CLASSTIME Pg. No. 1
Date / /

Week-1

Video-1 2/7/21

1. Introduction

- understand basic statistics (art of learning from data).
- learn to collect data ⇒ organize data ⇒ describe data
  descriptive statistics
  (probability th.) inferential stats ⇐

Video-2 L1.1

2. Week wise course

- intended for beginners, create & handle data sets (numerically and graphically), appropriate numerical summaries
- Graphical & Numerical Summaries Data | Probability Basic notions & conditional probability | Random vars. · Discrete → Binomial · Continuous → Normal

  Week ⇒ 1-4 | 5-7 | 8-12

Video-3 L1.2

3. Types of Data - Basic Definitions

- Statistics is the art of learning from data. It is concerned with the collecⁿ of data, their subsequent descrip. and their analysis, which often leads to the drawing of conclusions [Sheldon M. Ross]
- Description → Descriptive Stats - concerned with the descripⁿ and summarizaⁿ of data
  (Two branches)
- Inference → Inferential Stats - drawing of concluⁿ from data
  ↳ possibility of chance (Probability)

  Popula^n
  Represents
  Sample
- Popula^n - The total collecⁿ of all the elements that we are interested it.
- Sample - A subgrp. of popula^n that will be studied in detail.
- If the purpose of analysis is to examine and

explore info. for its own intrinsic interest only, the study is descriptive

- If info. is obtained from a sample of popula and purpose of study is to use that info. to draw conclusion abt the popula", the study is inferential.
- A descriptive may be performed either on a sample of popula".
- When an inference is made abt. the popula", based on info. from sample ⇒ inferential stats

Video 4 2.1.3

4. Intro. & types of data - Understanding data

- Data - are the facts and fig. collected, analysed and collected ⇒ summarized for presenta" and interpreta"
- stats relies on data, info. that is around us.
- why do we collect data?
  ⇒ interested in the charac. of some grp. or grps. of people, place, things or events.
  ⇒ Eg. to abt. T in a month in Chennai, etc.
- data collec" → data available → published data (data.gov.in) → " not " → need to collect, gen. data
- we assume that data available and our obj. is to do a statistical analysis of available data.
- unstructured / unorganized data ⇒ no meaning for us
- for the info. in a db to be useful, we must know the context of nos. and text it holds.
- when scattered ⇒ of very little use for us
- ∴, we need to organize data.
- Dataset ⇒ a structured collec" of data.
  ⇒ it is a collec" of values ⇒ nos., names, etc.
- case (obs.): A unit from which data are collected
- variable ⇒ a variable is that "varies."

O } this is a data

-- } data is not available in that regard

⇒ a charac. or attribute that varies across all units

→ in school data → case → each student

→ variable → name, marks, etc.

- rows ⇒ represent cases, for each case same attribute is recorded
- columns ⇒ variables. For each variable, same type of value for each case is recorded (consistency of units)

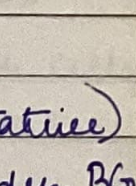
⇒ Each var. must have its own column.

⇒ Each obs. / case must have its own row.

Video 5. L1.4

5. Classifica^n of data

- data → categorical and numerical (quantitative)
- categorical ⇒ qualitative variables Eg. Gender, BG (M/F) (CBSE/ICSE)
  ⇒ identify group membership.
- numerical → quantitative variables → discrete (played match) / continuous (avg.)
  ⇒ describe numeric prop. of cases
  ⇒ have measurement units
- measurement units - scale that defines the meaning of numerical data, wt. (kg), prices (₹), ht. (cm)
- the data that make up a numerical variable in a data table must share a common unit
- one variable - time series - data recorded over a time
- timeplot - graph of a time series showing values in a chronological order
- cross-sectional - data observed at the same time



Video 6 L1.5

6. Scales of Measurement

- data collec^n req. one of the following scales of measurement - nominal, ordinal, interval or ratio.

- Nominal - when a data for a var. consists of labels or names used to identify the charac. of an obs. Eg. Name, Board, Gender, BP, etc.
  (order is not impt.) - no ordering
  - sometimes it can be numerically codes M→0 F→1
  - name categories w/o implying order
- ordinal - data exhibits prop. of nominal and order / rank of data is meaningful.
  - Eg. customer giving feedback in hotel 5>4>3>2>1
  - name categories that can be ordered
- interval - all the prop. of ordinal and the interval btw. values is expressed in terms of a fixed unit of measure
  - are always numeric. Can find diff.
  - ratios have no meaning, becoz value of 0 is arbitrary
  - numerical values can be added / subtracted
- Eg. Temperature
  - just feeling (comfortable or not) → nominal
  - a liq. (cold < warm < hot) → ordinal
  - AC room (T = 20°c), outside is 40°c, then the difference is 20°c ⟶ interval
    - but u cannot say outdoor is twice of AC.
- ratio - all prop. of interval and ratio of 2 values is meaningful. Eg. Ht, Wt., Highest score
  - numerical values → arithematic opⁿˢ +, −, ÷, ×

+ − ÷ ×	True 0 exist, ratio possible	Ratio	Age, Ht, Wt. Numerical
− +	No abs. 0, diff. exists	Interval	Temp., GPA
Order	Named + order	Ordinal	Ranking Categorical
Nothing	Named	Nominal	Name, BG

Google Sheet

Video-7

7. Tutorial - 1

- Google sheets → our course → Personal (logged in) → New spreadsheet
- each box is cell → navigate by arrow keys
- name → Name of spreadsheet
- grp → same vertical seq. (column), horizontal (row)
- enter text/nos. into these cells → Enter
- Eg. Simple Interest r of i = 0.5% / month

✓Month	✓Interest	✓Total Int.	Total Money Own
March	0	0	10000
April	0.5% = 50	50	10050
May	50	100	10100

- utility ⇒ autofill → click, hold drag ⇒ fills all the cells
- it also catches patterns

Video-8

8. Tutorial - 2

- represent data in a more meaningful way
- label the columns (✓)
- add row above first row → Right click [Insert]
- some spilling over
- Select cells (shift →) ⇒ Fill color (any color) ⇒ Ctrl B
- text wrapping → [...] → wrap (multiple line)
- inc. cell width ⇒ click, hold, drag
- select all cells ⇒ more ⇒ center align → vertical
- select no. cells ⇒ Format ⇒ Number ⇒ More Format ⇒ More Currency ⇒ Indian ₹

Video 9 Google sheet

9. Spreadsheet Formulae (Tutorial 3)
- remove old entry, add for march first
- $P = 10,000$, $r = 0.5\%$
- go to cell '=' 10000 (x) 0.5/100 ⇒ ₹50
  formula → * ↳ Autofill
→ next becomes total int. ⇒ = 63 + C2
- total money owed = P + R ⇒ = d2 + C3
  ↳ Problem it should be P + I
  (we need to change c and not d_2)
  ↳ so do it D$2 → keeps the cell fixed

- going back to interest paid make it = D$2 * 0.5/100
- cell ref. is very useful. P amt. all become automatically updated. Eg P = 20,000
- Rate also get fixed in F column → 0.5 % (G1)
- Ctrl c → ctrl V ⇒ copies format ⇒ looks nice
- D$2 * G1
- Edit ⇒ Paste spe ⇒ Ctrl + Alt + V ⇒ same format copy
- ₹ 10,000 P amt. ⇒ G2

Video 10

10. Downloading & uploading spreadsheets
- download spreadsheet File → Download (all kinds)
- data. gov. in ⇒ comes as .csv file
- home → blank spreadsheet → ctrl O → .csv file from net
- My Drive (Google) → New Folder (Create)
  Click, hold & drag in the new folder

Week 2

CLASSTIME Pg. No. 7
Date / /

L2.1 Describing categorical data → Freq. distribuⁿ

- descriptive, inferential stats, sample, populaⁿ, variables, cases, categorical, numerical, cross sectional, time series, measurement scales
- Freq. distribuⁿ - is a listing of distinct values & their freq. (count)
- each row of freq. table lists a category along with no. of cases (count) in this category
- construct a freq. distribuⁿ ⇒ AABCADDABBC

1 → list distinct values of obs.
2 → for each obs., place a tally mark
3 → count the no. of tallies

A	IIII	4
B	II	2
C	II	2
D	II	2

⇒ AABCADABDCABCDA ⟶ ABBCADBBDCABCDB

A	卌I	6
B	III	3
C	III	3
D	III	3
	15	15

⟺ 6 times

A	III	3
B	卌I	6
C	III	3
D	III	3
	15	15

Google Sheet

freq. table in googlesheet

1 → select the cells → highlight
2 → formatting bars → Data opⁿ
3 → data opⁿ → Pivot table opⁿ & create a sheet
4 → Pivot table editor and in that add rows & values (at side)

Category
A
A
B
C
A

another way to create freq. table

- we can do for categorical data → in hospital data set. Eg. BG
- gives count of each categorical variable
- Relative freq. - ratio of freq. to total no. of obs.

1 → obtain freq. distribuⁿ table
2 → divide each freq. by total no. of obs.

	F	RF
A	4	0.4
B	2	0.2
C	2	0.2
D	2	0.2
	10	1

	F	RF
A	6	6/15 = 0.4
B	3	3/15 = 0.2
C	3	3/15 = 0.2
D	3	3/15 = 0.2
	15	15/15 → 1

↳ always come to 1

- why do we need relative freq.?
  - → to compare 2 data sets (0 → 1) always
  - → provide a standard comparison
- to Rel. Freq. in google sheet Eg. 4/10 = B8/B12

→ ~~AAAAADAABDCABCDACDD~~

A	卌 I	6	6/18
B	III	3	3/18
C	IIII	4	4/18
D	卌	5	5/18
		18	1

in google sheet

L2.2 **Charts of Categorical Data**

- 2 most common ways to display categorical variable - bar chart & pie chart
- both describe its freq. table
- pie chart - a circle divides into pieces (wedges) $\propto$ to relative freq. of qualitative data

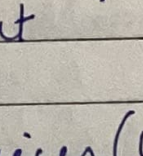
1 → obtain a rf. table
2 → divide the circle into rf
3 → label the wedges

rf	Degree
$0.4 \times 360°$	$= 144°$
0.2×360	$= 72°$
$0.2 \times$ "	= "
$0.2 \times$ "	= "
	$360°$

- pie chart → gives the share of a particular thing in π

Google Sheet

pie chart in google sheet
1 → select the cells
2 → Insert charts options
3 → change visualiza"
4 → chart editor chart type → π chart



- good way to show that 1 category make up more than half of the total
- bar chart - displays the distinct value of qualitative data on x-axis and rfs on a vertical (y-axis). → can be horizontal or vertical

Google sheet - google sheet 1 → highlight
2 → choose column chart in chart opⁿ

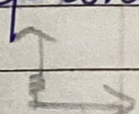
- in bar chart, we know exact count of that thing, easily annotate things, whereas, π chart gives %.
- ~~Pattos~~ Pareto chart - when the categories of bar chart are sorted by freq. → we called Pareto chart
  → if it ordinal, then bar chart must have ordering

L2.3 Best prac. while graphing data - (Part 1)
- have a purpose for every graph u create
- choose the graph to serve the purpose
- for just counting ⇒ tabulate the freq. count
- pie chart - compare parts of a whole
- bar chart - exact count, compare things btw. diff. grps
- big data, lot of categories ⟶ tabulate it
- google sheet - in google sheet, bring chart, chart customize, title, subtitle, legend, chart color (do not take same colour), do not move pies outside of a circle → to give some meaning, label, font-size
- add borders, 3D looks good but doesn't give extra value
- for column chart, we can do for several attributes, horizontal / vertical axis title, data labels, posⁿ of label in bar
- horizontal lines are called grid lines that is also choosable in customize
- many categories → the graph is cluttered

- In some case, gruping other smaller categories Together might be done. This helps by not concealing data.

L2.4 Best prac. - (Part - 2)

- area prin. → display of data must obey a fundamental rule c/d area prin
- area occupied by a part of the graph should correspond to the amt. of data it presents
- decorated graphics → charts decorated to attract attenⁿ often violate the area prin
- violaⁿˢ of area prin. leads to misleading graphs
- truncated graphs — when the baseline of bar chart is not at 0.
- for truncate graphs - indicate a y-axis break



L2.5 Mode & Median

- round-off errors - impt. to check for round-off errors → leads to wrong because they do not add up to 100% / 1.
- it'll not form a pie chart
- need for a compact measure
- descriptive measure - nos. used to describe data sets
- descriptive measure that indicate center or most typical value of set → measures of central tendency.
- arithmetic operaⁿ are not allowed

- Mode – most common category, the category with highest freq.
  - labels → 1. longest bar of bar chart
  - 2. widest slice of π chart
  - 3. Pareto chart, the mode is shown first
- if 2 or > categories tie with highest frequency, the data → bimodal or multimodal.
- Median – only ordinal data offers this, data has to be put in order. Of an ordinal variable is the category of middle obs. of sorted values

~~A, B, B, C, A, D, B, B, A, C, B, B, C, D, A~~

~~A, A, A, A, B, B, B, B, B, B, B, C, C, C, D, D~~ ⇒ B

- in some sense, median helps to divide in equal halves.

⇒ A, B, B, C, A, D, B, B, A, C, B, B, C, D, A

A = 4 Mode Median = B

B = 6

C = 3

D = 2

⇒ A B B C A D A B A C B A C D A

A = 6 B = 4 C = 3 D = 2 Mode = A

Median = B

- they can be different or same

⇒ Week 2 → Google sheet Tutorial 1

- no. of songs vs genres in playlist

(a) no. of songs → hindustani genre

$10x = 0.1111 \times 10$

$10x = 1.1111$

$9x = 1$

$x = \frac{1}{9}$

16.7%

16.7% × 90

$\frac{167}{10 \times 100} \times 90 = 15.03 \sim 15$ songs

CLASSTIME Pg No.
Date / /

(b) no. of songs — Indian Pop

$\frac{7.6}{100 \times 12} \times 92$ 78 $\times 9$ 702 ~ 7 songs

(c) modal genre → indian rock

(d) app. genres G_1, G_2, G_3, G_4

G_3 → Indi Rock
G_2 → Folk
G_1 → Punjabi
G_4 → Film soundtrack

$\frac{24}{90}$

$\frac{111}{100 \times 12} \times 92$ 9.99 ~10

$\frac{15.6}{100 \times 92} \times 92$ 156 $\times 9$ 1404

→ Tutorial
- misleading → not start with 0.
- select → Insert → chart → customize
- misleading → % do not add up to 100%

→ Tutorial - SUMIF in Google Sheets
- sumif google sheet → support → SYNTAX
- summing with an if, criterion applied

5 items 5 category 1 210 210
1 140 140
1 79 79
1 20 12 100 24 ← total of stationary

= sumif(B3 : B7, "Stationary", E3 : E7)
from where it is selected — string — summing range

= sumif(C3 : C7, "1", E3 : E7)
">1" ⇒ for more than quantity
↳ if not then error

→ Tutorial - VLOOKUP in Google Sheets

- func^n → VLOOKUP → vertical look up → support google sheets
- we want place of card no. 7 ?
  - vlookup ("Place", range, index, [is sorted])
    - "Place" → value to search for
    - range → range of data
    - index → index column
    - [is sorted] → if arranged in an order TRUE; if not in order → FALSE

"DOB"

- vlookup ("Phy", A1:AE7, 23, FALSE)
  - "Phy" → Type the actual value
  - FALSE → since the first column is not sorted

- If you do
  - vlookup ("Bio", . . . , . . . , False)
  - we didn't get any value → do Teuse in deep approx. value

- vlookup gives the value first found in the column → be careful.

