Textbook : Mathematical Statistics with applications in R

by Ramachandran & Tsokos

LEC-01

24/07/24

Focus of this course : Understanding basic concepts and developing intuition so that you can apply these to solve some problems in your area of research.

- Assignments : 30%
- Midterm : 20%
- End-semester : 50%
- (Class tests, take-home assignments)

- Descriptive statistics.

1. Graphical representation of data.

Data
- Quantitative
  - Continuous
  - Discrete
- Qualitative (Eg. Hair Color, Grades in exams)
  - Nominal (Hair color)
  - Ordinal (Edn. degree)

Population

Sample

A subset of population

↓

(very often it is randomly selected)

Quantitative data :

10, 11, 8, 7, 12, 16, 15, 25, 18, 7, 6, ...

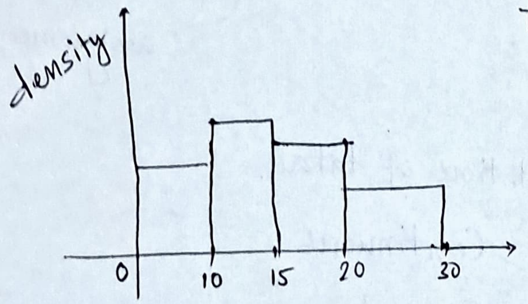
- Histogram :

Class intervals : $[0, 10)$, $[10, 15)$, $[15, 20)$, $[20, 30)$

Class intervals	Frequency	Relative Frequency	density
$[0, 10)$	f_1	f_1/n	$(f_1/n)/\omega_1$
$[10, 15)$	f_2	f_2/n	$(f_2/n)/\omega_2$
$[15, 20)$	f_3	f_3/n	$(f_3/n)/\omega_3$
$[20, 30)$	f_4	f_4/n	$(f_4/n)/\omega_4$
	n (sample size)	1	

ω_i = ith class width ~~density~~

Height of i-th class $= \dfrac{(f_i/n)}{\omega_i}$



- **Central tendency**:

Data 1: 10, 12, 8, 15, 5

- sample mean: $\bar{x} = \frac{1}{n}\sum x_i = \frac{10+12+8+15+5}{5} = 10$

Data 2: 10, 12, 8, 15, 5, 100

sample mean $= \frac{150}{6} = 25$

— Order the data from smallest to largest.

Data 1

5, 8, 10, 12, 15

sample median

Data 2

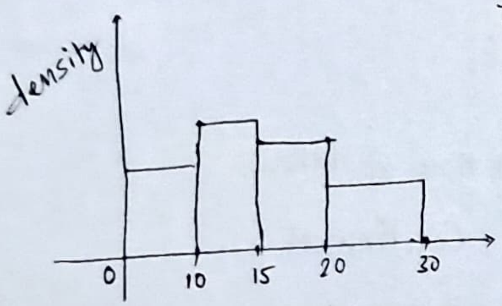
5, 8, $\boxed{10, 12}$ 15, 100

sample median $= \frac{10+12}{2} = 11$

Class intervals	Frequency	Relative Frequency	density
$[0, 10)$	f_1	f_1/n	$(f_1/n)/w_1$
$[10, 15)$	f_2	f_2/n	$(f_2/n)/w_2$
$[15, 20)$	f_3	f_3/n	$(f_3/n)/w_3$
$[20, 30)$	f_4	f_4/n	$(f_4/n)/w_4$
	n (sample size)	1	

w_i = ith class width ~~density~~

Height of i-th class $= (f_i/n)/w_i$



- Central tendency :

Data 1 : 10, 12, 8, 15, 5

- sample mean : $\bar{x} = \frac{1}{n}\sum x_i = \frac{10+12+8+15+5}{5} = 10$

Data 2 : 10, 12, 8, 15, 5, 100

sample mean $= \frac{150}{6} = 25$

— Order the data from smallest to largest.

Data 1

5, 8, 10, 12, 15

sample median

Data 2

5, 8, $\boxed{10, \ 12}$ 15, 100

sample median $= \frac{10+12}{2} = 11$

— Sample median is more robust than sample mean.

Calculating median : Order $x_{(1)} \le x_{(2)} \le x_{(3)} \le \cdots \le x_{(n)}$

$$\begin{cases} n \text{ is odd} & \text{— sample median } x_{(n/2)} \\ n \text{ is even} & \text{— sample median} = \frac{1}{2}\left(x_{(n/2)} + x_{(n/2+1)}\right) \end{cases}$$

- Trimmed mean — Trim α % of numbers from each end.
- Sample variance — $\{x_1, x_2, \ldots, x_n\} \leftarrow$ sample

