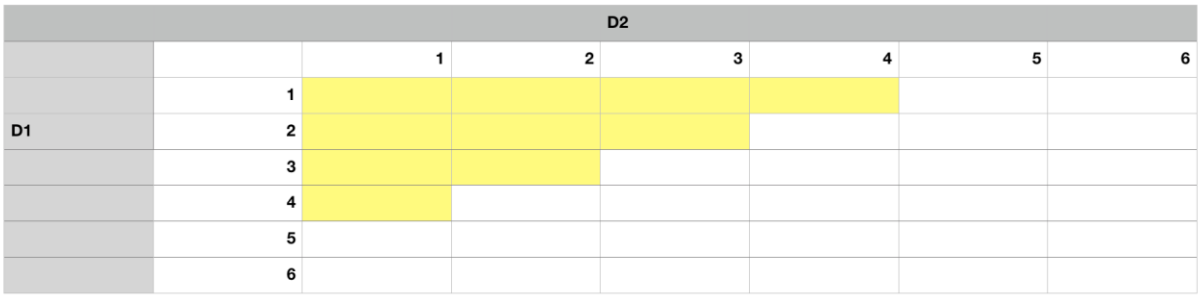
Suppose that somebody secretly rolls two fair six-sided dice,
Let D1 be the value rolled on die 1.
Let D2 be the value rolled on die 2.

What is the probability that D1 = 2

		D2					
		1	2	3	4	5	6
	1						
D1	2						
	3						
	4						
	5						
	6						

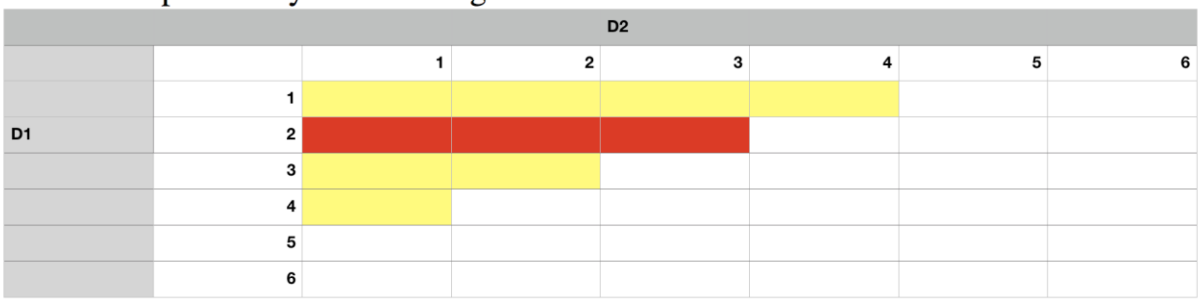
sample space = 36 outcomes. So D1 = 2 in exactly 6 of the 36 outcomes;
P(D1=2) = 6⁄36 = 1⁄6.

What is the probability that D1+D2 ≤ 5



D1+D2 ≤ 5 for exactly 10 of the same 36 outcomes, thus P(D1+D2 ≤ 5) = 10⁄36.

What is the probability that D1 = 2 given that D1+D2 ≤ 5



For D1 = 2, there are 3 out of10 outcomes,
So the conditional probability P(D1=2 | D1+D2 ≤ 5) = 3/10 = 0.3

Independence

Two events are said to be independent of each other, if the probability that one event occurs in no way affects the probability of the other event occurring, or in other words if we have observation about one event it doesn't affect the probability of the other. For Independent events A and B below is true

$$P(A, B) = P(A) * P(B) \quad where \quad P(A) \neq 0 \quad and \quad P(B) \neq 0$$

$$P(A \mid B) = P(A) \quad \text{and} \quad P(B \mid A) = P(A)$$

Example

Let's say you rolled a die and flipped a coin. The probability of getting any number face on the die is no way influences the probability of getting a head or a tail on the coin.

Conditional Independence

Two events A and B are conditionally independent given a third event C precisely if the occurrence of A and the occurrence of B are independent events in their conditional probability distribution given C. In

other words, A and B are conditionally independent given C if and only if, given knowledge that C already occurred, knowledge of whether A occurs provides no additional information on the likelihood of B occurring, and knowledge of whether B occurs provides no additional information on the likelihood of A occurring.

$$P(A \mid C, B \mid C) = P(A \mid C) * P(B \mid C) \quad where \quad P(A \mid C) \neq 0 \quad and \quad P(B \mid C) \neq 0$$

Example

A box contains two coins, a regular coin and one fake two-headed coin (P(H)=1P(H)=1). I choose a coin at random and toss it twice.

Let

A = First coin toss results in an HH.

B = Second coin toss results in an HH.

C = Coin 1 (regular) has been selected.