Week 3

Describe Numerical Data – Freq. Tables

L3.1

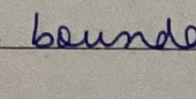
- numerical → discrete & continuous
  Marks scored (68, 75, 35) (68.5, 72.5) kg, wt
- discrete → count ; continuous → measurement
- grp. obs. into classes (categories / bins) and treat them as qualitative data
- construct, rf & freq. distribution table
- small no. of values → treat each distinct as a category → freq. table

~~2 x 3 4 5 2 3 3 3 4 x 2 3 4~~

we can plot it ~~in the bar chart~~ | each as a category

	f	rf
1	2/15	0.133
2	3/15	0.2
3	5/15	0.333
4	3/15	0.2
5	1/15	0.066
		1

- few guidelines for continuous data & large no. of categories
  1. no. of classes → 5 – 20
  2. each obs. should belong to some class
  3. no obs. should belong to >1 class
  4. common to choose class intervals of equal length
- lower class limit – smallest value gets into class
- upper c.l. – highest value of class
- class width – diff. b/w lower limit of a class and lower limit of next class
- class interval has its left end and not right end boundary pt.
  30 → 40
  30, 35, ~~40~~
- steps for a histogram (most popular for cont.
  1 → obtain freq. / rf.
  2 → x-axis place the classes, ↑ display freq.
  3 → each class, create a bar for the corresp. freq.



Google sheet — on google sheet, 50 students marks → Insert → Chart → Histogram → Range (Use Row 1 as Header) → it choose random Class Interval
- customize → bucket size → Auto → 10, Δ other attributes also
- Stem & leaf diag. Stemplot - each obs. is seperated into 2 parts
  1 → stem of all
  2 → leaf of right-most digit

eg. data are all 2-digits nos. stems → Tens place
75, 78 leaf → ones place

stem	leaf
7	5, 8

Steps
- 1 → each obs. as a stem & leaf with right most digit
  2 → write stems from smallest to largest in a vertical colum
  3 → each leaf acc. to appropriate stem
  4 → arrange leafs in asc. order

⇒ 15, 22, 29, 36, 31, 23, 45, 10, 25, 28, 48

Stem	Leaf	Final
1	5, 0	0, 5
2	2, 9, 3, 5, 8	2, 3, 5, 8, 9
3	6, 1	1, 6
4	5, 8	5, 8

L3.2 Mean

- measures to summarize data set
- measures of central tendancy - most typical value or center of dataset
- measures of dispersion - variabity / spread of a dataset
- most common central tendency → Mean
- Mean = $\frac{\text{Sum of Obs.}}{\text{Total obs.}}$ = Avg = $\bar{x} = \frac{x_1 + x_2 \dots x_n}{n}$

- $n \rightarrow$ sample size, $N \rightarrow$ population size
- population mean : $\mu = \frac{x_1 + x_2 \dots x_N}{N}$

$\Rightarrow$ 2, 12, 5, 7, 6, 7, 3 $\quad \bar{x} = \frac{\cancel{42}}{\cancel{7}} = 6$

$\Rightarrow$ 2, 105, 5, 7, 6, 7, 3 $\quad \frac{135}{7} = 19.28$

$\rightarrow$ 2, 105, 5, 7, 6, 3 $\quad \frac{128}{6} = 21.33$

Google Sheet — in google sheet, avg. command = mean

x_i	f_i	$f_i x_i$
1	2	2
2	3	6
3	5	15
4	4	16
5	1	5

$x_i = \frac{f_1 x_1 + f_2 x_2}{\cancel{n}}$

$$\bar{x} = \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i}$$

- mean for continous data

CI	f_i	m_i (mid-pt.)	$f_i m_i$
30-40	3	35	105

~~this~~ is approx. and not real mean value

$x_i = \frac{f_1 m_1 + f_2 m_2 \dots f_n m_n}{n}$

- what happens when we add a constant
  $y_i = x_i + c \Rightarrow \bar{y} = \bar{x} + c$
- multiply with a constant
  $y_i = x_i c \Rightarrow \bar{y}_i = \bar{x} c$

CLASSTIME	Pg. No.
Date / /	

L3.3 Median and Mode

- median divides the data set into half. Median of a data set is the middle value of a data set
- Steps 1 → arrange in ascending order
  2 → if n = (no. of obs) odd, the median is the middle most → $\frac{n+1}{2}$

  3 → if n = even, median is the mean of 2 middle elements → $\frac{n}{2}$ & $\frac{n}{2}+1$

Eg. 2, 12, 5, 7, 6, 7, 3 ~~2, 3, 5, 6, 7, 7, 12~~ → 6

Eg. 2, 105, 5, 7, 6, 7, 3 ~~2, 3, 5, 6, 7, 7, 105~~ → 6

Eg. 2, 105, 5, 7, 6, 3 ~~2, 3, 5, 6, 7, 105~~

$\frac{5+6}{2}$ $\frac{11}{2}$ → 5.5

Median need not belong to data set.

- The mean is sensitive to outliers, whereas median isn't sensitive to outliers.

Google Sheet = func^n → =median (Range)

- $y_i = x_i + c$ ⇒ new median = old median + c
- $y_i = x_i c$ ⇒ new median = old median × c
- mode → is the most freq. occurring value
- Steps 1. no value has > 1, then no mode
  2 → else with greatest freq. is the mode of dataset

Eg. 2, 12, 5, 7, 6, 7, 3 → 7

- $y_i = x_i + c$ ⇒ new mode = old mode + c
- $y_i = x_i c$ ⇒ new mode = old mode × c

Google Sheet = mode (Range)

L3.4 Measures of dispersion

	DS1	DS2
Mean	3	3
Median	3.	3
Mode	3	–

- DS 1 → 3, 3, 3, 3, 3
  DS 2 → 1, 2, 3, 4, 5

- the datasets are different
- to describe this diff. quantitatively, we use descriptive measures → measures of dispersion / measures of variaⁿ / measure of spread
- Range, Variance, Std. deviaⁿ, Interquartile Range (most common)
- Range - ~~diff.~~ btw. largest & smallest value
  DS1 (Range) = 0 DS2 = 4
  → It is sensitive to outliers
  → It takes only extreme data.
- Variance - take into accnt. all obs., we can measure variability is to consider deviaⁿ of data value from a central value (Mean)

Populaⁿ Var. $\sigma^2 = \dfrac{(x_1-\mu)^2+(x_2-\mu)^2+\ldots+(x_N-\mu)^2}{N}$

Sample Var. $s^2 = \dfrac{(x_1-\bar{x})^2+(x_2-\bar{x})^2+\ldots+(x_n-\bar{x})^2}{n-1}$

Q. 68, 79, 38, 68, 35, 70, 61, 47, 58, 66   Mean = 59

9 20 -21 9 -11 2 -12 1 0

81 400 441 81 121 4 144 1 = 1898

$s^2 = \dfrac{1898}{9} = 210.88$ $\sigma^2 = \dfrac{1898}{10} = 189.8$

Google Sheet

funcⁿ ⇒ VAR.S (Range) (sample), VAR.P (Range) (Populaⁿ)

- $y_i = x_i + c$ ⇒ old variance = new variance
- $y_i = x_i c$ ⇒ new variance = old variance × c^2

CLASSTIME Pg. No.
Date / /

- Std. Deviaⁿ - sq. root of sample varⁿ

$$s = \sqrt{\frac{(x_1-\bar{x})^2 + (x_2-\bar{x})^2 \ldots}{n-1}}$$

- unit of avg. of dataset is same as unit of data
- unit of variance $\Rightarrow (kg)^2$
  std. deviaⁿ $\Rightarrow \sqrt{(kg)^2} \Rightarrow kg$
  ↳ same units as original data.
- $y_i = x_i + c \Rightarrow$ new devⁿ = old devⁿ
- $y_i = x_i c \Rightarrow$ new devⁿ = old std devⁿ × c
- STDEV(Range) → Standard deviaⁿ funcⁿ

L3.5 Percentile, Quartile & Interquartile Range

- Percentile - The sample 100percent(p) percentile has prop at least 100p% of the data are less than or = to it and at least 100(1-p)% are greater than = to it
  $p = 1/2$ 50% | 50% (50 percentile) ↳ median of data set
  → that value which differentiates
  ↳ if 2 values satisfy this condⁿ, then 100p percentile is the arith. avg. of these values
- To find sample 100p percentile of a dataset n
  1. → order → ascending
  2. → if np is not an integer, determine > np. The data value is 100p percentile
  3. → np is Z, avg. of np & np+1 in the sample 100p percentile

Eg. n = 10   35, 38, 47, 58, 61, 66, 68, 68, 70, 79

p	np	
0.1	1	(35+38)/2 = 36.5
0.25	2.5	47
0.5	5	(61+66)/2 = 63.5
0.75	7.5	68
1	10	79

Google Sheet

1. Paste dataset in column
2. =Percentile (data range, percentile 0.1)

For google sheet ⇒ rank = [percentile × (n − 1)] + 1

3. split the rank into int & frac. part
4. compute the ordered data value x_i corres. to int. part rank
5. percentile = x_i + frac. part × [$x_{[i+1]} - x_{[i]}$]

- Quartile – sample 25th percentile → 1st Quartile
  " 50th " → median / 2nd "
  " 75th " → 3rd Quartile (upper)
  ↳ they break the data into 4 parts

min — Q_1 — median (Q_2) — Q_3 — max

- 5 No. summary
  - min
  - Q_1 quartile
  - median / Q_2 quartile
  - Q_3 quartile
  - max

- IQR (Interquartile Range)
  $$IQR = Q_3 - Q_1$$

Google Sheet

func^n ; Quartile (Range ; 1)
Quartile (Range ; 3)
IQR ($Q_3 - Q_1$)
Quartile (Range ; 2) ⇒ Median

→ Tutorial 1

Eq. (a) avg. speed of vehicles

speed	cars		CI	
10-20	4	×	15	60
20-30	7	×	25	175
30-40	5	×	35	175
40-50	3	×	45	135
50-60	6	×	55	330
60-70	5	×	65	325
	30			

avg. speed = 40 km/hr

(we will not take 40-50)

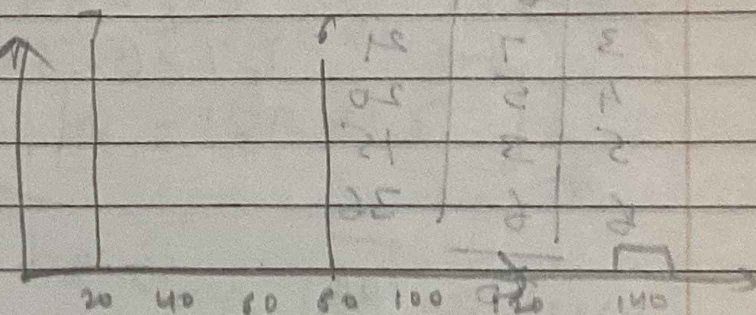
(b) exceed speed limit by atleast 10 kmph ~~3+~~6+5 = 14-3 = 11

(c) 40 > vehicles speed ≥ 20 kmp 7+5 = 12

→ Tutorial 2

Eq.

20	~~||||~~ ||
40	~~||||~~
60	||||
80	~~||||~~ |
100	||||
120	|||
140	|



→ Tutorial 3

Eq.

1	1 2 4
2	0 5 6 7
3	1 2 3 7 8
4	0 1 3 5
5	2

avg. attendance = ?

$\frac{527}{17} = 31$

Date / /

→ Tutorial -4

Eq. Mean of 25 obs. = 36 ← Mean First 13 obs. = 32
last 13 obs. = 39

What is 13th obs.

$\frac{sum}{25} = 36$ $\frac{sum_F}{13} = 32$. $\frac{sum_L}{13} = 39$

sum = 900 $sum_F = 416$ $sum_L = 507$

13th obs. = 923 - 900 = 23

→ Tutorial - 5

Eq.

No.	F	
1	4	4
2	6	12
3	7	21
4	5	20
5	3	15
6	6	36
	30	108

$\bar{x} = 3.4$

$4 \times (1 - 3.4)^2$
$6 \times (2 - 3.4)^2$
$7 \times (2 - 3.4)^2$

1.65 (a) sample std. deviaⁿ
2.731 (b) sample var
3 (c) sample mode

⇒ Tutorial -6

Eq. Rajan's no. is 80th percentile. n = 12
grades → 46, 85, 68, 93, 84, 70, 38, 66, 78, 75, 55, 60
Rajan's grade = ?

n = 12 p = 0.8
np = 9.6

38, 46, 55, 60, 66, 68, 70, 75, 78, 84, 85, 93
1 2 3 4 5 6 7 8 9 10 11 12

→ Tutorial – 7

eg. 5, 9, 13, 15, 17, 3, 5, 1 Std. devia^n

$\bar{x} = 8.5$

12.25
0.25
20.25
42.25
72.25
30.25
12.25
56.25

246

$\frac{246}{7}$, $\sqrt{35.142} = 5.928$

→ Tutorial (Outliers)

eg. if outliers are excluded → mean

$\bar{x}_{old} = 37.75$

$\bar{x}_{new} = 35.142$

$[Q_3 + IQR \times 1.5]$

$[Q_1 - IQR \times 1.5]$

→ for outliers

→ Tutorial (Box Plot)

eg. 12, 16, 22, 8, 36, 34, 40, 32, 28

8, 12, 16, 22, 28, 32, 34, 36, 40

8 12 28 34 40

Box → IQR

40 − 8

32

0.5×9

4.5

0.75×9

6.75

CLASSTIME Pg No.
Date / /

Week - 4

L4.1 Review of the course
- association btw. 2 variables
- stats - descriptive, inferential
- data collecⁿ, tabulaⁿ (row → cases)
- categorical / numerical → discrete, conti
- nominal, ordinal, interval & ratio
- describe categorical data → freq, rf., π chart, bar chart, mode, median
- numerical data → freq., mean, median, mode, range, variance, std. deviaⁿ, percentile, quartile, interquartile range, histogram, stem leaf plot

L4.2 Associaⁿ btw. 2 categorical variables - Intro
- use of 2-way contingency tables, scatter plots, compute & interpret correlaⁿ

Eg. Gender v/s use of smartphone - who owns a smartphone

100 college going students G Own
M/F Y/N

Gender - M/F (2) - Nominal
Own - Y/N (2) - Nominal

1. 44 F & 56 M] Total = 100 students
2. 76 Own a smartphone, 24 do not have
3. 34 F own, 42 M own smartphone

- a Contingency tables helps in the associaⁿ

(2-way table) 2nd var →

		N	Y	Row T
Gender (1st var)	I F	10	34	44
	II M	14	42	56
	Colum	24	76	100

Google Sheet: click the data, select ⇒ Data → Pivot Table

Date / /

- Existing → New locⁿ
- Rows → click 1st categorical var ; Columns → click 2nd categorical var. ⇒ Values → add

Eg. Income v/s use of smartphone
High Med. Low (Ordinal var.) — Y N (categorical var.)

- in ordinal data, u should maintain the order
- coded → high → 1, med → 2, low → 3

4.3 Relative freq.
- how many proporⁿ who own a phone → 76/100
- propⁿ of female who own → 34/44
- Row relative freq. = divide each cell freq. in a row by its row total. ⇒ we can do it in % also
- prop. of total are female → 44/100
- prop. of female who are phone owners → 34/76
- column relative freq. ⇒ divide each freq. by its column-total
- knowing info. abt. 1 var. provides info. abt. the other variable ; associaⁿˢ are described by fu
- if row / column relative freq. are same for rows / columns then we say that 2 var. are not associated & vice versa.