Use $\bar{x}$: sample mean as center

$$(x_1-\bar{x})^2, (x_2-\bar{x})^2, \ldots, (x_n-\bar{x})^2$$

- sample variance $= \frac{1}{(n-1)} \sum (x_i - \bar{x})^2$
- Sample standard deviation $= \sqrt{\frac{1}{(n-1)} \sum (x_i-\bar{x})^2}$
- $\frac{1}{(n-1)} \sum (x_i - \bar{x})^2 = \frac{1}{(n-1)} \left[\sum x_i^2 - \frac{1}{n} \left(\sum x_i\right)^2 \right]$
- Inter-quartile range:

$x_1 \le x_2 \le \cdots \quad \cdots \le x_n$

sample median in 2nd half Q_3

Sample median in first half Q_1

sample median $= \tilde{x}$

Inter-quartile range $= Q_3 - Q_1$

1st quartile → median of

2nd quartile → median

3rd quartile → median

Eg: 8, 9, 11, 15, 18, 20, 21

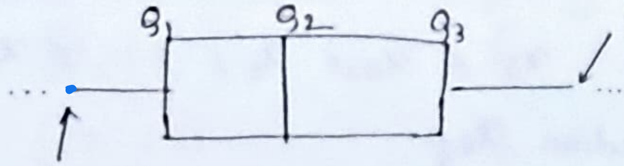
$Q_2 = 15$

$Q_1 = 10$

$Q_3 = 19$

$IQR = 19 - 10 = 9$

Boxplot:



max point that is not an outlier

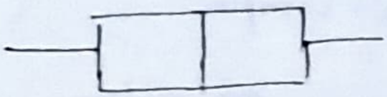
min point that is not an outlier

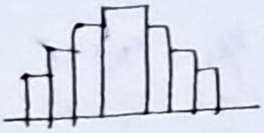
Outliers:

$> Q_3 + 1.5 \times IQR$

or

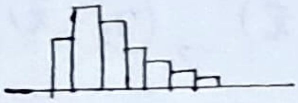
$< Q_1 - 1.5 \times IQR$

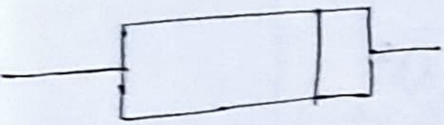




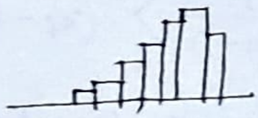
symmetric

right skewed

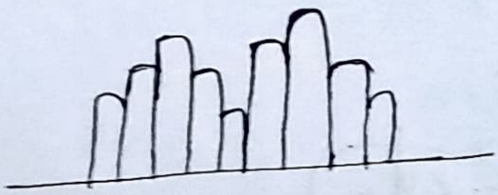




left skewed



— Bimodal distribution



Reading assignment — Chapter 1.

LEC-02

26/07/24

Sample space: Set of all possible outcomes in an experiment

Eg
- Tossing a coin. $S = \{H, T\}$
- Rolling a dice $S = \{1, 2, 3, 4, 5, 6\}$
- Max temp tomorrow $S = (10, 50)$

Event: A subset of the sample space.

A = the roll of a dice results in an odd number.

$A = \{1, 3, 5\}$, a subset of $\{1, 2, 3, 4, 5, 6\}$

Event operations:

S = Sample Space

Suppose A is an event. $A \subseteq S$.

- **Complement**

A^c or $\bar{A}$ (Complement of the event) $= S \setminus A$

$= \{$all elements of the sample space which are not in $A\}$

- Union: Let A and B are two events.

$A \cup B = \{$set of all elements either in A or in B or in both$\}$

- Intersection –

$A \cap B = \{$set of all elements in both A and $B\}$

- $(A \cup B)^c = A^c \cap B^c$
- $(A \cap B)^c = A^c \cup B^c$
- $A = (A \cap B) \cup (A \cap B^c)$
- If $U \cap V$ ~~are~~ $= \phi$, then U and V are said to be disjoint.

- **Probability:** $P(A)$ is a function. Domain $= \{$set of events$\}$

Range $= [0, 1]$

Axioms

i) $P(A) \geq 0$ for any event A

ii) $P(S) = 1$ $\qquad S =$ sample space

iii) If $A_1, A_2, \ldots$ are mutually disjoint events,

then $P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$

- $P(\phi) = 0$

Take $A_1 = A$, any event other that ϕ

$A_2 = A_3 = \cdots = \phi$

$P(A) = P(A) + P(\phi) + P(\phi) + \cdots$

This is possible when $P(\phi) = 0$

- $P(A^c) = 1 - P(A)$

$A \cup A^c = S \Rightarrow P(A \cup A^c) = P(S) = 1$

$A_1 = A, \quad A_2 = A^c, \quad A_3 = A_4 = \cdots = \phi$

$P(S) = P(A) + P(A^c) + P(\phi) + P(\phi) + \cdots$

$\Rightarrow 1 = P(A) + P(A^c)$

$\Rightarrow P(A^c) = 1 - P(A)$

- If A and B are two disjoint events, then

$P(A \cup B) = P(A) + P(B)$.

- $A_1, A_2, \ldots, A_n$ disjoint events.