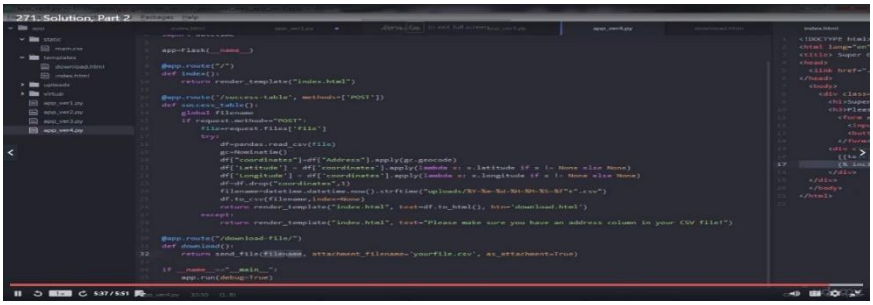
If C is already observed i.e. we already know whether a regular coin is selected or not, the event A and B becomes independent as the outcome of 1 doesn't affect the outcome of other event.

AFTERNOON SESSION

Date:	**16 JUNE 2020**	**Name:**	**PAVITHRAN S**
Course:	**R PROGRAMMING**	**USN:**	**4AL17EC068**
Topic:	**R PROGRAMMING**	**Semester & Section:**	**6TH B**
Github Repository:	**Pavithran**		

FORENOON SESSION DETAILS

Image of session



Report – Report can be typed or hand written for up to two pages.

A data frame is a table or a two-dimensional array-like structure in which each column contains values of one variable and each row contains one set of values from each column.

Following are the characteristics of a data frame.

- The column names should be non-empty.
- The row names should be unique.
- The data stored in a data frame can be of numeric, factor or character type.
- Each column should contain same number of data items.

Create Data Frame

```
# Create the data frame.
emp.data <- data.frame(
   emp_id = c (1:5),
   emp_name = c("Rick","Dan","Michelle","Ryan","Gary"),
   salary = c(623.3,515.2,611.0,729.0,843.25),

   start_date = as.Date(c("2012-01-01", "2013-09-23", "2014-11-15",
"2014-05-11",
      "2015-03-27")),
   stringsAsFactors = FALSE
)
# Print the data frame.
print(emp.data)
```

When we execute the above code, it produces the following result −

```
  emp_id    emp_name     salary     start_date
1      1      Rick        623.30     2012-01-01
2      2      Dan         515.20     2013-09-23
3      3      Michelle    611.00     2014-11-15
4      4      Ryan        729.00     2014-05-11
5      5      Gary        843.25     2015-03-27
```

Get the Structure of the Data Frame

The structure of the data frame can be seen by using **str()** function.

```
# Create the data frame.
emp.data <- data.frame(
   emp_id = c (1:5),
   emp_name = c("Rick","Dan","Michelle","Ryan","Gary"),
   salary = c(623.3,515.2,611.0,729.0,843.25),

   start_date = as.Date(c("2012-01-01", "2013-09-23", "2014-11-15",
"2014-05-11",
      "2015-03-27")),
   stringsAsFactors = FALSE
)
# Get the structure of the data frame.
str(emp.data)
```

When we execute the above code, it produces the following result −

```
'data.frame':   5 obs. of  4 variables:
 $ emp_id    : int  1 2 3 4 5
 $ emp_name  : chr  "Rick" "Dan" "Michelle" "Ryan" ...
 $ salary    : num  623 515 611 729 843
 $ start_date: Date, format: "2012-01-01" "2013-09-23" "2014-11-15"
"2014-05-11" ...
```

Summary of Data in Data Frame

The statistical summary and nature of the data can be obtained by applying **summary()** function.

```
# Create the data frame.
emp.data <- data.frame(
   emp_id = c (1:5),
   emp_name = c("Rick","Dan","Michelle","Ryan","Gary"),
   salary = c(623.3,515.2,611.0,729.0,843.25),

   start_date = as.Date(c("2012-01-01", "2013-09-23", "2014-11-15",
"2014-05-11",
      "2015-03-27")),
   stringsAsFactors = FALSE
)
# Print the summary.
print(summary(emp.data))
```

When we execute the above code, it produces the following result −

```
     emp_id    emp_name             salary        start_date
 Min.   :1   Length:5           Min.   :515.2   Min.   :2012-01-01
 1st Qu.:2   Class :character   1st Qu.:611.0   1st Qu.:2013-09-23
 Median :3   Mode  :character   Median :623.3   Median :2014-05-11
 Mean   :3                      Mean   :664.4   Mean   :2014-01-14
 3rd Qu.:4                      3rd Qu.:729.0   3rd Qu.:2014-11-15
```

```
 Max.   :5                      Max.   :843.2   Max.   :2015-03-27
```

Extract Data from Data Frame

Extract specific column from a data frame using column name.

```
# Create the data frame.
emp.data <- data.frame(
   emp_id = c (1:5),
   emp_name = c("Rick","Dan","Michelle","Ryan","Gary"),
   salary = c(623.3,515.2,611.0,729.0,843.25),

   start_date = as.Date(c("2012-01-01","2013-09-23","2014-11-15","2014-
05-11",
      "2015-03-27")),
   stringsAsFactors = FALSE
)
# Extract Specific columns.
result <- data.frame(emp.data$emp_name,emp.data$salary)
print(result)
```