Eg. our Gender v/s Phone, ownership patterns are consistent, hence, gender and phone are not related to each other

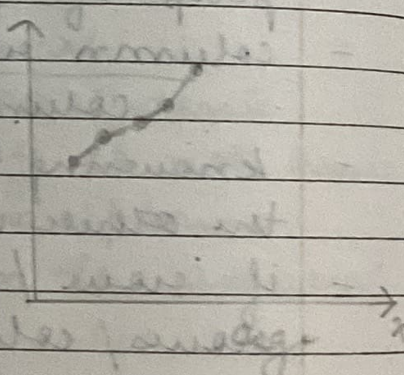
- stacked / segmented bar chart — counts for a particular category, each bar is further segmented into smaller segments representing the freq. of that particular category within the segment

Google Sheet - select contingency table → Insert chart → stacked column chart (std.) → count
helpful in knowing part-to-whole relationship { (100%) → % how much proportion
↳ the association is much more clearer

L4.4 Scatterplot

- It is a graph that displays pair of values as pts. on 2d plane

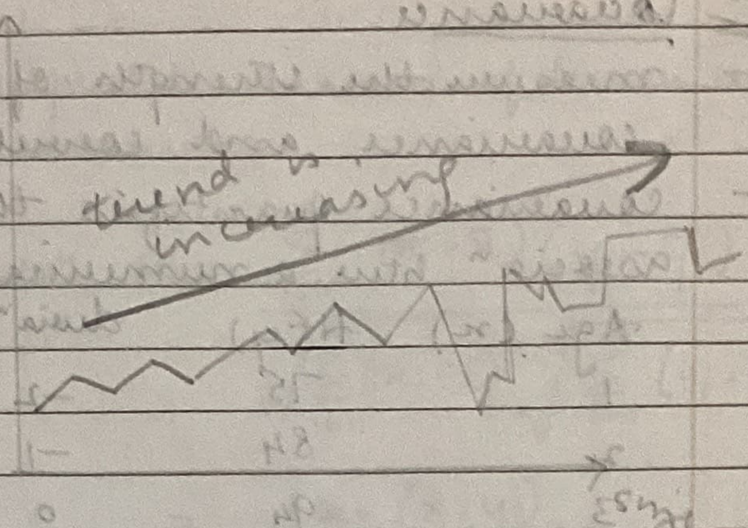
Age (yrs)	ht (cms)	
1	75	(1, 75)
2	85	(2, 85)
3	94	(3, 94)
4	101	(4, 101)
5	108	(5, 108)



- x-axis — explanatory var.
  y-axis — response var.

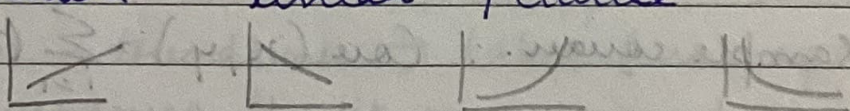
Eg. relationship btw. price of home (in lakhs) → response
v/s size of home (1000 sq. ft.) ↳ explanatory

15 data pts.

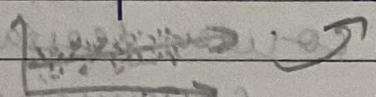


Google Sheets
- select data → choose → Insert chart → scatterplot → series → y → response
- visual test for associa^n → find patterns

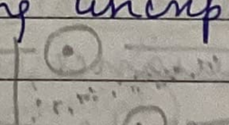
L4.5 Describing associa^n
- 4 ques → direc^n → pattern trend ↑, ↓
  - curvature → linear / curve
  - varia^n → tight / clustered
  - outliers → anything unexpected
- direc^n → as size ↑, price ↑, uptrend
- curvature → linear / curve



- varia^n → pts. are tightly clustered / scattered



- outliers → anything unexpected



Date / /

L4.6 **Covariance**

- measure the strength of association by - covariance and correlation
- covariance quantifies the strength of linear association btw. 2 numerical variables

Eg. Age (x)	Ht(y)	deviaⁿ of x $x_i - \bar{x}$	devⁿ of y $y_i - \bar{y}$
1	75	-2	-17.6
2	84	-1	-7.6
3	94	0	1.4
4	101	1	8.4
5	108	2	15.4
$\bar{x}$ 3	$\bar{y}$ 92.6		

$(x_i - \bar{x})(y_i - \bar{y})$
35.2
7.6
0
8.4
30.8

- When large (small) values of x_i are associated large (small) values of y, the signs of deviaⁿs are also same. & vice-versa
- Populaⁿ covariance: $Cov(x,y) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{N}$

  sample covar.: $Cov(x,y) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n-1}$

Google Sheet - select the data ⇒ covar / covariance.P ⇒ popⁿ
covariance.S ⇒ sample
(Range, Range)

- size of covariance, is difficult to interpret as covariance has units

L4.7 Correla^n

- more easily interpreted measure of linear rela^n
- derived from covariance

Pearson correla^n $r = \dfrac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})}\sqrt{\sum_{i=1}^{n} (y_i - \bar{y})}}$

the units cut out

$= \dfrac{cov(x, y)}{S_x S_y}$ std. devia^n

$\boxed{-1 \le r \le 1}$

Google sheet: =correl(data y, data x)

strong -ve linear association: -1 —— 0 (no associa^n) —— 1: strong +ve linear associa^n

L4.8 Fitting a line

- linear associa^n using eq^n of line

Google sheet: select → Insert → scatter plot → customize series → trendline → label (use equa^n) → show R^2

- R^2 → goodness of fit measure [0 → 1] → good fit
  - ↳ [0 → 1]

L4.9 Associa^n btw. numerical & categorical

- assume categorical var. has 2 subtypes (dichotomous)
- Gender (cat) v/s Marks (numerical) → gender → coded $F \to 1$, $M \to 0$
- =correl(Gender, Marks) → gives an error
- scatterplot

CLASSTIME Pg No.
Date / /

- Point Biserial correla^n coefficient
1. categ. variable $Y \leq^{0}_{1}$
2. cal. mean of 2 grps $\bar{Y}_0$ $\bar{Y}_1$
3. p_0 & p_1 be the proportion of obs. in a grp.
4. S_x → std. dev. of numerical var.

$$\boxed{r_{pb} = \frac{\bar{Y}_0 - \bar{Y}_1}{S_x} \sqrt{p_0 p_1}}$$

$p_0 = \frac{n_0}{n}$

$p_1 = \frac{n_1}{n}$

⇒ Tutorial-1

① $\frac{835}{900}$ $\frac{77}{100}$ $\frac{\cancel{110}\ 55}{\cancel{200}\ 100}$

0.927 0.77 0.55

Student

② $\frac{835}{1022} = 81.7\%$ → ~~students of total~~

③ ~~S~~ 100 × 835 = 83,500 Student

S 340 × 77 = 26,180

Rs 240 × 110 = 26,400

⇒ Tutorial-2

Eg. sample correlation coeff.

$$\text{correl} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum x_i - \bar{x}}\ \sqrt{\sum y_i - \bar{y}}} = \frac{\text{cov}(x, y)}{S_x S_y}$$

	$T(x)$	$N(y)$	a $(x_i - \bar{x})$	b $(y_i - \bar{y})$	$(a \times b)$
1	293	1500	-8.125	611.75	–
2	295	1300	-6.125	411.75	–
3	299	1150	-2.125	261.75	–
4	300	800	-1.125	-88.25	+
5	305	500	3.875	-388.25	–
6	297	1200	-4.125	311.75	–
7	308	456	6.875	-432.25	–
8	312	200	10.875	-688.25	–

$$\frac{-21196.25}{7} = -3028.03$$

✓ $\bar{x} = 301.125$

✓ $\bar{y} = 888.25$

$$S_x = \sqrt{\frac{\sum_{i=1}^{8} (\bar{x} - x_i)^2}{7}} = 6.621$$

$$\text{correl} = \frac{-3028.03}{6.621 \times 467.291} \approx -0.978$$

$S_y = 467.291$

$r \approx -1$ strong linear rela^n T ↑ , N ↓

⇒ Tutorial - 3

1. associa^n +ve , fairly strong
2. $S_x = 1.93$ $\quad$ cov $(x, y) = 13.96$

   $S_y = 9.36$

$$\text{correl} = \frac{13.96}{1.93 \times 9.36} = \frac{13.96}{18.0648} = 0.7727$$

fairly strong, +ve associat^n

⇒ Tutorial - 4

F	46
F	47
F	~~2~~40
M	34
M	18
M	22
F	45
F	50
F	55
F	60
F	69
M	34
M	36
M	35
F	70
F	75
F	80
M	28
M	44
M	33

11 → F
9 → M

$\bar{F} = \frac{637}{11} = 57.9$

$\bar{M} = \frac{284}{9} = 31.55$

$r_{pb} = \sqrt{pq} \; \frac{\bar{Y}_0 - \bar{Y}_1}{S_x}$

$S_x = 17.47 \Rightarrow \frac{57.9 - 31.55}{S_x} \sqrt{0.55 \times 0.45}$

$= \frac{26.35}{17.47} \times 0.497$

$= \underline{\underline{0.75}}$

fairly large, +ve
gender & monthly income

$F_{\sim} > M_{\sim}$

Week-5

CLASSTIME Pg. No.
Date / /

L 5.1 Permutaⁿ & Combinaⁿ – Basic Prin. of Counting

- counting with order (permutaⁿ) & counting w/o order (combination)
- 4 shirt choices or 3 pant choice → 4+3=7
  ↳ the actions are dependent on each other
- Addⁿ rule → if acⁿ A can occur in n_1 ways & acⁿ B can occur in n_2 ways, total ways of occurance of acⁿˢ A or B is $\boxed{n_1 + n_2}$
- what happens if 1 shirt & 1 pant ?
  4 shirts | 3 pant   1≤ 1≤ 1≤ 1≤ → 12
- Multiplicaⁿ rule → if acⁿ can occur in n_1 diff. way, & acⁿ B in n_2 diff. way, so together A and B is $\boxed{n_1 \times n_2}$
- if we have n acⁿˢ in a definite order ...
  $n_1 \times n_2 \times \dots \times n_n$

Eg. 6 digit password → first 2 alpha + 4 digits

$\underbrace{\_\ \_}_{26\ \ 26}\ \underbrace{\_\ \_\ \_\ \_}_{\text{numbers } (0 \to 9)}$

repⁿ ⇒ 26 × 26 × 10 × 10 × 10 × 10
no repⁿ ⇒ 26 × 25 × 10 × 9 × 8 × 7

L5.2 Factorials

Eg. 100 m race → 8 people → possible ways to finish
(no ties) → clear ranking
(8!) 8 × 7 × 6 × 5 × 4 × 3 × 2 × 1 = 40,320

- Factorial – prod. of 1ˢᵗ n +ve integers is c/d n factorial and is denoted by n!
  ⇒ $n! = n \times (n-1) \times \dots \times 1$
  ⇒ $0! = 1$; $1! = 1$

Eg. 3 shirts → 3 people → possible ways
$3! = 3 \times 2 \times 1 = 6$

Eg. $5! = 5\times4\times3\times2\times1 = 120$; $5! = 5\times4!$

$6! = 6\times\underbrace{5\times4\times3\times2\times1}_{5!} = 6\times5!$

$$\boxed{n! = n\times(n-1)!}$$

- for $i \le n$

$n! = n\times(n-1)\ldots\times(n-i+1)\times(n-i)!$

Eg. $\frac{6!}{3!}$ $= \frac{6\times5\times4\times\cancel{3\times2\times1}}{\cancel{3\times2\times1}} = 120$

Eg. $\frac{6!\times5!}{3!\times4!}$ $= \frac{6\times5\times\cancel{4!}\times5\times4\times\cancel{3!}}{\cancel{3!}\times\cancel{4!}}$

$= 600$

Eg. Express $25\times24\times23$ in terms of factorials

$\frac{25!}{22!}$

⇒ Tutorial 1

Eg. In a building 6 men & 4 women. ⇒ 3 married couples. In how many ways u can guess the couples?

M_1W_1 $\quad m_1\ m_2\ m_3\ m_4\ m_5\ m_6$

M_2W_2 $\quad w_1\ w_2\ w_3\ w_4$

M_3W_3

3 womens are married

$w_1w_2w_3$

$w_2w_3w_4$

$w_1w_3w_4$

$w_1w_2w_4$

① → 6
② → 5
③ → 4

$120 \times 4 = 480$

Date / /

→ Tutorial 2

Q. 4 dice a rolled. no. of obs. on each face of dice. How many ways each dice shows diff. no.

$$\_\ \_\ \_\ \_$$

$$6 \times 5 \times 4 \times 3 = 360$$

→ Tutorial 3

Q. dice & coin is tossed. diff. pairs of outcome are there.

1 → 6

T / H

$$6 \times 2 = 12$$

→ Tutorial 4

Q. 4 digit nos. are possible by 0 → 9. if 1st & 2nd nos. are 2 & 8 respectively. (Rep^n not allowed)

$$\underline{2}\ \underline{8}\ \underset{8}{\_} \times \underset{7}{\_} = 56$$

→ Tutorial 5

Q. 5 letter blocks. How many diff. kind of words >4 letter w/o repitition

$$\underset{20}{5 \times 4} \times \underset{60}{3} \times \underset{120}{2} \times 1 = 120\ (5!) \qquad 4! \times 5 = 120$$

$$\Rightarrow 240$$

→ Tutorial 6

Q. 4 digit password, nos. (0-9) or alpha (small & large) (Rep^n not allowed)

$$26+26+10 \Rightarrow 62$$

$$(52 \times 51 \times 50 \times 49) + (10 \times 9 \times 8 \times 7)$$

→ Tutorial-7

Eg. 4 digit password → alphanumeric password
Can't use small alphabets & only 1st digit as alphabet (rep^n not allowed)

$$26 \times 10 \times 9 \times 8$$

⇒ Tutorial-8

Eg. same ques^n but rep^n allowed

$$26 \times 10 \times 10 \times 10$$

Week - 6

L6.1 Permutation, distinct objects

- permutation - an ordered arrangement of all or some of n objects (distinct) [repⁿ not allowed]

A | B | C
3 distinct people

Eg. Take A, B & C. Possible - taking all at a time

$n=3$, 1 ABC, 2 ACB, 3 BAC, 4 BCA, 5 CAB, 6 CBA — possible arrangements = 6

- taking two at a time

$n=3$, $r=2$: 1. AB, 2. BA, 3. BC, 4. CB, 5. AC, 6. CA — possible arrangements = 6

- There are n objs, 1 in choosing r obj. from n, ..., $r \le n$
- The no. of possible permutation of r obj from a collection of n distinct objects is given by, $n \times (n-1) \times (n-2) \times \ldots \times (n-\{r-1\})$

^nP_r

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

$0! = 1$

Spl. case

- $^nP_0 = \frac{n!}{(n-0)!} = \frac{n!}{n!} = 1$, ⇒ only 1 ordered arrangement of 0 obj.
- $^nP_1 = \frac{n!}{(n-1)!} = n$ ⇒ There are n ways of choosing 1 obj. from n obj.
- $^nP_n = \frac{n!}{(n-n)!} = n!$ ⇒ multiplication rule of counting

Eg. comittee of 8 people, choose chairman & vice-chairman, and post shouldn't be repeated

$n=8$, $r=2$: $\frac{8!}{(8-2)!} = \frac{8!}{6!} = 8 \times 7 = 56$

Eg. find no. of 4 digits no. from 1, 2, 3, 4, 5

$^5P_4 = 5! = 120$

Eg. How many of above are even?

last

$n = 3 \Rightarrow {}^4P_3$

× ___ ___ ___ [2]

× ___ ___ ___ [4]

$\Rightarrow {}^4P_3$

$4 = \frac{4!}{(4-3)!} = 4! = 4 \times 3 \times 2 = 24 \times 2 = 48$

Eg. 6 go to cinema. They sit in a row with 10 seats

(i) sit anywhere

$n = 10$

$r = 6$

$^{10}P_6 = \frac{10!}{4!} = 10 \times 9 \times 8 \times 7 \times 6 \times 5 = 1,51,200$

(ii) all the empty seats are next to each other

$n = 7$

$r = 6$

$^7P_6 = 7! = 720 \times 7 = 5,040$

- if somehow rep^n is allowed, like we have A, B & C, taking all at a time $3^3 = 3 \times 3 \times 3 = 27$
- A, B & C, taking 2 at a time, rep^n allowed $3^2 = 9$
- n objects choose r objects (rep^n allowed) $r \leq n$ $\boxed{= n^r}$

L 6.2 Permutation: Objects not distinct & circ. permutation

Eg. DATA. How many ways the letters can be arranged

$\frac{4!}{2!} = 12$

D, A, T

- the no. of permutations of n objects when p of them are of 1 kind and rest distinct,

$$\boxed{\frac{n!}{p!}}$$

Eg. STATISTICS . Ways?