$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i)$

If A, B are any two events, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

— $A \cup B = (A \cap B^c) \cup (A \cap B) \cup (A^c \cap B)$

$\therefore \quad P(A \cup B) = P(A \cap B^c) + P(A \cap B) + P(A^c \cap B)$

Also $\quad P(A) = P(A \cap B^c) + P(A \cap B)$

$P(B) = P(A^c \cap B) + P(A \cap B)$

Substituting,

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

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

- $P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{1}^{n} P A_i - \sum_{i<j} P(A_i \cap A_j) + \sum_{i<j<k} P(A_i \cap A_j \cap A_k) - \cdots + (-1)^{n+1} P(A_1 \cap A_2 \cap \cdots \cap A_n)$

- Counting Techniques:

1. Sample with replacement, order is important.
   — How many ways can we select k objects out of n? $\longrightarrow n^k$.

2. Sample without replacement, order is important.

$$n(n-1) \cdots (n-k+1) = \frac{n!}{(n-k)!}$$

$= \#$ of permutations

3. Sample without replacement, order is not important.

$$\frac{n!}{(n-k)!\; k!} = \# \text{ no of combinations}$$

$= \binom{n}{k}$;

(n choose k)

— Axioms i) $P(A) \geq 0$ for any event A

ii) $P(S) = 1$ S = sample space

iii) If $A_1, A_2, \dots$ are mutually disjoint events,

then $P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$.

- $P(\phi) = 0$

Take $A_1 = A$, any event ~~other~~ than ϕ.

$A_2 = A_3 = \dots = \phi$

$P(A) = P(A) + P(\phi) + P(\phi) + \dots$

This is possible when $P(\phi) = 0$

- $P(A^c) = 1 - P(A)$

$A \cup A^c = S \Rightarrow P(A \cup A^c) = P(S) = 1$

$A_1 = A$, $A_2 = A^c$, $A_3 = A_4 = \dots = \phi$

$P(S) = P(A) + P(A^c) + P(\phi) + P(\phi) + \dots$

$\Rightarrow 1 = P(A) + P(A^c)$

$\Rightarrow P(A^c) = 1 - P(A)$

- If A and B are two disjoint events, then

$P(A \cup B) = P(A) + P(B)$.

- $A_1, A_2, \dots, A_n$ disjoint events.

$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i)$

- If A, B are any two events, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

— $A \cup B = (A \cap B^c) \cup (A \cap B) \cup (A^c \cap B)$

$\therefore \quad P(A \cup B) = P(A \cap B^c) + P(A \cap B) + P(A^c \cap B)$

Also $\quad P(A) = P(A \cap B^c) + P(A \cap B)$

$\quad P(B) = P(A^c \cap B) + P(A \cap B)$

Substituting,

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

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

- $P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{1}^{n} P A_i - \sum_{i<j} P(A_i \cap A_j) + \sum_{i<j<k} P(A_i \cap A_j \cap A_k) - \dots + (-1)^{n+1} P(A_1 \cap A_2 \cap \dots \cap A_n)$

- Counting Techniques:

1. Sample with replacement, order is important.
   — How many ways can we select k objects out of n? $\longrightarrow n^k$.

2. Sample without replacement, order is important

$$n(n-1) \dots (n-k+1) = \frac{n!}{(n-k)!}$$
$$= \# \text{ of permutations}$$

3. Sample without replacement, order is not important.

$$\frac{n!}{(n-k)!\ k!} = \# \text{ no of combinations}$$
$$= \binom{n}{k}$$

(n choose k)

4. Sample with replacement, order is not important.

— = # of ways we can arrange (n−1) sticks and k stars. Order is not important.

$$| \quad |**| \quad |*|$$

$$= \binom{n+k-1}{k}$$

— $x_1 + x_2 + \cdots + x_n = k$, $x_i \geq 0$, x_i integer

no of solutions

- Suppose there are 6 black balls and 4 white balls. You randomly select 2 balls without replacement. What is the probability that one of them is black and other is white?

— $$\frac{\binom{6}{1}\binom{4}{1}}{\binom{10}{2}}$$

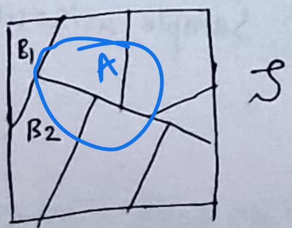
- Conditional Probability.

$P(A|B)$ (conditional probability of A given B)

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

$$\Rightarrow P(A \cap B) = P(A|B)\,P(B) = P(B|A)\,P(A)$$ ← Multiplication rules (provided $P(A) > 0$)

$$A = (A \cap B_1) \cup (A \cap B_2) \cup \cdots \cup (A \cap B_k)$$



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

$S = B_1 \cup B_2 \cup \cdots \cup B_k$

$B_i \cap B_j = \emptyset \quad i \neq j$

$$P(A) = \sum_{i=1}^{k} P(A|B_i)\, P(B_i)$$ (Law of total probability)

- Toss a coin, where $P(\text{Head}) = 0.6$

Toss a coin, if head select a ball from bag 1, if tail from bag 2.