When we execute the above code, it produces the following result −

```
  emp.data.emp_name emp.data.salary
1              Rick          623.30
2               Dan          515.20
3          Michelle          611.00
4              Ryan          729.00
5              Gary          843.25
```

Extract the first two rows and then all columns

```
# Create the data frame.
emp.data <- data.frame(
   emp_id = c (1:5),
   emp_name = c("Rick","Dan","Michelle","Ryan","Gary"),
   salary = c(623.3,515.2,611.0,729.0,843.25),

   start_date = as.Date(c("2012-01-01", "2013-09-23", "2014-11-15",
"2014-05-11",
      "2015-03-27")),
   stringsAsFactors = FALSE
)
# Extract first two rows.
result <- emp.data[1:2,]
print(result)
```

When we execute the above code, it produces the following result −

```
  emp_id    emp_name   salary    start_date
1      1     Rick      623.3     2012-01-01
2      2     Dan       515.2     2013-09-23
```

Extract 3rd and 5th row with 2nd and 4th column

```
# Create the data frame.
emp.data <- data.frame(
```

```
   emp_id = c (1:5),
   emp_name = c("Rick","Dan","Michelle","Ryan","Gary"),
   salary = c(623.3,515.2,611.0,729.0,843.25),

       start_date = as.Date(c("2012-01-01", "2013-09-23", "2014-11-15",
"2014-05-11",
      "2015-03-27")),
   stringsAsFactors = FALSE
)

# Extract 3rd and 5th row with 2nd and 4th column.
result <- emp.data[c(3,5),c(2,4)]
print(result)
```

When we execute the above code, it produces the following result −

```
  emp_name start_date
3 Michelle 2014-11-15
5     Gary 2015-03-27
```

Expand Data Frame

A data frame can be expanded by adding columns and rows.

Add Column

Just add the column vector using a new column name.

```
# Create the data frame.
emp.data <- data.frame(
   emp_id = c (1:5),
   emp_name = c("Rick","Dan","Michelle","Ryan","Gary"),
   salary = c(623.3,515.2,611.0,729.0,843.25),

   start_date = as.Date(c("2012-01-01", "2013-09-23", "2014-11-15",
"2014-05-11",
      "2015-03-27")),
   stringsAsFactors = FALSE
)

# Add the "dept" coulmn.
emp.data$dept <- c("IT","Operations","IT","HR","Finance")
v <- emp.data
print(v)
```

When we execute the above code, it produces the following result −

```
  emp_id   emp_name     salary     start_date       dept
1      1    Rick         623.30     2012-01-01       IT
2      2    Dan          515.20     2013-09-23       Operations
3      3    Michelle     611.00     2014-11-15       IT
4      4    Ryan         729.00     2014-05-11       HR
5      5    Gary         843.25     2015-03-27       Finance
```

Add Row

To add more rows permanently to an existing data frame, we need to bring in the new rows in the same structure as the existing data frame and use the **rbind()** function.

In the example below we create a data frame with new rows and merge it with the existing data frame to create the final data frame.

```
# Create the first data frame.
emp.data <- data.frame(
   emp_id = c (1:5),
   emp_name = c("Rick","Dan","Michelle","Ryan","Gary"),
   salary = c(623.3,515.2,611.0,729.0,843.25),

   start_date = as.Date(c("2012-01-01", "2013-09-23", "2014-11-15",
"2014-05-11",
      "2015-03-27")),
   dept = c("IT","Operations","IT","HR","Finance"),
   stringsAsFactors = FALSE
)

# Create the second data frame
emp.newdata <-      data.frame(
   emp_id = c (6:8),
   emp_name = c("Rasmi","Pranab","Tusar"),
   salary = c(578.0,722.5,632.8),
   start_date = as.Date(c("2013-05-21","2013-07-30","2014-06-17")),
   dept = c("IT","Operations","Fianance"),
   stringsAsFactors = FALSE
)

# Bind the two data frames.
emp.finaldata <- rbind(emp.data,emp.newdata)
print(emp.finaldata)
```

When we execute the above code, it produces the following result −

```
  emp_id     emp_name     salary     start_date       dept
1      1     Rick         623.30     2012-01-01       IT
2      2     Dan          515.20     2013-09-23       Operations
3      3     Michelle     611.00     2014-11-15       IT
4      4     Ryan         729.00     2014-05-11       HR
5      5     Gary         843.25     2015-03-27       Finance
6      6     Rasmi        578.00     2013-05-21       IT
7      7     Pranab       722.50     2013-07-30       Operations
8      8     Tusar        632.80     2014-06-17       Fianance
```

CERTIFICATE FOR DIGITAL MARKETING