S, T, A, I, C
↓ ↓ ↓ ↓ ↓
3 2 1 2 1

$\frac{10!}{3!\,3!\,2!}$ $\qquad \frac{10\times9\times8\times7\times6\times5\times4\times3\times2}{3\times2\times3\times2\times2}$

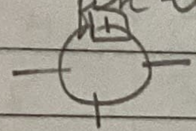
$10\times9\times8\times7\times10 = 50,400$

- the no. of permutations of n obj. where p_1 is of one kind, p_k of k kind,

$$\boxed{\frac{n!}{p_1!\,p_2!\ldots p_k}}$$

Eg. ways four people sit in a round table?
↳ clockwise & anticlockwise are diff.

like this



then, $3! = 3\times2 = 6$ ways

- When there are n obj. in a circle, also, clockwise & anticlockwise are diff. is given by, $\boxed{(n-1)!}$

- when clockwise & anti are same, then.

$$\boxed{\frac{(n-1)!}{2}}$$

Eg. n, $^nP_4 = 20\,^nP_2$ $\qquad ^nP_r = \frac{n!}{(n-r)!}$

$\frac{\cancel{n!}}{(n-4)!} \quad 20\frac{\cancel{n!}}{(n-2)!} \Rightarrow (n-2)! = 20(n-4)!$

$(n-2)(n-3) = 20$

$n^2 - 5n + 6 = 20 \qquad n^2 - 5n - 14 = 0$

$(n+2)(n-7) = 0$

$n = 7, -\cancel{2}$

Eg. $\dfrac{^nP_4}{^{n-1}P_4} = \dfrac{5}{3} \Rightarrow 3\ ^nP_4 = 5\ ^{n-1}P_4$

$\dfrac{3 \times n!}{\cancel{(n-4)!}} = \dfrac{5\ \cancel{(n-1)!}}{\cancel{(n-4)!}} = 3 \times n \times \cancel{(n-1)!}$

$\dfrac{n}{n-4} = \dfrac{5}{3}$

$3n = 5n - 20$

$2n = 20 \qquad n = 10$

Eg. $^5P_r = 2 \times\ ^6P_{r-1} \qquad \dfrac{5!}{(5-r)!} \quad \dfrac{2 \times 6!}{(7-r)!}$

$\dfrac{1}{\cancel{(5-r)!}} = \dfrac{2 \times 6}{(7-r)(6-r)\cancel{(5-r)!}}$

$(7-r)(6-r) = 12$

$r^2 - 13r + 30 = 0$

$r^2 - 3r - 10r + 30 = r(r-3) - 10(r-3)$

$r = 3, \cancel{10}$

$r \leq n$

6.3 Combina^n

Eg. we select 2 students from a grp. of 3 students

A, B, C $\qquad \underbrace{AB \quad BC \quad AC}_{3}$ order is not impt

- each selec^n is c/d a combina^n
- each combina^n of r obj. from n obj. can give rise to r! arrangements, no. of possible combina^n of r obj. from n distinct obj

CLASSTIME Pg. No.
Date / /

$$\boxed{{}^nC_r = \frac{n!}{r!\,(n-r)!}}$$

- another common notaⁿ is $\binom{n}{r}$ and is also c/d binomial coefficient.
- $\boxed{{}^nC_r = {}^nC_{n-r}}$ if I select r obj., then I'm rejecting $(n-r)$ obj.
- $\boxed{{}^nC_n = 1}$ & $\boxed{{}^nC_0 = 1}$
- $\boxed{{}^nC_r = {}^{n-1}C_{r-1} + {}^{n-1}C_r}$ $1 \leq r \leq n$

Eg. ques. paper - 12 ques, I → 7, II → 5. A student has to attempt 8 ques, selecting at least 3 from each. In how many way?

$$8 \begin{cases} P_I & P_{II} & \\ 3 & 5 & {}^7C_3 \times {}^5C_5 \\ 4 & 4 & {}^7C_4 \times {}^5C_4 \; + \\ 5 & 3 & {}^7C_5 \times {}^5C_3 \end{cases}$$

35 + 175 + 210 = 420

- deck of 52 playing cards ⟹ 4 suits

♣ ♠ ♥ ♦

club (13), spade (13) — Black; heart (13), diamond (13) — Red

52 → 26 Black, 26 Red

Eg. Choose any 4 cards from deck
$${}^{52}C_4 = 2,70,725$$

Eg. choose 4 cards from same suit
$${}^{13}C_4 = \frac{13 \times \cancel{12} \times 11 \times \cancel{10}^{5}}{\cancel{4 \times 3 \times 2}} = 13 \times 55 = 715 \times 4 = 2860$$

55
× 13
165
55×

Eg. all cards from same colour

$$^2C_1 \times {}^{26}C_4 = 29,900$$

Eg. cricket team of 11 ← 17 players, only 5 can bowl. The req. exactly 4 bowlers. How many ways?

$$^5C_4 \times {}^{12}C_7 = 5 \times 792 = 3960 \text{ ways}$$

17
12 5
7 4

Eg. n pts. on a circle, how many lines segment can be drawn connecting these pts

$\boxed{{}^nC_2}$

n = 2 n = 3

 line = 1

 lines = 3

16.4 Applicaⁿ

- permutaⁿ - order matters; combinaⁿ - order doesn't matter

Eg. 8 athletes with several rounds

1. How many diff. ways can u award G, S, B
2. How " ways can u choose top 3 for next round

1. $^8P_3 = \frac{8!}{(8-3)!} = 8 \times 7 \times 6 = 56 \times 6 = 336$

2. $^8C_3 = \frac{8!}{3!5!} = \frac{8 \times 7 \times \cancel{6}}{\cancel{6}} = 56$

Eg. class with 40

(i) choose 2 leaders

(ii) choose a captain & vice-captain

(i) $^{40}C_2 = \frac{40!}{2!\,38!} = \frac{\cancel{40}^{20} \times 39}{\cancel{2}} = 780$

(ii) $^{40}P_2 = \frac{40!}{38!} = 1560$

Eq. given n pts, how many lines can be drawn

$$^nC_2 \times 2! = 2\,^nC_2 = \,^nP_2$$

⇒ Tutorial-1

Eq. permutations of word "GRAPH" are arranged in dictionary of word. What is 73rd word

$5! = 5 \times 4 \times 3 \times 2 = 120$

1 AGHPR

2 AGHRP

3 AGPHR

:

120 RPHGA

_ _ _ _ _

A $4 \times 3 \times 2$ = 24

G = 24 48

H = 24 72

$\boxed{\text{PAGHR}}$ → 73rd word

⇒ Tutorial -2

Eq. 5 spl dishes in 10 dishes. 7 dishes such that at least 3 dishes are spl and spl. are served consequtively

I 5 spl. 2 ordinary $\boxed{5}$ O O → 3!

5! 1 $^5C_2 = 10$ 1 2 3

$3! \times 5! \times 10$

$= 7200$

II 4 spl 3 ordinary

4! 1 $^5C_3 = 10$

$4! \times 4! \times 5C_4 \times 5C_3$

$= 28800$

III 3 spl 4 ordinary

3! 1 $^5C_4 \quad \frac{5!}{4!} = 5 \quad 5 \times 3! \times 5! \times {}^5C_3$

$5 \times 6 \times 120 \times 10 = 36000$

Ans $= 7200 + 28800 + 36000 = 72000$ ways

CLASSTIME Pg No.
Date / /

⇒ Tutorial - 3

Eg. - coin tossed 7 times, no. of outcome with utmost 3 heads is 64 T/F ✓

0 heads $^7C_0 = 1$
1 heads $^7C_1 = 7$
2 heads $^7C_2 = 21$
3 heads $^7C_3 = 35$

$\} 64$

- fair die is rolled 3, outcome sum of the results should be odd ✗

$\underbrace{x_1 + x_2 + x_3}_{\text{odd or 2 even \& 1 odd}} = \text{odd}$

$x_1 = 3$
$x_2 = 3$
$x_3 = 3$
$= 27$

$x_1 = 3$
$x_2 = 3$
$x_3 = 3$
$3 \times 3 \times 3 \times 3C_1$
$= 81$

$27 + 81 = 108$

- no. of ways 1 Indian & 1 American ~~&~~ grp. from 3 Indian & 4 American is 105 ✓

3 I, 4 A ⇒ 7

$^7C_1 + {^7C_2} + {^7C_3} + \dots + {^7C_7} = 2^7 - 1$
$= 127$

1 Indian $3C_1$
2 Indian $3C_2$
3 Indian $3C_3$

$\} = 2^3 - 1$
$= 7$

$4C_1 \cdots 4C_4 = 2^4 - 1$
$= 15$

$127 - 22 = 105$

- 2 adults & 3 children in 12 ways such adults are always together ✓

CLASSTIME Pg. No.
Date / /

Four adults

A 1 1 4 $(4-1)! = 3! \times 2 = 6 \times 2$

1 6 $= 12$

→ Tutorial - 4

Eq. 4 black & 16 brown dogs, no three in a straight line. Suddenly, 2 brown followed 1 black ⇒ collinear. 3 blacks are following 1 brown. How many straight lines possible

$${}^{n}C_2 = {}^{20}C_2 = \frac{20!}{2! \times 18!} = \frac{20 \times 19}{2} = 190$$

$$- {}^{3}C_2 + 1 \qquad 190 - 3 + 1 - 6 + 1$$

$$- {}^{4}C_2 + 1 \qquad 190 - 7 = \underline{183}$$

→ Tutorial - 5

Eq. arrange r from n people along circular, seating is same if person has same neighbours

$$\begin{matrix} & A & \\ B & & C \\ & D & \end{matrix} = \begin{matrix} & A & \\ C & & B \\ & D & \end{matrix} \qquad {}^{n}P_r = \frac{n!}{(n-r)!}$$

(For this ques) $\boxed{\frac{{}^{n}P_r}{2r!}}$

→ Tutorial - 6

Eq. A → 2, first is vowel & exactly 8 letters btw. A & B

$\underset{1}{\boxed{A}} \underbrace{\quad}_{8} \underset{10}{\boxed{B}} \underbrace{\quad\quad}_{16}$

$4 \times 16 \times 2 \times 23!$

$\Rightarrow 24! + 23! [128]$

$24!$

CLASSTIME Pg. No.
Date / /

Week-7

L7.1 Probability - Basic Defⁿ

- the determinaⁿ of chance, that an event will occur is the matter of probability
- expt. is any process that prod. an obs. or outcome
- random expt. is an expt. whose outcome is not predictable with certainty. We know the results of expt. in advance.

Eg. 1 guess ans. four mcq → A, B, C, D
2. order of finish race of six students - A, B, C, D, E, F
3. Toss 2 coins → HH, HT, TH, TT

- sample space (Ω / S) : collecⁿ of all basic outcom
- Basic outcome - the possible outcome that can occur must be -
  1. mutually exclusive - only 1 basic outcome can occur
  2. exhaustive - one basic outcome must occur

Eg. 1. S = {A, B, C, D}   2. S = {A, B, C, D, E, F} 6! Permutaⁿ of   3. {HH, HT, TH, TT}

L7.2 Events

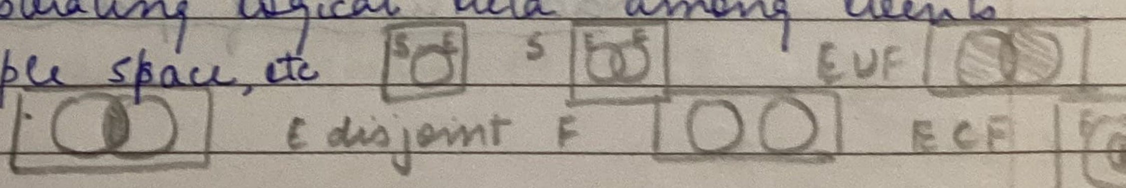
- Event E is a collecⁿ of basic outcome
- it is a subset of sample space
- we say event has occurred if the outcome is contained in the subset
- For events E & F, $\boxed{E \cup F}$, to consist all outcomes that are in E or in F or in both E & F
  - ↳ E ∪ F will occur if E / F occurs
- $\boxed{E \cap F}$, consists of all outcomes that are in E and F

Date / /

- <u>Null Event</u> - Event w/o any outcomes ϕ
- <u>Disjoint / Mutually Exclusive Event</u> - if $E \cap F$ is null event, then E and F can't occur simultaneously, thus, E and F are disjoint. Occurrence of E disallows the occurrance of F.
- $\boxed{E^c}$ - all outcomes in the sample space that are not in E
- S^c is always ϕ (null set).
- $\boxed{E \subset F}$ - if all outcomes of E are in F, we say E is contained in F or E is a subset of F

L7.3 <u>Venn Diag.</u>

- Ace, 2 — 10, Jack, Queen, King
- random expt. will 1 card from these 52 cards
- sample $S = \{$collec^n of all 52 cards$\}$
- $E_1 = \{$heart king$\}$ $E_2 = \{$a king $= 4\}$ $E_3 = \{$hearts $= 13\}$
- $E_2 \cup E_3 = \{16$ cards$\}$
- ~~A~~ $E_2 \cap E_3 = \{1$ heart king$\} = E_1$
- $E_4 = \{\overset{4}{\text{Ace}}\}$ $E_3 = \{13$ hearts$\}$ ⇒ no they are not mutually exclusive
- $E_5 = \{4$ Queens$\}$ $E_2 = \{4$ kings$\}$ ⇒ yes, they are mutually exclusive
- <u>Venn diag</u> - a graphical representa^n that is useful for illustrating logical rela^n among events
- for sample space, etc



L 7.4 **Prop. of Probability**

- classical / Apriori / theoretical - S sample space, in n equally likely outcome, and event E consist of exactly m of these outcomes, $\boxed{P(E) = \frac{m}{n}}$

- relative / aposteriori / emperical - prob. of an E in an expt. is the propor^n of times the E occurs in a very long series of repi^n of expt. $\boxed{P(E) = \lim_{n \to \infty} \frac{n(E)}{n}}$

- subjective - p of E is 'best guess' by a person. The probability measures an indi's degree of belief in an event.