B_1	B_2
6B, 4W	3B, 5W

$$P(\text{selected ball is white}) = P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2)$$

$$= \left(\frac{4}{10}\right)(0.6) + \left(\frac{5}{8}\right)(0.4)$$

Also,
$$P(B_j|A) = \frac{P(A \cap B_j)}{P(A)} = \frac{P(A|B_j)\,P(B_j)}{P(A)},$$

(Bays Theorem)

2.2.3 A pair of six-sided balanced dice are rolled.

LEC-03

31/07/24

$S = \{(1,1), (1,2), \dots, (6,6)\}$ $\quad |S| = 36$

(a) $A = \{x_1 + x_2 = 8\}$

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

$|A| = 5$

$\therefore P(A) = P(x_1 + x_2 = 8) = \dfrac{|A|}{|S|} = \dfrac{5}{36}$

(b) $B = \{x_1 + x_2 = 6 \text{ or } 9\}$

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

$\therefore P(B) = \dfrac{9}{36} = \dfrac{1}{4}$

(2.2.13) (A) $B = A \cup (B \cap A^c)$

$B \cup \phi \cup \cdots = A \cup (B \cap A^c) \cup \phi \cup \cdots$

$P(B) + P(\phi) + \cdots = P(A) + P(B \cap A^c) + P(\phi) + \cdots$

$\Rightarrow P(B) + 0 + 0 \cdots = P(A) + P(A^c \cap B) + 0 + \cdots$

Since $P(\phi) = 0$

$\Rightarrow P(B) \geq P(A)$. (need to be proved)

(2.2.16) $P(A) = 0.17$, $P(B) = 0.46$, $A \cap B = \phi$

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

(b) $P(A^c) = 1 - P(A)$

(c) $P(A^c \cup B^c) = P((A \cap B)^c) = 1 - P(A \cap B) = 1$

2.3.12

$15 \to$ 13 good, 2 defective

(a) P (none of the selected apples is defective)

$$= \frac{\binom{13}{4}}{\binom{15}{4}}$$

(b) P (at least one of the selected apples is defective)

$$= 1 - \frac{\binom{13}{4}}{\binom{15}{4}}.$$

2.3.19 (a) $|S| = 6^5$

~~(b)~~ • **Def^n:** A and B are said to be independent iff $P(A\cap B) = P(A)\cdot P(B)$, or $P(A|B) = P(A)$

2.4.5

(1) $P(A^c|B) = \dfrac{P(A^c\cap B)}{P(B)}$ $\qquad B = (A\cap B)\cup(A^c\cap B)$

$= \dfrac{P(B) - P(A\cap B)}{P(B)}$ * $\rightarrow$ $\dfrac{P(B)\left(1 - \dfrac{P(A\cap B)}{P(B)}\right)}{P(B)}$

$= \dfrac{P(B)(1-P(A))}{P(B)}$ $\qquad = \dfrac{P(B)(1-P(A))}{P(B)}$ [$\because$ A and B are independent].

$= 1 - P(A)$

$= P(A^c)$

2.4.13 $\quad$ D : Disease $\qquad$ + : detected disease

$P(+|D) = 0.98$

$P(+|D^c) = 0.005$

$P(D) = 0.02$

$$P(D|+) = \frac{P(+|D)\cdot P(D)}{P(+)}$$

$$= \frac{P(+|D)P(D)}{P(+|D)\cdot P(D) + P(+|D^c)\cdot P(D^c)}$$

$$= \frac{0.98\times 0.02}{(0.98)(0.02) + (0.005)\times \cancel{0.02}\,(1-0.02)}$$

2.4.20

P(1 was choosen | at least one white ball was removed) $\qquad \boxed{\begin{matrix}2W\\2B\end{matrix}}$

$= \dfrac{P(\text{at least one white ball was removed} \mid 1 \text{ was choosen})\cdot P(1 \text{ was choosen})}{P(\text{at least one white ball was removed})}$

$$= \frac{1/4\cdot 2/4}{1/4\cdot 2/4 + 1/4\cdot 5/6 + 1/4\cdot 1 + 1/4\cdot 1} = \frac{1/2}{1/2 + 5/6 + 1 + 1} = \frac{3}{20}$$

(2.2.13) (A) $B = A \cup (B \cap A^c)$

$B \cup \phi \cup \cdots = A \cup (B \cap A^c) \cup \phi \cup \cdots$

$P(B) + P(\phi) + \cdots = P(A) + P(B \cap A^c) + P(\phi) + \cdots$

$\Rightarrow P(B) + 0 + 0 \cdots = P(A) + P(A^c \cap B) + 0 + \cdots$

Since $P(\phi) = 0$

$\Rightarrow P(B) \geq P(A)$. (need to be proved)

