Recap → Covariance & Correlation

x, y → f_1, f_2 → Bivariate

Bivariate → Through $-1 \le r \le 1$

Pearson Corr & Spearman

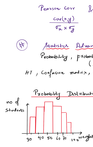
$\frac{cov(x,y)}{\sigma_x \sigma_y}$ $\frac{cov(r_x, r_y)}{\sigma_{r_x} \sigma_{r_y}}$

(18) Analysis Advanced

Probability, Probability Distribution fn. (pdf, cdf, pmf)

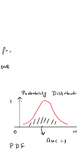
AI, Confusion matrix, error analysis, ROC

Probability Distribution Function



30-40 =
40-50
50-60
60-70

Probability Distribution



Area = 1

Histogram → Smoothening → PDF

★ Rolling a die (unbiased)

Possible outcomes	Probability
1	1/6
2	1/6
3	1/6
4	1/6
5	1/6
6	1/6

(9) another example

Pair distribution of getting sums of rolling two dice.

possible outcomes → {(1,1),(1,2),(1,3),(1,4),(1,5),(1,6), (2,1), (2,2), (2,3), (2,4), (2,5), (2,6), ... (6,6)}

possible sums	occurrence	Probability
2	1	0.03
3	2	0.06
4	3	0.08
5	4	0.11
6	5	0.14
7	6	0.17
8	5	0.13
9	4	0.11
10	3	0.08
11	2	0.06
12	1	0.03

no of outcomes = 36

Prob Dist fn

Probability mass function (pmf)

Discrete R.V

Probability Density fn. (pdf)

Continuous R.V

Ex ① Tossing a coin H, T

② Rolling a die 1, 2, 3, 4, 5, 6

1.75

Ex ① weight of students

② height " "

★ Pmf

Ex No of heads in an exp of tossing 2 coins

possible outcomes {TT, TH, HT, HH}

x	0	1	2
P(x)	1/4	2/4	1/4

P(x=0) P(x=1) P(x=2)

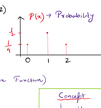
no of outcomes = 4

① → possible outcomes

② → x value

③ → " Prob Distribution

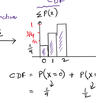
P(x) → Probability mass function



Σ

CDF

Σ F(x)



CDF (Cumulative Distribution Function)

↳ Σ pmf

Example

$\underset{0}{\uparrow}$ $\underset{1}{\uparrow}$ $\underset{2}{\uparrow}$

CDF(y) = P(x=0) + P(x=1) + . . . + P(x=y)

CDF(2) = Σ P(x) = 1

CDF = P(x=0) + P(x=1) + P(x=2)

$= \frac{1}{4} + \frac{1}{2} + \frac{1}{4}$

CDF(x=0) = P(x=0)

$= \frac{1}{4}$

CDF(x=1) = P(x=0) + P(x=1)

$= 3/4$

CDF(x=2) = P(x=0) + P(x=1) + P(x=2)

$= \frac{1}{4} + \frac{1}{2} + \frac{1}{4} = 1$

= CDF(x=1) + P(x=2)

Σ p = 1

↳ 0 = 1

pdf

↳ continuous R.V

P(x=a) ✗ discrete

↳ not valid for continuous R.V.

P(a)

$$P(a) = \int_{a}^{a+\Delta a} f(x)\,dx$$

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

pdf

x → continuous Random variable

→ $P(a) = \int_{-\infty}^{a} f(x)\,dx$

integration = Area

Σ area

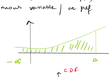
$$P(x) = \int_{\alpha}^{a} f(x)\,dx$$

10-20
20-30

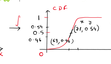
$$P(x) = \int_{30}^{40} f(x)\,dx$$

CDF (Cumulative Distribution fn)

↳ Continuous variable, or pmf



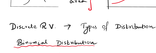
Normal Distribution



→ 'S' shaped Curve

Slope = $\frac{\Delta y}{\Delta x} = \frac{0.94 - 0.46}{70 - 40}$

$= \frac{0.48}{30} = 0.04$



(19) Discrete R.V → Types of Distribution

Binomial Distribution

Ex If coin is flipped 5 times, what is the no of heads from flipping

X = # of H from flipping 5 coins

possible outcomes = 2 × 2 × 2 × 2 × 2

$= 2^5 = 32$

$P(x=0) = \frac{1}{32}$

$P(x=1) = \frac{5}{32} = \frac{5C_1}{2^5}$

TTTTT

HTTTT
THTTT
TTHTT

$5C_1 = \frac{5!}{1!\,4!} = \frac{5 \times 4!}{4!} = 5$

H T T T T
- - - - -

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

$P(x=2) = \frac{5C_2}{2^5}$

HHTTT
THHTT
HTHTT

$5C_2 = 1$

$5C_1 = 5$

$5C_2 = \frac{5!}{2\times3!} = \frac{5\times4\times\cancel{3!}}{2\times\cancel{3!}} = 10$

$5C_3 = 5C_2$

$^nC_r = \frac{^nC_r}{n-r}$

$P(x=0) = \frac{5C_0}{2^5} = \frac{1}{32}$

$P(x=1) = \frac{5C_1}{2^5} = \frac{5}{32}$

$P(x=2) = \frac{5C_2}{2^5} = \frac{10}{32}$

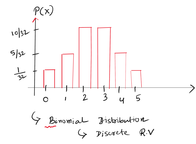
$P(x=3) = \frac{5C_3}{2^5} = \frac{10}{32}$

$P(x=4) = \frac{5C_4}{2^5} = \frac{5}{32}$

$P(x=5) = \frac{5C_5}{2^5} = \frac{1}{32}$

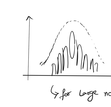
^nC_k

↳ Binomial coefficient



↳ Binomial Distribution

↳ Discrete R.V



↳ for large no. of flip