- 3 prop. (axioms) of P(E) →
① for any event, $0 \leq P(E) \leq 1$
② $P(S) = 1$, probability of sample space. The outcomes are elements of sample space
③ for disjoint events, $E_1 E_2$ ...

$$\Rightarrow \boxed{P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i)}$$

- $P(E^c) = 1 - P(E)$
- $P(\phi) = 0 \Rightarrow \phi \Rightarrow$ null event
- E_1, E_2 are non-disjoint events, then

$$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$$

↳ Add^n rule of probability

L7.5 Applica^n of Probability

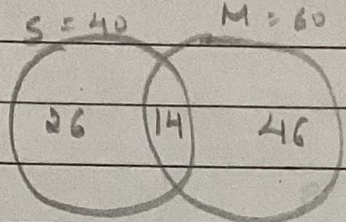
Eg. $P_S = 0.3$ $P_P = 0.2$ $P_{P\cap S} = 0.1$ $P\left(\substack{\text{neither}\\ S/P}\right) = ?$

$0.3 + 0.2 - 0.1 = 0.4$ $P(S \cup P)^c$

↳ 0.6

Eg. 40% A in stats, 60% A in maths, 86% A in stats / maths

1. does not receive an A in stats / maths 0.14
2. receive A in both stats & maths. 0.14



L7.6 Equally likely outcomes

- sample space, S has N equally likely outcome
  $S = \{1, 2, 3, \dots, N\}$,
  $P(\{1\}) = P(\{2\}) = \dots = P(\{N\})$
- thus, probability of any event from n outcome $= \boxed{\frac{1}{n}}$

$$P(A) = \frac{\text{no. of outcomes in } S \text{ that are in } A}{N}$$

Eg. probability of a card drawn from 52 is either red or queen?

26 red 4 queen ⇒ $26 + 2 = \frac{\cancel{28}\ \cancel{14}\ 7}{\cancel{52}\ \cancel{26}\ 13} = \frac{7}{13}$ ✓

→ Tutorial -1

Eg. ~~4 letters~~ 3 digits followed by 4. P → first 2 letters as ~~vowels~~ vowels and later 2 are consonants (rep^n not allowed)

□ □ □ □
V V C C

$$\frac{\cancel{5}}{26} \times \frac{\cancel{4}}{\cancel{25}_{5}} \times \frac{\cancel{21}}{\cancel{24}_{6}} \times \frac{\cancel{20}^{4}}{23} = \frac{7}{13 \times 23} = \frac{7}{299}$$

≈ 0.02341

→ Tutorial - 2

Eg. Engineer inspects pair of nut & bolt. In 200 pairs, 8 are defectives. Randomly selected 10 pairs, what is prob. at exactly 3 pairs would be defectives in 10 pairs

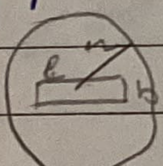
200 —

$\frac{8}{200}$

$$\frac{{}^{8}C_3 \times {}^{192}C_7}{{}^{200}C_{10}} \approx 0.00426$$

→ Tutorial -3

Eg. Circ. board of rad. r, a player wins if she hits a central rec. (l × b). Counting cases where dart hits the board, what is winning probabity



$$\frac{lb}{\pi r^2}$$

→ Tutorial - 4

Eg. 30000 students join a course. Gives 1000 students a feedback form. S_1 & S_2 are in course. What is probability S_1 got feedback form & S_2 didn't

S_1 $\quad \frac{1000}{30000}$ $\qquad$ S_2 $\quad \frac{29000}{30000}$

Together $\frac{1}{30} \times \frac{29}{30} = \frac{29}{900}$

→ Tutorial - 5

Eg. $P(C) = 0.3$

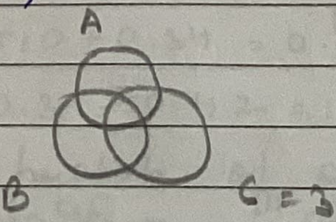
$P(A \cup B) = 0.6$

$P(A \cup B \cup C) = 0.8$

$P((A \cup B)^c \cap C) = ?\ 0.2$

$0.8 - 0.3 = 0.5$

$A + B = 6$



$0.6 - 0.5 = 0.1$

$0.3 - 0.1 = 0.2$

$A + B + C = 8$

→ Tutorial - 6

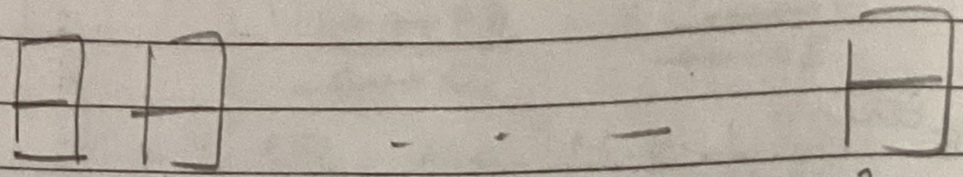
Eg. upper shelf - 7 books ME & 5 books EE
He selected 3 books at random. What is prob. 1 of ME & 2 EE

$$\frac{{}^5C_2 \times {}^7C_1}{{}^{12}C_3} = \frac{10 \times 7}{220} = \frac{7}{22} = 0.318$$

Date / /

→ Tutorial – 7

Eg – 20 rooms with 2 beds/room. 20 civil & 20 CS students are allotted. If pairing is random, prob. of no civil & CS room pair



$^{40}C_2$ $\quad$ $^{38}C_2$ $\quad \dots \quad$ $^{2}C_2$

$$^{40}C_2 \times {}^{38}C_2 \times \dots \times {}^{2}C_2 = \frac{40!}{2^{20}}$$

civil

$$\frac{20!}{2^{10}}$$

CS

$$\frac{20!}{2^{10}}$$

$$\frac{\frac{20!}{2^{10}} \times \frac{20!}{2^{10}} \times {}^{20}C_{10}}{\frac{40!}{2^{20}}}$$

$$= \frac{(20!)^3}{40!\,(10!)^2}$$

Week-8

L8.1 Conditional Probab. – Contingency Tables

- find the probab. of event given another event has occured

count	N	Y	
M	10	34	44
F	14	42	56
	24	76	100

- Joint probabilities

	N	Y	
F	0.1	0.34	0.44
M	0.14	0.42	0.5
	0.24	0.76	1

$P(F \& N) = 0.1$
$P(M \& Y) = 0.42$

- Represents the probab. of an intersec^n of ≥ 2 events
- Marginal probab.

$P(F) = 0.10 + 0.34 = 0.44$
$P(Y) = 0.34 + 0.42 = 0.76$

- it is the probab. of obs. an outcome with a single attribute, regardless of other attributes
- Conditional probab.
  - among female buyers → owns a phone ($34/44$)
  - among who don't have phone → who're male ($\frac{14}{24}$)
  - we restrict the sample space

$P(N|F) = \frac{10}{44} = \frac{P(F \cap N)}{P(F)}$

$P(M|N) = \frac{14}{24}$

L8.2 Conditional Probab. Formula

- in determining probab. when some partial info. concerning the outcome of expt. is available

Eg. Roll a dice twice

$$S = \begin{Bmatrix} (1,1)\ (1,2) \ldots \\ (4,2) \ldots \\ 6,6 \end{Bmatrix}$$

6 - sided

$\{S\} = \{1, 2, 3, 4, 5, 6\}$

36 outcomes

→ each outcome is equally likely to occur

~~~ ... of 1/36~~

→ further, first dice ~~shows~~ a 4 ...

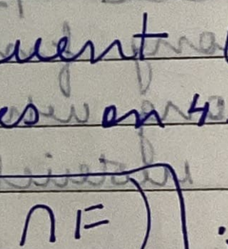
↳ what is probab. that sum of dice is 10

$P(d_1 + d_2 = 10 \mid E) = $

- if each outcomes of S are equally likely then the subset of, those are also equally likely
- Let E be conditioning event $(d_1 + d_2 = 10)$ and F be event first dice was 4.

$$\boxed{P(E \mid F) = \frac{P(E \cap F)}{P(F)}} \; ; \; P(F) > 0$$

probab of E given F



$\frac{1}{36}$

$P(E) = \{(4,6)\ (5,5)\ (6,4)\}$

$P(F) = \frac{6}{36}$

$P(F) = \{(4,1)\ (4,2) \ldots (4,6)\}$

$$P(E \mid F) = \frac{1}{6}$$
~~~

L8.3 Multiplicaⁿ Rule

$$\boxed{P(E|F) \times P(F) = P(E \cap F)}$$

- probab. of E & F occurs is equal to probab. of F occurs multiplied by conditional probab.
- useful to compute probab. of intersecⁿ

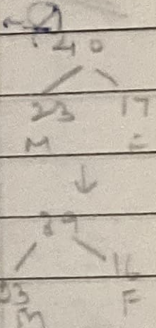
Eg. $T = 40$, $M = 23$, $F = 17$, 2 students are selected. sampling is w/o replacement. Find probab. that 1st is female & 2nd is M

$S = \{M_1M_2,\ M_1F_2,\ F_1M_2,\ F_1F_2\} \Rightarrow E = F_1M_2$

$P(F_1 \cap M_2) = P(F_1) \times P(M_2|F_1)$

$= \frac{17}{40} \times \frac{23}{39}$

$= 0.254$



- generalized multiplicaⁿ rule

$$P(E_1 \cap E_2 \cap \ldots \cap E_n) = P(E_1)\, P(E_2|E_1)\, P(E_3|E_1 \cap E_2) \times P(E_n| E_1 \cap E_2 \ldots \cap E_{n-1})$$

Eg. 52 → 4 piles of 13 cards. Probab. that each pile has exactly 1 ace.

1	2	3	4
13	13	13	13

$E_1 = \{$ace (spade) in pile 1$\}$

$E_2 = \{$ace " + heart in 2 piles$\}$

$P(E_1 \cap E_2 \cap E_3 \cap E_4) = P(E_1) \times P(E_2|E_1) \times P(E_3|E_1 \cap E_2)$

$P(E_1) = 1 \qquad P(E_2|E_1) = \frac{P(E_1 \cap E_2)}{P(E_1)} = \frac{39}{51}\left(1 - \frac{12}{51}\right)$

$P(E_3 | E_1 \cap E_2) = \left(\frac{12}{50} + \frac{12}{50}\right) = \frac{24}{50} \Rightarrow 1 - \frac{24}{50} = \frac{26}{50}$

$P(E_4 | E_1 \cap E_2 \cap E_3) = \frac{12}{49} + \frac{12}{49} + \frac{12}{49} = \frac{36}{49} = 1 - \frac{36}{49} = \frac{13}{49}$

L8.4 Independent Events

Q. Will the cond. probab. that E occurs given F has occured equal to unconditional probab. of E?

$P(E|F) = P(E)$?

↳ will knowing F has occured gen. Δ the chances of E's occurence?

Coss eg. $P(E) = P(E|F) = \frac{1}{2}$

we say that E is independent of F.

- $P(E \cap F) = P(F) \times P(E|F)$

we see that E is independant of F if

$$\boxed{P(E \cap F) = P(F) \times P(E)}$$

- 2 events that are not independant is c/d dependant events

If, $P(E \cap F) = P(E) \times P(F)$

then E & F are independent events

CLASSTIME Pg. No.
Date / /

18.5 Independent Events → Egs

Eg. rolling dice twice, equally likely

E_1 → first outcome is 3

E_2 → sum of outcomes is 8 (dep on 1st throw)

E_3 → " " " " 7 (indep of 1st throw)

are E_1 & E_2 independent (Not) & E_1 & E_3 indep. (yes)?

E_2:

1st	2nd
2	6
3	5
4	4
5	3
6	2

E_3:

1st	2nd
1	6
2	5
3	4
4	3
5	2
6	1

$P(E_1 \cap E_2) = P(\{3,5\}) = \frac{1}{36}$ $\quad P(E_1) = \frac{1}{6}$ $\quad P(E_2) = \frac{5}{36}$

$\frac{1}{36} \neq \frac{1}{6} \times \frac{5}{36}$

$P(E_1 \cap E_3) = P(\{3,4\}) = \frac{1}{36} = \frac{1}{6} \cdot \frac{1}{6}$

Eg. 1 card from 52 $\quad E_1$ = face card

E_2 = king $\quad E_3$ = heart

are E_1 & E_2 are indep.

" E_2 & E_3 are indep.

$E_1 = \frac{12}{52}$ $\quad E_2 = \frac{4}{52}$ $\quad E_3 = \frac{13}{52}$

$E_1 \cap E_2 = \frac{4}{52}$ $\quad E_2 \cap E_3 = \frac{1}{52}$

Date / /

L8.6 Independent Events → Properties

- If E & F are independent, then so are E and F^c

$$\boxed{P(E \cap F^c) = P(E) \times P(F^c)}$$

Eg. roll a dice twice. sum of 7 is independent of first & second throw also

E = sum of die 7 $\frac{1}{6}$

F = first dice is 4 $\frac{1}{6}$

G = second die is 3 $1/6$

$F \cap G = \frac{1}{36}$ $P(E / F \cap G) = 1$

event E is not independent of $(F \cap G)$

- 3 events E, F & G are said to be indep. if

1. $P(E \cap F \cap G) = P(E) \times P(F) \times P(G)$
2. $P(E \cap F) = P(E) \times P(F)$
3. $P(E \cap G) = P(E) \times P(G)$
4. $P(F \cap G) = P(F) \times P(G)$

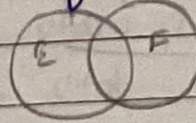
Eg. 3 children, find probab. all 3 are girls

E_1 = 1 girl → 1st
E_2 = 1 girl → 2nd
E_3 = 1 girl → 3rd

$\frac{1}{2}$ $P(E_1 \cap E_2 \cap E_3) = \frac{1}{8}$

L8.7 Bayes' Rule

- Law of total probab.



$E = (E \cap F) \cup (E \cap F^c)$

$\rightarrow P(E) = P(E \cap F) + P(E \cap F^c)$

$= P(E|F)P(F) + P(E|F^c)P(F^c)$

- in order for an outcome to be in E, it must be either in both E & F or be in E but not in F
- events $F_1, F_2, \ldots F_k$ are mutually exclusive
Then, for any event E

$$E = (E \cap F_1) \cup (E \cap F_2) \cup \ldots \cup (E \cap F_k)$$

$$P(E) = P(E|F_1)P(F_1) + \ldots + P(E|F_k)P(F_k)$$

$$\boxed{P(E_i) = \sum_{i=1}^{k} P(E|F_i)P(F_i)}$$

Eg. 2 → prone to accident & those who do not.
Accident prone → 0.1, the probab. for other 0.05. If probab. is 0.2 is accident prone. What is the probab. that a new policyholder will have accident?