(2.2.16) $P(A) = 0.17$, $P(B) = 0.46$, $A \cap B = \phi$

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

(b) $P(A^c) = 1 - P(A)$

(c) $P(A^c \cup B^c) = P((A \cap B)^c) = 1 - P(A \cap B) = 1$

2.3.12

$15 \to$ 13 good, 2 defective

(a) P (none of the selected apples is defective)

$$= \frac{\binom{13}{4}}{\binom{15}{4}}$$

(b) P (at least one of the selected apples is defective)

$$= 1 - \frac{\binom{13}{4}}{\binom{15}{4}}.$$

2.3.19 (a) $|S| = 6^5$

(b) • Def^n: A and B are said to be independent iff $P(A \cap B) = P(A) \cdot P(B)$, or $P(A \mid B) = P(A)$

<u>2.4.5</u>

(1) $P(A^c \mid B) = \dfrac{P(A^c \cap B)}{P(B)}$ $B = (A \cap B) \cup (A^c \cap B)$

$= \dfrac{P(B) - P(A \cap B)}{P(B)}$

$= \dfrac{P(B)(1 - P(A))}{P(B)}$

$= 1 - P(A)$

$= P(A^c)$

<u>2.4.13</u> D : Disease + : detected disease

$P(+ \mid D) = 0.98$

$P(+ \mid D^c) = 0.005$

$P(D) = 0.02$

$$P(D \mid +) = \frac{P(+ \mid D) \cdot P(D)}{P(+)}$$

$$= \frac{P(+ \mid D) P(D)}{P(+ \mid D) \cdot P(D) + P(+ \mid D^c) \cdot P(D^c)}$$

$$= \frac{0.98 \times 0.02}{(0.98) \times (0.02) + (0.005) \times \cancel{0.02}\,(1 - 0.02)}$$

<u>2.4.20</u>.

P (1 was choosen | at least one white ball was removed) $\begin{array}{|c|} 2W \\ 2B \end{array}$

$= \dfrac{P(\text{at least one white ball was removed} \mid 1 \text{ was choosen}) \cdot P(1 \text{ was choosen})}{P(\text{at least one white ball was removed})}$

$$= \frac{1/4 \cdot 2/4}{1/4 \cdot 2/4 + 1/4 \cdot 5/6 + 1/4 \cdot 1 + 1/4 \cdot 1} = \frac{1/2}{1/2 + 5/6 + 1 + 1} = \frac{3}{20}$$

Random Variables

Discrete random variable : Probability mass function (pmf) of X = $P(X=x)$, $x \in S$.

Eg Rolling a dice $P(X=i) = \frac{1}{6}$ $i = 1, 2, 3, 4, 5, 6$
(pmf of X)

$$\boxed{E(X) = \sum_{x \in S} x \cdot P(X=x)} \; (= \mu_X)$$

(or mean of X)

Eg $E(X) = \sum_{i=1}^{6} i \cdot \frac{1}{6} = 3.5$

$$\boxed{Var(X) = V(X) = \sum_{x \in S} (x-\mu_X)^2 P(X=x)}$$

Variance of the random variable X

$$= \sum x^2 P(X=x) - \left(\sum x P(X=x)\right)^2$$

$$\sum (x-\mu_X)^2 P(X=x)$$

$$= \sum (x^2 - 2\mu_X x + \mu_X^2) P(X=x)$$

$$= \sum x^2 P(X=x) - 2\mu_X \left(\sum x P(X=x)\right) + \mu_X^2 \left(\sum P(X=x)\right)$$

$$= \sum x^2 P(X=x) - 2\mu_X^2 + \mu_X^2$$

$$= \sum x^2 P(X=x) - \mu_X^2$$

$$= \sum x^2 P(X=x) - \left(\sum x P(X=x)\right)^2$$

- $E(h(x)) = \sum_{x \in S} h(x) P(X=x)$
- $E(X^2) = \sum x^2 P(X=x)$
- $E(aX) = a E(X)$ | $V(aX) = a^2 V(X)$
- $E(X+b) = E(X) + b$ | $V(X+b) = V(X)$
- $E(aX+b) = a E(X) + b$ | $V(aX+b) = a^2 V(X)$

$\boxed{\text{LEC-04}}$

02/08/24

$S / \Omega \longleftarrow$ sample space

$\omega \in \Omega$.

RV: $X : \Omega \to \mathbb{R}$ (mapping the sample space (outcomes) to Real Number).

$X(\omega) : \Omega \to \mathbb{R}$

Eg $\Omega = \{HH, HT, TH, TT\}$ X : No of H's

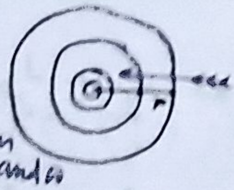
$X(\omega_1) = 2$, $X(\omega_2) = 1$, $X(\omega_3) = 1$, $X(\omega_4) = 0$

Eg Toss a coin until we get a Head

$\Omega = \{H, TH, TTH, \dots\}$

$X(\omega_1) = 0$, $X(\omega_2) = 1$, $X(\omega_3) = 2, \dots$

Eg



$0 \le R(\omega) \le r$ $R(\omega)$ = dist between center and ω

$R(\omega) \in [0, r]$

$R \longleftarrow$ a continuous random variable.