$P(F) = 0.2$

$P(E|F) = 0.1 \qquad P(E|F^c) = 0.05 \qquad P(F^c) = 0.8$

$P(E) = 0.02 + 0.040 = 0.06$

- Bayes' Rule

$$P(F|E) = \frac{P(F \cap E)}{P(E)} = \frac{P(E|F)P(F)}{P(F)P(E|F) + P(F^c)P(E|F^c)}$$

CLASSTIME Pg. No.
Date / /

$$\boxed{P(E_i|E) = \frac{P(E|E_i)P(E_i)}{\sum_{i=1}^{k} P(E|E_i)P(E_i)}}$$

Eg. prev. ques → if new holder had the accident, probab. of accident-prone

$$P(E|E) = \frac{0.1 \times 0.2}{0.07} = \cancel{\frac{20}{63}} \quad 0.33$$

⇒ Tutorial 1

Eg. disk defective, manual detec^n → 0.70
that passes manual, electronic detec^n → 0.80
What % defective disk not detected?

| M | | E |

nde 0.3 × 0.2 = 0.06 ⇒ 6%

⇒ Tutorial - 2

Eg. A & B are independent. $P(A) = 0.8$ $P(B^c) = 0.4$

(i) $P(B)$ (ii) $P(A \cap B)$ (iii) $P(A \cup B)$

$1 - 0.4$
$= 0.6$

0.8×0.6
$= 0.48$

1.40
− 0.48
0.92

⇒ Tutorial -3

Eg. 54% women, 46% men | 68% of registed women
54 46 | 62% of " men voted
28.52

(i) women voted (registed) $\frac{68}{100} \times 54 = 36.72$

(ii) man not voted) 17.48

(iii) this person is a man given that this person voted in last elec^n

$$\frac{28.52}{28.52 + 36.72}$$

CLASSMATE Pg. No.
Date / /

→ Tutorial-4

Eg. A → ace card & B → spade. A & B are indep?
(i) a std. deck of 52 card Yes
(ii) a std. deck, w/o hearts 7s
(iii) a std. deck, 2–9 hearts removed. No

(i) $A = \frac{\cancel{4}}{\cancel{52}} = \frac{1}{13}$, $B = \frac{\cancel{13}}{\cancel{52}} = \frac{1}{4}$; $P(A \cap B) = \frac{1}{52}$

(ii) $P(A \cap B) = \frac{1}{39}$ | $\frac{\cancel{13}}{\cancel{39}} = \frac{1}{3}$; $\frac{\cancel{3}}{39} = \frac{1}{13}$

$52 - 13 = 39$. $\frac{3}{39}$

(iii) $A = \frac{\cancel{4}}{\cancel{44}} = \frac{1}{11}$ $B = \frac{13}{44}$ | $P(A \cap B) = \frac{1}{44}$

$52 - 8 = 44$

→ Tutorial-5

Eg. New battery over 10,000 miles – 0.8, 20000 – 0.4 & 30,000 – 0.1. New battery is working after 10,000
(i) total life will exceed 20,000 miles $P(B/A)$
(ii) add^{nal} life will exceed 20,000 miles $P(C/A)$

A → >10000
B → >20000
C → >30000

$P(B/A) = \frac{P(B \cap A)}{P(A)} = \frac{\cancel{0.4}}{\cancel{0.8}} = \frac{1}{2}$

$P(C/A) = \frac{0.1}{0.8} = \frac{1}{8}$

→ Tutorial-6

Eg. 2% of age 45 → have breast cancer, 90% of have +ve m
10% of women who do not have has +ve m. If women has +ve m, what is probab. breast

$\frac{\cancel{2}}{\cancel{100}} \times \frac{\cancel{90}}{} = 0.9$

A → BC
B → +ve m

$0.018 = 0.02 \times 0.9$
$0.098 = 0.98 \times 0.1$

$\frac{P(A \cap B)}{P(B)}$

$= \frac{0.018}{0.018 + 0.098} = \frac{0.018}{0.116} = 0.1552$

Week-9

L9.1 Random Var. - Introducⁿ

- When expt. is performed, sometimes, we are interested in the value of some numerical quantity determined by the result
- Eg. rolling twice, we care abt. sum of outcome and not indi. values
- Random Var - these quantities of interest, these real-valued funcⁿ defined on sample space.
- we may probab. to the possible values of random var.

Eg. roll dice twice ⟹ S = 36 outcomes

1. how many outcome will give sum 7? 1/6
2. how many outcome will have smaller of outcome as 3? (Eg. (1,3) ⇒ 1) 7/36

S 1. X ⇒ sum of 2 rolls   Y ⇒ denotes the lesser of 2 outcome

X ⇒ 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12

Value of X	Relevant Event
2	{(1,1)}
3	{(1,2)(2,1)}
⋮	⋮
12	{(6,6)}

Ans $= \frac{1}{6}$

$P(X=2)$ $P\{(1,1)\}$ $\frac{1}{36}$

$P(X=3)$ $P\{(1,2)(2,1)\}$ $\frac{1}{18}$

Y ⇒ lesser of 2 outcome

Y ⇒ 1, 2, 3, 4, 5, 6

Value of Y	Relevant Event
1	{(1,1)(1,2) .. (6,1)}

P ⇒ $P(Y=1)$ P(")   11/36

Ans = $\frac{7}{36}$

Eg. toss coin 3 times ⟹ S = 8 outcomes

1. of the 3 tosses, how many tosses will be head
2. " " " ", which toss results in a heads first

X ⟹ no of heads that appear   Y = toss in which heads appear first

1.

Outcome	X	Y
HHH	3	1
HHT	2	1
⋮		⋮

X = 0, 1, 2, 3

X	Relevant Event
0	{TTT}
1	{(HTT), (THT), (TTH)}

$P(X=0)$ $P(\{(TTT)\})$ 1/8

$P(X=1)$ $P(\{(HTT),(THT),(TTH)\})$ 3/8

2.   Y R   Relevant Event
1   {(HHH) (HHT) (HTH) (HTT)}

$P(Y=1)$ $P(\text{Relevant})$ = $\frac{4}{8} = \frac{1}{2}$

L9.2 Applicaⁿ

Eg. 2 elder clients, pays ₹1 lakh upon death

A → younger one dies in following year

B → elder one dies in " "

2	= Both dies	→ A + B	$A \cap B$
1	= younger	→ A	$A \cap B^c$
1	= elder	→ B	$A^c \cap B$
0	= not dies	→ X	$A^c \cap B^c$

⟹ Both are independent

$P(A) = 0.05$

$P(B) = 0.10$

CLASSTIME | Pg No.
Date / /

$X \to$ total amt. paid to beneficiaries for death

$X = 0, 1, 2$ $P(1/4)\,0$ | $(A^c \cap B^c)$
$P(2/4)\,1$ | $(A^c \cap B)\ (A \cap B^c)$
$P(1/4)\,2$ | $(A \cap B)$

$P(A) = 0.05$ $P(A^c) = 0.95$
$P(B) = 0.1$ $P(B^c) = 0.9$

$P(X=0) = P(A^c \cap B^c) = 0.855$
$P(X=1) = (0.05 \times 0.9) + (0.95 \times 0.1)$
$= 0.045 + 0.095 = 0.140$
$P(X=2) = 0.005$

- Discrete RV - that takes at most a countable no. of possible values
  ↳ only a finite no. / countably ∞ no. of diff. values
- Continuous RV - outcomes of random event are numerical, but can't count & are infinitely divisible

L9.3 Discrete & conti. RV
- discrete RV - is one that has possible values that are discrete pts. along the $\mathbb{R}$ no. line
  ↳ counting
- conti. RV - that form an interval along the $\mathbb{R}$ no. line
  ↳ measurement

Eg. 4 floors in apartment, each floor has 3 apartment – 1, 2 & 3 bedroom apartment, we have the data

Ap. No.	Floor No.	No. of Bed.	Size of Ap. (sq. ft.)	Dist. from lift
1	1	1	900.23	503.5
2	1	2	1175.34	325.6
3	1	3	1765.15	480.88

- Randomly selecting 1 apt. in complex of 12 apt.
  $S = \{1, 2, 3, \ldots, 12\}$
- Let RV be no. of bedrooms, what can be the values? $X = 1, 2, 3 \Rightarrow$ Discrete RV
- Let RV be no. of floors, what can be the values? $Y = 1, 2, 3, 4 \Rightarrow$ Discrete RV
- Let RV be the size of apt., what can be the possible values? $\Rightarrow$ Continuous RV
  No discrete pts possible $[900, 1800]$ sq. ft
- Let RV be the dist. of apt. from lift, what is the value possible? $[324, 505]$ m
  $\Rightarrow$ Continuous RV

- Eg. of discrete – no. of people, no. of spelling mistakes, no. of accidents, etc.
- Eg. of Conti. → Temp., Ht., speed, time taken by a person

L9.4 Probab. Mass Func – Prop.

- Let X be discrete RV & suppose it has n possible values → $x_1, x_2, x_3, \ldots, x_n$

$$\boxed{p_{x_i} = P(X = x_i)}$$

Probability mass func^n $f(x)$ of X

$$X \quad x_1 \quad x_2 \quad x_3$$
$$p(X = x_i) \quad p(x_1) \quad p(x_2) \quad p(x_3)$$

- Prop. of p.m.f:

1. $p(x_i) \geq 0, \; i = 1, 2, \ldots$
2. $p(x) = 0$ for all other values of x
3. Since, X must take one of the values x_i

$$\boxed{\sum_{i=1}^{\infty} p(x_i) = 1}$$

$\hookrightarrow$ p_i also

Eg. X is RV has 3 values 0, 1, 2 $p(0) = 1/4$
$p(1) = 1/2$, $p(2) = 1/4$

X	0	1	2
$P(X = x_i)$	$1/4$	$1/2$	$1/4$

Eg. $X \Rightarrow 1, 2, 3, 4, 5$, which are p.m.f.

1. X 1 2 3 4 5
   $P \Rightarrow$ 0.4 0.1 0.2 0.1 0.3 ✗ 1 cond^n ✓ 2 cond^n ✗
2. $P \Rightarrow$ 0.2 0.3 0.4 0.1 -0.2 ✗ 1 cond^n ✗
3. $P \Rightarrow$ 0.3 0.1 0.2 0.4 0.0 ✓ 1st & 2nd cond^n ✓

Eg. $X \to 0, 1, 2 \ldots$ with probab.

$$p(i) = \frac{c\lambda^i}{i!} \quad \text{for some +ve } \lambda$$

what value of c makes this p.m.f

$p(x_i) \geq 0$

$$e^x = \sum_{i=0}^{\infty} \frac{x^i}{i!}$$

$$\sum_{i=0}^{\infty} \frac{c\lambda^i}{i!} = 1 \Rightarrow c \sum_{i=0}^{\infty} \frac{\lambda^i}{i!} \Rightarrow c e^{\lambda} = 1$$

$$\boxed{c = e^{-\lambda}}$$

CLASSTIME Pg. No.
Date / /

Eg. roll a dice twice, X is a sum of outcomes (2, 3, ..., 12)
for $X = 2$; $p(X=2) = 1/36$ ...... & so on.
Yes, it is p.m.f. (check both cond^{ns})

Y is lesser of the 2 values

Y	1	2	3	4	5	6	
$P(Y_i)$	$\frac{11}{36}$	$\frac{9}{36}$	$\frac{7}{36}$	$\frac{5}{36}$	$\frac{3}{36}$	$\frac{1}{36}$	$= \frac{36}{36}$

Yes, it is a p.m.f

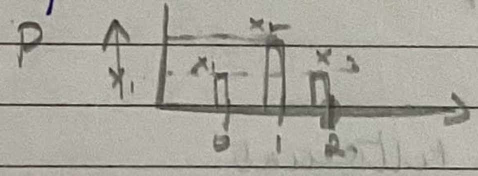
Eg. Tossing a coin 3-times
X ⇒ counts the no. of heads in tosses (0, 1, 2, 3)

X	0	1	2	3
p_i	1/8	3/8	3/8	1/8

Yes, it is p.m.f

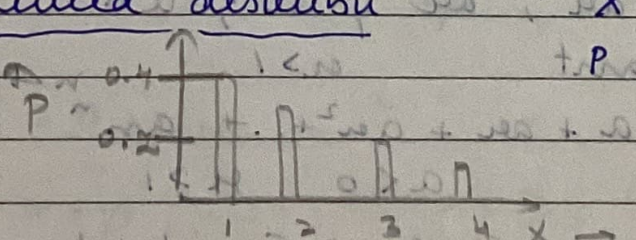
Y ⇒ the toss in which head appears first
Yes, it is also p.m.f

L9.5 Graph of p.m.f
- $P(X = x_i)$ on y-axis | x_i on the x-axis



Eg. +ve skewed distribution

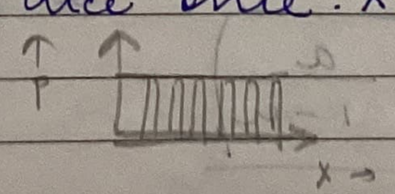
X	1	2	3	4
P	0.4	0.3	0.2	0.1



(opp. is -ve skewed distribution)

Eg. Rolling a dice once : X outcome

X	1	2	3	4	5	6
P	1/6					



degree of uniformity

Eg. Toss a coin $S = \{H, T\}$ $X\{H\} = 1$, $X\{T\} = 0$

X	0	1
P	0.5	0.5

L9.6 Cumulative distribuⁿ funcⁿ

$F : \mathbb{R} \to [0,1]$

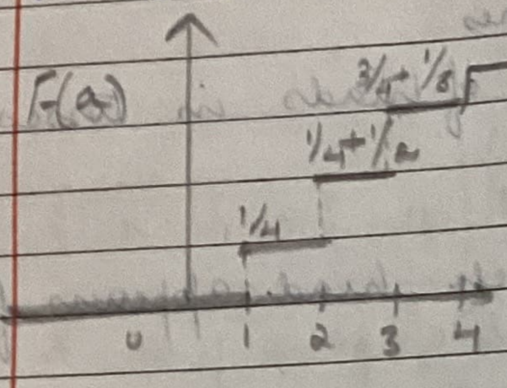
- cdf, $F \to$ $F(a) = P(X \leq a)$

Eg. $P(X \leq 1) = P(X=0) + P(X=1)$

Eg.

X	1	2	3	4
P	1/4	1/2	1/8	1/8

$$F(a) = \begin{cases} 0 & a < 1 \\ 1/4 & 1 \leq a < 2 \\ 3/4 & 2 \leq a < 3 \\ 7/8 & 3 \leq a < 4 \\ 1 & 4 \leq a \end{cases}$$



Step Funcⁿ

- Size of the step at any values 1, 2, 3 & 4 is equal to the probab. that X assumes at that particular values.

→ ∞ series tutorial

$a, ar, ar^2, ar^3, \ldots, ar^n$

$a \to$ constant $\quad r > 1$

$a + ar + ar^2 + \ldots + ar^n \xrightarrow{n \to \infty} +\infty$

$a < 0 \quad r > 1 \longrightarrow -\infty$

$r < 1 \quad r = 0.5 \quad a = 0.5 \quad \frac{1}{2} + \frac{1}{4} + \frac{1}{8} \ldots \to$ con. value

$r < 1 \to \boxed{= \frac{a}{1-r}}$

CLASSTIME Pg. No.
Date / /

$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots + \frac{x^n}{n!} \to \infty$$

$$\left.\begin{aligned} \frac{1}{1-x} &= 1 + x + x^2 + x^3 + x^4 \dots \infty \\ \frac{1}{1+x} &= 1 - x + x^2 - x^3 + x^4 - x^5 \dots \infty \end{aligned}\right\} |x| < 1$$

→ Tutorial - 1

Eg. ½ life U-235 → 25,000 yrs. Prob. of atom after k half lives ⇒ $f(k) = \frac{c}{2^k}$

1.1 find c for which f(k) is p.m.f

1.2 for c, find prob. after 10 half lives

1.1 $n \xrightarrow[25k\text{ yrs}]{} \frac{n}{2} \xrightarrow[50k\text{ yrs}]{} \frac{n}{4} \dots$ after k half lives alive atoms $= \frac{n}{2^k}$

Prob. of survival $= \frac{\frac{n}{2^k}}{n} = \frac{1}{2^k} = \frac{c}{2^k}$

⇒ c = 1

1.2 $\frac{1}{2^{10}}$

→ Tutorial - 2

Eg. p.m.f $P(X=x) = \alpha p^x$, $x = 0, 1, 2, 3 \dots$ find α

$\alpha p^x > 0 \ \forall x$

$\alpha > 0$ (x = 0)

$\alpha \geq 0$, $p \geq 0$ $\sum_{i=0}^{\infty} (P(X=x)) = 1$

$\alpha = 1 - p$ $\alpha + \alpha p + \alpha p^2 + \alpha p^3 \dots = 1$

$\alpha + p = 1$ $\alpha = 1$

⇒ Tutorial - 3

Eg. verify $f(x) = \frac{e^{-\lambda}\lambda^{x}}{x!}$ is pmf / not for $x = 0, 1, 2, 3..$

$f(x) \geq 0$

always +ve

$\frac{e^{-\lambda}\lambda^{x}}{x!} \geq 0$

$\sum_{x=0}^{\infty} f(x) = 1$

$= \sum_{x=0}^{\infty} \frac{e^{-\lambda}\lambda^{x}}{x!} = 1$

$\lambda^{x} > 0 \ \forall x$

$= e^{-\lambda}\left(1 + \frac{\lambda}{1!} + \frac{\lambda^2}{2!} \dots\right)$

Yes,

$= e^{-\lambda} \times e^{\lambda}$

$f(x) = \frac{e^{-\lambda}\lambda^{x}}{x!}$ is p.m.f

⇒ Tutorial - 4

Eg. X → no. of sand grains → discrete RV

Y → wt. of sand grains → conti. RV

⇒ Tutorial - 5

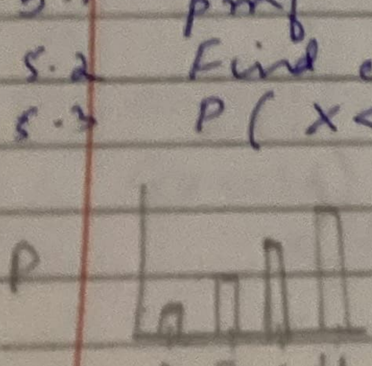
$$P(X = x) = \begin{cases} 0.1 & x = 0 \\ 0.2 & x = 1 \\ 0.4 & x = 2 \\ 0.3 & x = 3 \\ 0 & \text{others} \end{cases} \qquad P(Y) \begin{cases} 0.1 & Y = 2 \\ 0.2 & Y = 3 \\ 0.3 & Y = 6 \\ 0.4 & Y = 11 \end{cases}$$

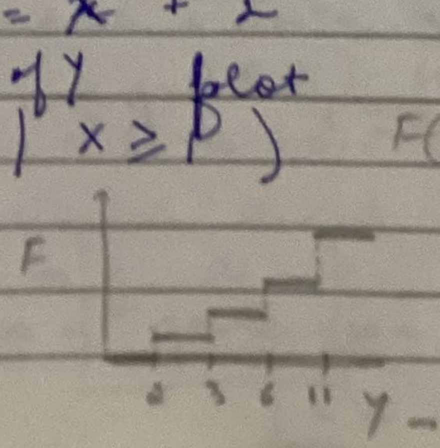
5.1 pmf $Y = X^2 + 2$

5.2 Find cdf of Y plot

5.3 $P(X < 3 \mid X \geq 1)$

$$F(Y) = \begin{cases} 0 & Y < 2 \\ 0.1 & 2 \leq Y < 3 \\ 0.3 & 3 \leq Y < 6 \\ 0.6 & 6 \leq Y < 11 \\ 1 & 11 \leq Y \end{cases}$$





$$P(X < 3 \mid X \geq 1) = \frac{0.2}{0.5} = \frac{2}{5} = \frac{4}{10} = 0.4$$

⇒ Tutorial - 6

cdf

$$F(X = x) = \begin{cases} 0 & x < 1 \\ 4/10 & 1 \leq x < 2 \\ 4/10 & 2 \leq x < 3 \\ 7/10 & 3 \leq x < 10 \\ 1 & x \geq 10 \end{cases}$$

$x = 1 \Rightarrow 4/10$ $\quad$ $x = 3 \Rightarrow 3/10$

$x = 2 \Rightarrow 0$ $\quad$ $x = 10 \Rightarrow 3/10$

X	1	2	3	10
$P(X = x)$	$\frac{4}{10}$	0	$\frac{3}{10}$	$\frac{3}{10}$

CLASSTIME Pg No.
Date / /

Week-10

L10.1 Discrete Random Var - Applicaⁿ

Eg. No. of credit cards owned by people
- Random Expt.: Selecting an adult at random
- Random Var: No. of credit cards owned by person

↳ Discrete random var.

X	0	1	2	3	4
$P(X=x_i)$	0.42	0.36	0.14	0.06	0.02

- The peak is at 0 and decreasing further.
  Skewed formation → Skewed Right
- If a person is chosen random. Is person like to have 0 card or 2 or more card

$P(0) = 0.42$ $P(\geq 2) = 0.14 + 0.06 + 0.02 = 0.22$

More likely to have 0 card than 2 or more credit card

- 500 people & how many card they own.
  Would you be surprised -
  (i) Everyone owns a card Yes
  (ii) 72% say that they have 2 cards No
  $0.14 \times 500 = 70$

L10.2 Expectaⁿ of a Random Var.

Eg. Rolling a fair dice once
Even outcome - u lose an amt. equal to n
Odd " - u win an amt. equal to n

1	2	3	4	5	6
+1	-2	+3	-4	+5	-6

CLASSTIME Pg No.
Date / /

- Would you play this game? If I accept to win this game.
- Roll the dice 100 times - - . Observe the outcome

1	+1	16	0.16
2	-2	10	0.10
3	+3	16	0.16
4	-4	21	0.21
5	+5	23	0.23
6	-6	14	0.14
		100	1

Avg. Winning $\bar{x} = -0.09$

- for 1000 times, average winnings = -0.451
- The r.f. of each of the six possible outcomes is close to probab. of 1/6 of respective outcome
- Hence, if I repeat large no. of times, our avg. gain is $\frac{1}{6} - \frac{2}{6} + \frac{3}{6} - \frac{4}{6} + \frac{5}{6} - \frac{6}{6} = -0.5$
- Expectaⁿ of discrete RV $\Rightarrow$ $\boxed{E(x) = \sum_{i=1}^{\infty} x_i P(X=x_i)}$

- It is a long-run avg. value of RV. in repeated independent observaⁿˢ

Eg. | X | 1 2 3 4 5 6 |
$P(X=x_i)$ | 1/6 each

$E(X) = \frac{1}{6}(21) = \frac{7}{2} = 3.5$

- If we die many no. of time, avg. of roll could be 3.5.
- Expected value of X is a theoretical avg.

Date / /

Eg. Rolling a dice twice
$X =$ sum of outcome

$$E(x) = \frac{2+6+12+20+30}{36} = \underset{\sim}{7}$$

- ⇒ Centre of mass expectaⁿ

Eg. Tossing coin thrice

X	0	1	2	3
P	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

$E(x) = \frac{3}{2} = \underset{\sim}{1.5}$

- A RV that take either the value 1/0 is c/d Bernoulli RV

X	0	1
P	$(1-p)$	p

$\boxed{E(x) = p}$

- Discrete Uniform RV

X	0	1	...	n
P	$\frac{1}{n}$	$\frac{1}{n}$	...	$\frac{1}{n}$

$E(X) = \frac{1}{n}\left(\frac{n \times (n+1)}{2}\right)$

$\boxed{E(x) = \frac{n+1}{2}}$

L10.3 Properties of Expectaⁿ

- Let g be any real values funcⁿ.

$$\boxed{E(g(X)) = \sum_i g(x_i) P(X = x_i)}$$

- if a & b are constants, $\boxed{E(aX+b) = aE(X)+b}$

Eg. | X | -1 | 0 | 1 |
|---|---|---|---|
| P | 0.2 | 0.5 | 0.3 |

$Y = g(X) = X^2$

What is $E(Y)$

X^2	1	0
P	0.5	0.5

$E(Y) = 0.2 + 0.3$
$= 0.5$

CAUTION $\Rightarrow E(X^2) \neq (E(X))^2$

Eg. Anitha's RV whose $E(X) = 15{,}000$. Sanjay's Bonus $= 90\%$ of Anita, find $E(Y)$

$Y = 0.9X$

$E(Y) = 0.9 \times E(X)$
$= 13{,}500$

If Sanjay's Bonus is ₹1000 more than Anita's
$Y = X + 1000$
$E(Y) = E(X) + 1000$
$=$ ₹$16{,}000$

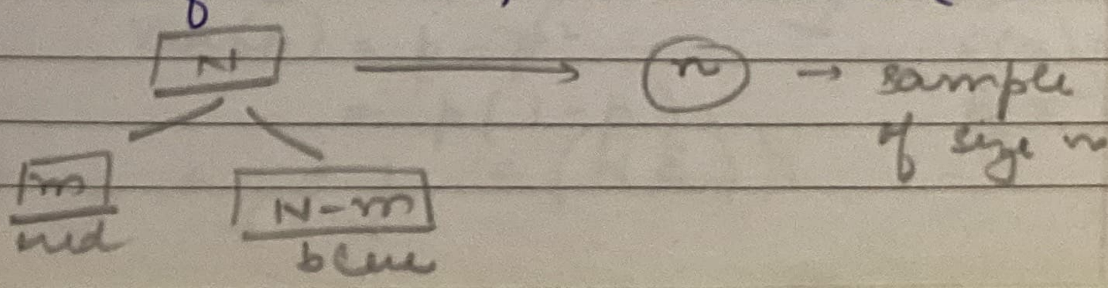
- $\boxed{E(X+Y) = E(X) + E(Y)}$

Eg. Rolling a dice. $X =$ outcome of fair dice
$Y =$ " " another dice

$E(X) = E(Y) = 3.5$

$E(X+Y) = 7$

- Hyper Geometric RV - suppose a size n from a box of N balls, m is red & $(N-m)$ blue



CLASSTIME Pg No.
Date / /

$$P(X=i) = \frac{\binom{m}{i}\binom{N-m}{n-i}}{\binom{N}{n}}, \quad i = 0, 1, 2 .. n$$

$$\boxed{E(X) = \frac{nm}{N}}$$

X is said to be a hypergeom.. var. for some values of n, m & N

Eg. 2 students from 20 boys & 10 girls.
Let X is no. of boys & Y is no. of girls

$E(X) = \frac{nm}{N} = \frac{2 \times 2\cancel{0}}{3\cancel{0}} = \frac{4}{3}$

$E(Y) = \frac{2 \times 1\cancel{0}}{3\cancel{0}} = \frac{2}{3}$

$E(X+Y) = \frac{4}{3} + \frac{2}{3} = \frac{\cancel{6}}{\cancel{3}} = 2$

- Expecta^n of sum of many RVs

$$E\left(\sum_{i=1}^{k} X_i\right) = \sum_{i=1}^{k} E(X_i)$$

Eg. Tossing a coin 3 times

X = no. of heads
$E(X) = \frac{3}{2}$

	Toss 1		Toss 2		Toss 3	
	X_1		X_2		X_3	
	H	T	H	T	H	T
$P(x)$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$
$E(X)$	$\frac{1}{2}$		$\frac{1}{2}$		$\frac{1}{2}$	

$$P(3H) = \frac{1}{2} + \frac{1}{2} + \frac{1}{2} = \frac{3}{2}$$

L.10.4 Variance of a RV

- The expected value of RV gives weighted avg. of possible values, it doesn't tell about the variation or spread of these values.

X	= 0	P(1)	E(X) = 0
Y	-2	P(1/2)	E(Y) = 0
	2	P(1/2)	
Z	-20	P(1/2)	E(Z) = 0
	20	P(1/2)	

- $E(X) = \mu$ X is a RV

$$\boxed{Var(X) = V(X) = E(X-\mu)^2}$$

- Computational formula for Var(X)

$$\boxed{Var(X) = E(X^2) - \mu^2}$$

Eg. Rolling a die once

X	1	2	3	4	5	6
X^2	1	4	9	16	25	36
P(X)	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

$E(X) = \mu = 3.5$
$E(X^2) = 15.167$

$Var(X) = 15.167 - 3.5^2$
$= 2.917$

- Bernoulli RV — $0 \to (1-p)$
  $1 \to (p)$

X	0	1
X^2	0	1
P	(1-p)	p

$E(X) = p$
$E(X^2) = p$

$Var(X) = p - p^2$
$= p(1-p)$

- Discrete uniform RV

X	1	2	...	n
X^2	1	4	...	n^2
P	$\frac{1}{n}$	$\frac{1}{n}$	...	$\frac{1}{n}$

$E(X) = \frac{n+1}{2}$

$E(X^2) = \frac{(n+1)(2n+1)}{6}$

$$Var(X) = \frac{n^2-1}{12}$$

L10.5 Prop. of Variance

- $Var(cX) = c^2 Var(X)$
- $Var(X+c) = Var(X)$
- $Var(aX+b) = a^2 Var(X)$
- $Var(X+Y) \neq V(X) + V(Y)$
- RV X & Y are independent, if knowing the value of one of them doesn't Δ the probab. of the other.
- If X & Y are indep. RV,

$$Var(X+Y) = Var(X) + Var(Y)$$

Eg. Rolling a dice twice

$X \to$ first dice outcome

$Y \to$ 2nd dice outcome

$E(X) = E(Y) = 3.5$ $\qquad E(X+Y) = 7$

$V(X) = V(Y) = 2.917$ $\qquad V(X+Y) = 5.83$

- Hypergeom. R.V => $E(X) = \frac{nm}{N}$

CLASSTIME Pg. No.
Date / /

L10.6 Standard Deviation of RV

- $SD(X) = \sqrt{Var(X)}$
- The std. deviation has units same as the units of RV.
- $SD(cX) = c\, SD(X)$; c is any real no.
- $SD(c+X) = SD(X)$

Eg. $V(X) = 4$, $SD(3X) = ?$

$V(X) = 4 \Rightarrow SD(X) = 2 \Rightarrow SD(3X)$
$= 3 \times 2 = 6$

Eg. $V(2X+3) = 16$, $SD(X) = ?$

$4V(X) = 16$

$V(X) = 4$ $\quad SD(X) = 2$

Eg. Anitha RV whose $E(X) = 15{,}000$ & $SD(X) = 3{,}000$
Sanjay RV whose $E(Y) = 20{,}000$ & $SD(Y) = 4{,}000$
What is $E(X+Y)$ & $SD(X+Y)$

$E(X+Y) = 35{,}000$ $\quad V(X) = 90{,}00{,}000$
$V(Y) = 1{,}6\,0{,}0\,0{,}000$

$V(X+Y) = 2{,}50{,}00{,}000$
$SD(X+Y) = 5{,}000$

Eg. Fixed fee of ₹25,000 , or $\quad$ win ₹50,000 $\frac{1}{2}$
loose 0 $\frac{1}{2}$

$E(X) = 25{,}000$ $\quad E(Y) = 25{,}000$
$SD(X) = 0$ ✓ $\quad SD(Y) = 25{,}000$

Date / /

⇒ Expectaⁿ of Hypergeometric RV

Eg. n balls from a box of N balls of which m are red and (N − m) blue.

$X = \{0, 1, 2, \ldots, m\}$

$$\boxed{P(X = i) = \frac{{}^{m}C_i \times {}^{N-m}C_{n-i}}{{}^{N}C_n}}$$

$$\boxed{E(X) = \frac{nm}{N}}$$

Week-11

CLASSTIME Pg. No.
Date / /

L11.1 Binomial distribution - Bernoulli distribution

- A trial on an expt. whose outcome can be classified as either success / failure → Bernoulli Trail
- $S = \{Success, Failure\} \mid \{0, 1\}$

Eg. - Tossing a coin → Head = Success
- Rolling a die → 6 = Success
- Opinion Poll → Yes = Success
- Salesperson selling → Sale = Success
- Drug effectiveness → Effective = Success

- not bernoulli trials → where outcomes are anything other 2.
- Bernoulli RV → X → 1 → p, $E(X) = p$; 0 → $1-p$

$V(x) = p(1-p)$

- $V(x) = p - p^2$ ⇒ quadratic equation
- What value of p maximises the $V(x)$?

$$f(x) = p - p^2 \Rightarrow f'(x) = 1 - 2p = 0$$
$$\boxed{p = \frac{1}{2}}$$

- The largest variance occurs when $p = 1/2$, when success & failure are equally likely.