— $P(X=1) = P(X^{-1}(\{1\}))$

- **Discrete RV** : A random variable taking only countable (finite or infinite) number of isolated points $x_1, x_2, \dots$ with $P[X = x_i] > 0$ is called a discrete random variable.

$P(X=x_i) > 0$ $\qquad$ x_i : mass point / jump point

- PMF (Probability Mass function)

$f/p : \mathbb{R} \to [0,1]$

- Eg: Suppose $p(x) = \begin{cases} 2^{-x} & \text{if } x \in \{1,2,\dots\} \\ 0 & \text{o/w} \end{cases}$

Compute $P(X \in A)$ where
i) $A = \{1,2,3\}$
ii) $A = [5,8]$
iii) $A = \{1,3,5\}$
(iv) $A = \{1,3,5,\dots\}$

Soln i) $P(X \in \{1,2,3\})$

$= P(X \in \{1\}) + P(X \in \{2\}) + P(X \in \{3\})$

$= \frac{1}{2} + \frac{1}{2^2} + \frac{1}{2^3}$

(ii) $P(X \in [5,8])$

$= P(X=1) + P(X=2) + \dots + P(X=8)$

$= \frac{1}{2} + \frac{1}{2^2} + \dots + \frac{1}{2^8}$

(iii) $P(X \in \{1,3,5\})$

$= P(X=1) + P(X=3) + P(X=5) = \frac{1}{2} + \frac{1}{2^3} + \frac{1}{2^5}$

(iv) $P\left(X \in \{1,3,5,\dots\}\right)$

$= \cancel{P(X \in \Omega \setminus \{2,4\}) = 1 - \frac{1}{2^2} - \frac{1}{2^4}}$

$= P(X=1) + P(X=3) + P(X=5) + \cdots$

$= \sum_{n=1}^{\infty} \frac{1}{2^{2n-1}} = \frac{2}{3}$

<u>Alt def^n of pmf</u> $\quad p : \mathbb{R} \to [0,1]$ is a pmf iff it satisfies

I. Countable set $S \subseteq \mathbb{R}$ s.t $\forall x \notin S \quad P(X=x) = 0$

II. $\forall x \in S, \quad p(x) > 0$.

III. $\sum_{x \in S} p(x) = 1$

<u>Eg</u> $p(x) = \begin{cases} c \cdot \frac{\theta^x}{x} & , x = 1,2,\dots \\ 0 & \text{o/w} \end{cases}$

$\sum_{x=1}^{\infty} c \frac{\theta^x}{x} = 1.$

$c \cdot \sum_{x} \frac{\theta^x}{x} = 1$

For what values of θ and c, $p(x)$ is a pmf?

— $S = \{1,2,\dots\}$

$\begin{cases} c\frac{\theta}{1} > 0 \Rightarrow c\theta > 0 \\ c\frac{\theta^2}{1} > 0 \Rightarrow c\theta^2 > 0 \end{cases}$

$\Rightarrow c > 0, \theta > 0$

$\sum_{x \in S} p(x) = 1 \Rightarrow \sum_{n=1}^{\infty} c \cdot \frac{\theta^n}{n} = 1$

$\Rightarrow c(-\ln(1-\theta)) = 1.$ $\quad \cancel{(\text{if } |\theta| < 1)}$

$\Rightarrow c = \frac{1}{-\ln(1-\theta)}$

<u>Eg</u> Find the value of c for which $f(x) = \begin{cases} c\left(\frac{2}{3}\right)^x ; & x = 1,2,\dots \\ 0 & \text{o/w} \end{cases}$ is a pmf.

<u>Soln.</u> $f(1) > 0 \Rightarrow c > 0$

and $\sum_{n=1}^{\infty} f(n) = 1 \Rightarrow \sum_{n=1}^{\infty} c \cdot \left(\frac{2}{3}\right)^n = 1$

$\Rightarrow c \cdot \frac{2}{3} \cdot \frac{1}{1 - 2/3} = 1$

$\Rightarrow c = 1/2$

- Continuous random variable : A r.v X is said to be continuous random variable if it takes any value from its range of variation

Remark : For a continuous r.v $P(X = x) = 0 \quad \forall x$

- PDF (probability density function)

$f: \mathbb{R} \to [0, \infty)$ if $\forall\, a, b \in \mathbb{R}$, $P(X \in [a, b]) = \int_a^b f(x)\, dx$