Eg. Toss of fair coin

L11.2 Independent Bernoulli Trials

- $n = 3$, independent bernoulli trials
p is the probab. of success

$(s, s, s) \rightarrow p \times p \times p$
$(s, s, f) \rightarrow p \times p \times (1-p)$

- 1, 2, 3, ... n

  S S S ... S $\rightarrow p$

  F F F ... F $\rightarrow (1-p)$

- The probab. of i successes in n trials is,

$$\boxed{P(X=i) = {}^{n}C_i \times p^i \times (1-p)^{(n-i)}}$$

$i = 0, 1, 2, \ldots, n$.

41.3 Distribuⁿ of Binomial RV

- X is a binomial R.V. with parameters n & p, that expresents the no. of success in n independent bernoulli trials, when each trial is a success with probab. p.

$$P(X=i) = {}^{n}C_i \times p^i \times (1-p)^{(n-i)}$$

Eg. Tossing coin three times

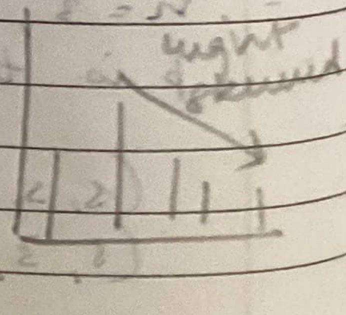
X	0	1	2	3
	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

$P(X=0) = {}^{3}C_0 \times 0.5^0 \times (0.5$

- a binomial distribuⁿ is
  - right skewed if $p < 0.5$
  - sym. if $p = 0.5$
  - left skewed if $p > 0.5$

Eg. $n = 4$, $p = 0.3$

$P(X=0) = 1 \times 1 \times (0.3)^4 = 0.241$



CLASSTIME Pg. No.
Date / /

- when n increases → sort of symmetricity but large skewness towards right / left
↪ approaches symmetric behaviour

L11.4 Modelling Situations as Binomial Distribuⁿ

Eg. Pack of 3 goods ⟶ 10% of goods are defective. The customers won't complain if 1 out of 3 are of bad quality. The company wants to keep the no. of complaints low → at 3%. How to analyze the situaⁿ?

defective = 0.1 $P(X=0) = 1 \times 1 \times (0.1)^3$
good = 0.9 $= 0.001$
$P(X=1) = 0.027$

$P(X \leq 1) = 0.028 \sim 2.8\%$

⟶ Company keeps the customers happy

Eg. Rolling 4 dice. 5 → 6
(a) 6 appears at least once
(b) 6 appears exactly once $P(X=1)$
(c) 6 appears at least twice $P(X = 2, 3, 4)$

$X \sim BN(4, \frac{1}{6})$ X = no. of 6s in 4 rolls

$P(X=0) = {}^nC_i \times p^i \times (1-p)^{n-i} = 0.4823$
$P(X=1) = 0.3858$ $P(X=3) = 0.0154$
$P(X=2) = 0.1157$ $P(X=4) = 0.0008$

Google Sheet

set up parameters $n = 4$, $p = 0.5$

	0	1	2	3	4
= binomial dist. probab. (x, n, p, 0)	0.0625	0.25	0.375	0.25	0.0625

= BINOM.DIST

CLASSTIME Pg No.
Date / /

Eg. defective = 0.05, sample = 5

(a) none are defective ($P(X=0)$)

X	1	0
	0.05	0.95

(b) 2 or more are defective

($P(X=2+3+4+5)$)

$X \sim B(5, 0.05)$

X	0	1	2	3	4	5
	0.7738					

$p(x=0) = 1 \times 1 \times (0.95)^5$

Eg. $n = 4$ comp., func^n if at least 2 are working

$p = 0.8$, work

X = no. of comp. working

X	0	1	2	3	4
	0.0016	0.0256	0.153	0.409	0.4096

To system to work $p(X = 2+3+4)$

Eg. 4 possible ans. for each 5 ques. What is prob. that student will get 4 or more correct ans.

X = correct respons $X \sim B\left(5, \frac{1}{4}\right)$

X	X
4	5

$p(X=4) = {}^nC_i \times p^i \times (1-p)^{n-i}$

$= {}^5C_4 \times (0.25)^4$

$= 5 \times (0.25)^4 \times (0.75)^1 = 0.0146$

$= P(X=5)$ $= 0.0010$.

$\frac{5!}{1!4!}$

0.0156

CLASSTIME Pg. No.
Date / /

11.5 Expecta^n & Var of Binomial distribution

$X \sim Bin(n, p) \rightarrow n$ independent trials

$X = X_1 + X_2 + \dots + X_n$ $\quad X_i = 1$, if success

$E(X_i) = p$

$E(X) = E(X_1 + X_2 + \dots + X_n)$

$= \underbrace{p + p \dots + p}_{n \text{ times}}$

$$\boxed{E(X) = np}$$

$V(X_i) = p(1-p)$

$V(X) = V(X_1 + X_2 + \dots X_n)$

$= V(X_1) + V(X_2) \dots + V(X_n)$

$= [p(1-p)] + [p(1-p)] \dots$

$$\boxed{V(X) = np(1-p)}$$

Eg. Tossing a coin 500 times $n = 500$, $p = 0.5$

What is std. devia^n of no. of times head appears?

$V(X) = 500 \times 0.5 \times 0.5 = 125$

$E(X) = 500 \times \frac{1}{2} = 250$

$SD(X) = \sqrt{125} = 5\sqrt{5} = 11.1803$

Eg. $n = 10$, $E(X) = 6$, p of 8 heads

$E(X) = np$ $\quad p = \frac{6}{10} = 0.6$

$P(X = 8) = {}^{10}C_8 \times (0.6)^8 \times (0.4)^2$

$= 45 \times 0.16 \times (0.6)^8$

$= 0.121$

CLASSTIME Pg. No.
Date / /

Eg. $E(X) = np = A$
$V(X) = np(1-p) = B$

$\boxed{p = 1 - \frac{B}{A}}$ $\boxed{n = \frac{A}{\left(1 - \frac{B}{A}\right)}}$

$E(X) = 4.5$, Var. $= 0.45$

$\frac{\cancel{45}}{10\cancel{2} \times \cancel{45}} \times 1\cancel{2} = 0.9 = p$ $n = 5$

$P(X = 3) = 10 \times (0.9)^3 (0.1)^2$

$= 0.0729$

→ Tutorial -1

Eg. 10% are left handed. Find in every 20.
→ how many are exactly 3 left handers

$p = 0.1$
$n = 20$

$P(X=3) = {}^{20}C_3 \times (0.1)^3 \times (0.9)^{17}$

$= ✓$

⇒ Tutorial -2

Eg. 80% disapprove abt. attendance. $n = 10$ find p. that no. disapprove is atmost 7

$p = 0.8$
$n = 10$

$P(X \leq 7) = 0 + 1 + 2 + 3 + 4 + 5 + 6 + 7$

or $1 - (X = 8 + 9 + 10)$

is same

→ Tutorial - 3

Eg. $E(x) = 9$, $V(x) = 4.95$. Find $P(x = 12)$

$np = 9$

$\cancel{np}(1-p) = 4.95$ $\qquad P(x = 12)$, $n = 20$

$1 - p = \frac{4.95}{9}$ $\qquad = {}^{20}C_{12} \times (0.45)^{12} \times (0.55)^{8}$

$p = \frac{9 - 4.95}{9} = \frac{4.15}{9}$

$= 0.45$

→ Tutorial - ~~8~~4

Eg. will qualify is 0.4. out of 10, no one will qualify

$p = 0.4$

$n = 10$ $\qquad P(x = 0) = {}^{10}C_0 \times (0.4)^0 \times (0.6)^{10}$

→ Tutorial - 5

Eg. 10 comp. if > 4 out of 10 can work, $p = 0$

$n = 10$ $\qquad P(x \geq 4)$

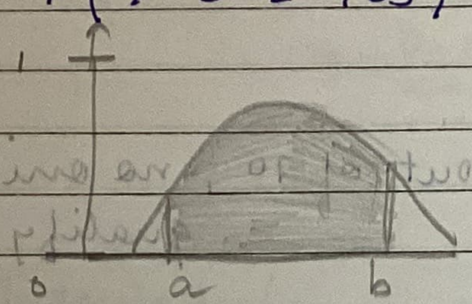
$p = 0.7$ $\qquad 1 - P(x = 0) + 1 + 2 + 3$

Week-12

L12.1 Introductⁿ to Continous RV

- when outcomes for random event, but can't be counted and are infinitely div., we have continous RV. It has ~~entervta~~ interval along the real line.
- every contd. RV X has a curve associated with it. → Probab. Density funcⁿ ($f(x)$)
- The probab. the X assumes a values btw a & b = equal to curve btw. a & b

$$P(X \in [a,b]) = P(a \leq X \leq b)$$
$$= \boxed{\int_a^b f(x)\,dx}$$



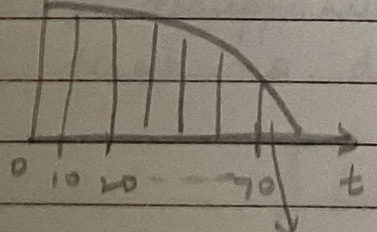
① The area under pdf btw. any 2 pts. is btw. 0 & 1

② Total under the pdf is always 1.

$P(a \leq X \leq b) = P(a < X < b)$
↳ remains same whether endpts. are included in it or not.

Eg. prob. less than 20 mins → 0.29

less than 40 min. → $0.29 + 0.27 = 0.56$



$0.29 + 0.27 + 0.29 = 0.83$

0.17

- cumulative distribuⁿ funcⁿ

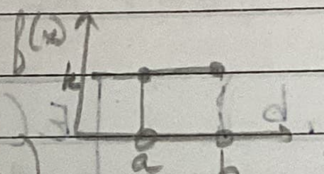
$$F(a) = P(X \le a) = \int_{-\infty}^{a} f(x)\,dx = P(X<a)$$

- the probab. that a conti. RV X assumes a single values as 0.
- $E(X) = \int x\, f(x)\, dx$
- $V(X) = \int (x - E(x))^2 f(x)\, dx$

L12.2 Uniform distribuⁿ

- $X \sim U(a,b)$

$f(x) = k \quad (a \le x \le b)$



$\int_a^b f(x) = 1$

$U[a,b] = U(a,b)$

$$1 = \int_a^b f(x)\,dx = k(b-a) = 1 \Rightarrow \boxed{k = \frac{1}{b-a}}$$

$$f(x) = \begin{cases} \frac{1}{b-a}, & a < x < b \\ 0, & \text{otherwise} \end{cases}$$

Eg. $U(2,4)$ $\quad k = \frac{1}{4-2} = \frac{1}{2}$

Eg. $U(-2,2)$ $\quad k = \frac{1}{2-(-2)} = \frac{1}{4} = 0.25$

- Std. Uniform distribuⁿ with min 0 & max is 1

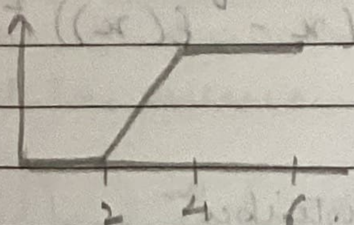
$$f(x) = \begin{cases} 1 & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

→ This plays an impt. role in random variate generation.

- For $X \sim U(a,b)$

$$F(x) = \begin{cases} 0 & \text{for } x < a \\ \dfrac{x-a}{b-a} & \text{for } x \in [a,b] \\ 1 & \text{for } x \geq b \end{cases}$$

Eg. cdf for $X \sim U(2,4)$



- $X \sim U(a,b)$ $\boxed{E(x) = \dfrac{a+b}{2}}$

$\boxed{V(x) = \dfrac{(b-a)^2}{12}}$

L12.3 Uniform Distributions Applica^{ns}

Eg. X is uniform RV over (0,1). Find

1. $P(X > \frac{1}{3})$ $\quad P(X > \frac{1}{3}) = 1 - P(X \leq \frac{1}{3})$

$F(x) = x \quad = F(\frac{1}{3}) = \frac{1}{3}$

$\frac{2}{3}$

2. $P(X \leq 0.7) = 0.7$
3. $P(0.3 < X \leq 0.9) = 0.9 - 0.3 = 0.6$
4. $P(0.2 \leq X < 0.8) = 0.6$

Eg. meet at 2pm, btw 2 - 3 pm

X = amt. of wait

$0 \longrightarrow 60$

$X \sim U(0, 60)$

1. at least 30 min $(X \geq 30) = 1 - P(X \leq 30)$
$= 1 - \frac{30}{60} = 1 - \frac{1}{2} = \frac{1}{2}$

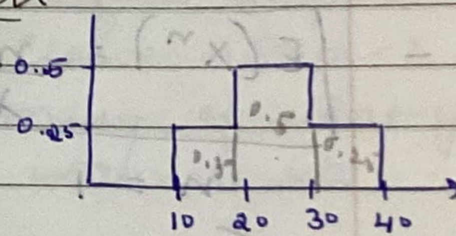
2. less than 15 min $(X \leq 15) = \frac{15}{60} = \frac{1}{4}$

3. btw. 10 & 35 min $(10 \leq X \leq 35) = \frac{35}{60} - \frac{10}{60} = \frac{25}{60} = \frac{5}{12}$
$= \frac{5}{12}$

4. less than 45 min $(X < 45) = \frac{45}{60} = \frac{3}{4}$

L12.4 Non Uniform & Δ Distribution

Eg. min. of playing time



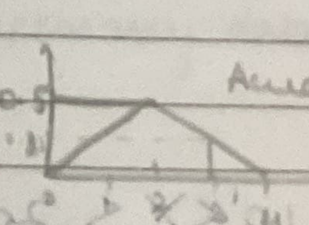
(i) over 20 min $P(X > 20)$
$1 - P(X \leq 20)$ $1 - 0.25$
0.75

(ii) ~~less than~~ 25 min $(X < 25)$ $0.25 + 0.25 = 0.5$

(iii) 15 → 35
0.125
0.5
0.125
0.75

(iv) $X > 35$ 0.125

Eg. 2pm → study until 6pm
X = time spend studying



Area of Δ = 1

(a) height of the curve at value 2

$\frac{1}{2} \times 4 \times h = 1$
$h = 0.5$

(b) probab. more than 3 hrs
0.125

CLASSTIME Pg No.
Date / /

(c) what will be prob. blue. $1 \to 3$

$0.5 + \frac{1}{2} \times 2 \times 0.25 = 0.75$

L12.4 Exponential Distribuⁿ

- a conti. RV, for some $\lambda > 0$; by

pdf $\rightarrow f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$

$X \sim Exp(\lambda)$

- $F(a) = 1 - e^{-\lambda a} \implies$ cdf
- $E(X) = \frac{1}{\lambda}$, $V(X) = \frac{1}{\lambda^2}$
- $\boxed{E(X^n) = \frac{n}{\lambda} E(X^{n-1})}$

$n = 1$

$\hookrightarrow \frac{1}{\lambda} = E(X^1)$

Thus, the mean of exponential is the reciprocal of its λ, & var is mean sq.

Eg. $\lambda = 0.1$, find the probab. u have to wait

(a) $X \geq 10$ min

$= 1 - P(X \leq 10)$

$= 1 - F(10)$

$= 1 - (1 - e^{-0.1 \times 10}) = 1 - (1 - e^{-1})$

$= e^{-1} = 0.368$

(b) $10 < X < 20$

$F(20) - F(10)$

$1 - e^{-2} - (1 - e^{-1})$

$= -e^{-2} + e^{-1} = 0.2325$