- Ex. $f(x) = \begin{cases} e^{-x}, & x > 0 \\ 0, & \text{o/w.} \end{cases}$

Compute $P(X \in A)$, where A is (i) $[2, 5]$
(ii) $\{3\}$
(iii) $[2, 3)$

Soln : (i) $P(X \in [2, 5]) = \int_2^5 e^{-x}\, dx$

$= -e^{-5} + e^{-2}$

(ii) $P(X \in \{3\}) = P(X = 3) = 0$

(iii) $P(X \in [2, 3)) = \int_2^3 e^{-x}\, dx = -e^{-3} + e^{-2}$

RV

- Discrete
  - pmf
- continuous
  - pdf exists
  - pdf doesnot existe
- others

Att $f: \mathbb{R} \to [0, \infty)$ is a pmf of some random variable X

iff (I) $\forall x \; f(x) \geq 0$

(II) $\int_{-\infty}^{+\infty} f(x) = 1$

Eg $f(x) = \begin{cases} x/2, & 0 < x \leq 1 \\ 1/2, & 1 \leq x \leq 2 \\ \frac{3-x}{2}, & 2 < x \leq 3 \\ 0 & o/w \end{cases}$

Check whether $f(x)$ is a pmf.

Eg: $f(x) = \begin{cases} k & \text{if } x \in [0, 1/2] \\ 0 & o/w. \end{cases}$ is a pdf. Find k.

CDF (cumulative ~~density~~ distribution function)

X, $F : \mathbb{R} \to [0,1]$, $F(x) = P(X \le x) \; \forall x$.

Def: A real valued function F

(I) $x_1 < x_2$, then $F(x_1) \le F(x_2)$

(i.e. F is monotonically non-decreasing)

(II) $F(-\infty) = \lim_{x \to -\infty} F(x) = 0$

(III) $F(\infty) = \lim_{x \to \infty} (F(x)) = 1$

(IV) F is right continuous.

$F(x+0) = \lim_{k \to 0^+} F(x+k) = F(x)$

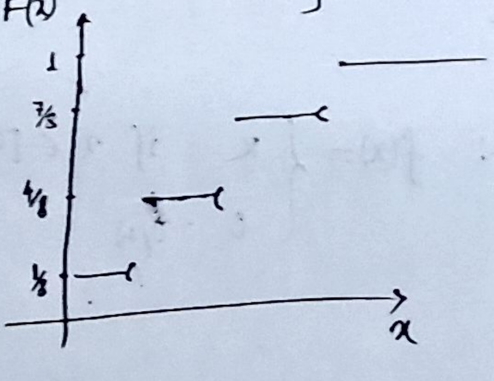
Remark I. F is not necessarily continuous.

II. A necessary and sufficient condition for a r.v X or its CDF F to be continuous at $X = x$ if $P(X = x) = 0$

Eg $X :=$ # Heads

$\Omega = \{ HHH, HHT, HTH, HTT, THH, THT, TTH, TTT \}$

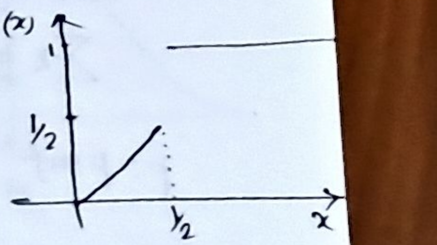
$$F(x) = \begin{cases} 0 & x < 0 \\ 1/8 & 0 \le x < 1 \\ 4/8 & 1 \le x < 2 \\ 7/8 & 2 \le x < 3 \\ 1 & x \ge 3 \end{cases}$$



$$F(x) = \begin{cases} \ldots & ; x < 0 \\ x & ; 0 \le x < \frac{1}{2} \\ 1 & ; x \ge \frac{1}{2} \end{cases}$$

Check whether the above function is a CDF ?

<u>soln</u> Right continuity at 0 :



$$\lim_{h \to 0^+} F(0+h) = \lim_{h \to 0^+} h = 0 = F(0)$$

Right continuity at $\frac{1}{2}$:

$$\lim_{h \to 0^+} F(\tfrac{1}{2}+h) = \lim_{h \to 0^+} 1 = 1 = F(\tfrac{1}{2})$$

~~Right continuity at~~ +

$\therefore$ F is right continuous.

Also $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to +\infty} F(x) = 1$.

F is also monotonically nondecreasing.

Hence f is a CDF.

Expectation and Variance of Random Variables

LEC-05

06/08/24

Discrete | **Continuous**

$\sum x\, p(x)$ $\quad x \in S$

p: pmf

- $P(x < X \leq x+h)$, $h>0$

$= P(X \leq x+h) - P(X \leq x)$

$= F_X(x+h) - F_X(x)$

$$\boxed{\begin{aligned} &\frac{d}{dx} F(x) = f(x) \\ &F(x) = \int_{-\infty}^{x} f(t)\,dt \\ &F_X(x) = P(X \leq x) \end{aligned}}$$

— $\frac{1}{h} P(x < X \leq x+h) = \frac{1}{h}\left(F_X(x+h) - F_X(x)\right)$

$$\lim_{h \downarrow 0} \left(P(x < X \leq x+h)\right) = \lim_{h \downarrow 0} \frac{1}{h}\left(F_X(x+h) - F_X(x)\right)$$

$$= \frac{d}{dx} F_X(x) \quad \text{(pdf)}$$

$\boxed{\text{NOTE}}$ pdf gives the information about the randomness of X.

$$\mu_X = E(X) = \int_{-\infty}^{+\infty} x f(x)\, dx$$

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

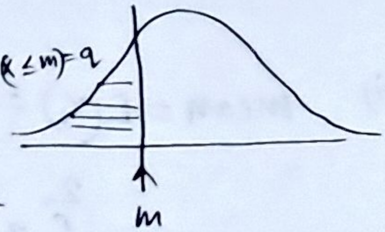
$$V(X) = E(X^2) - (E(X))^2 = E(X - E(X))^2 = \int_{-\infty}^{+\infty} (x-\mu_X)^2 f(x)\, dx$$

- Quantiles:

q — quantiles of X is m such that — (Assume X is continuous)

$P(X \le m) = q$

(100 q percentile)



0.5 — quantile of X

m such that — $P(X \le m) = 0.5$

$P(X > m) = 0.5$

$\shortparallel$

$P(X \geq m)$ ($\because X$ is continuous)

- Eg: pdf of X is $f(x) = \begin{cases} e^{-x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$

Calculate the median of X.

— Let m be the median of X.

Then $\int_{-\infty}^{m} f(x)\, dx = 0.5 \Rightarrow \int_{0}^{m} e^{-x}\, dx = 0.5$

$\Rightarrow -e^{-x}\Big|_0^m = 0.5 \Rightarrow 1 - e^{-m} = 0.5$

$\Rightarrow e^{-m} = \frac{1}{2} \Rightarrow -m = \ln \frac{1}{2} \Rightarrow m = -\ln \frac{1}{2} = \ln 2.$

(0.5 — q $\equiv$ 50 percentile)

- Suppose the pdf of X is $f(x) = \begin{cases} cx^3 & 1 \le x \le 2 \\ 0 & o/w \end{cases}$

i) find c.

ii) Find the mean and median of X.

Expectation and Variance of Random Variables

$\boxed{\text{LEC-05}}$

06/08/24

Discrete | **Continuous**

$\sum x\, p(x)$

$x \in S$

p: pmf

- $P(x < X \leq x+h)$, $h>0$

$= P(X \leq x+h) - P(X \leq x)$

$= F_X(x+h) - F_X(x)$

$$\frac{d}{dx} F(x) = f(x)$$

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

$$F_X(x) = P(X \leq x)$$

— $\frac{1}{h} P(x < X \leq x+h) = \frac{1}{h}\left(F_X(x+h) - F_X(x)\right)$

$$\lim_{h \downarrow 0} \left(P(x < X \leq x+h)\right) = \lim_{h \downarrow 0} \frac{1}{h}\left(F_X(x+h) - F_X(x)\right)$$

$$= \frac{d}{dx} F_X(x) \quad (\text{pdf})$$

$\boxed{\text{NOTE}}$ pdf gives the information about the randomness of X.

$$\mu_X = E(X) = \int_{-\infty}^{+\infty} x f(x)\,dx$$

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

$$V(X) = E(X^2) - (E(X))^2 = E(X - E(X))^2 = \int_{-\infty}^{+\infty} (x-\mu_X)^2 f(x)\,dx$$

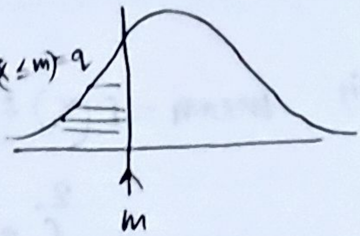
- Quantiles:

(Assume X is continuous)

q — quantile of X is m such that —

$$P(X \le m) = q$$

(100 q percentile)



0.5 — quantile of X

m such that — $P(X \le m) = 0.5$

$P(X > m) = 0.5$

$\parallel$

$P(X \ge m)$ ($\because X$ is continuous)

- Eg: pdf of X is $f(x) = \begin{cases} e^{-x}, & x \ge 0 \\ 0, & x < 0 \end{cases}$

Calculate the median of X.

— Let m be the median of X.

Then $\int_{-\infty}^{m} f(x)\,dx = 0.5 \Rightarrow \int_0^m e^{-x}\,dx = 0.5$

$\Rightarrow -e^{-x}\Big|_0^m = 0.5 \Rightarrow 1 - e^{-m} = 0.5$

$\Rightarrow e^{-m} = \frac{1}{2} \Rightarrow -m = \ln \frac{1}{2} \Rightarrow m = -\ln\frac{1}{2} = \ln 2.$

(0.5-q $\equiv$ 50 percentile)

- Suppose the pdf of X is $f(x) = \begin{cases} cx^3 & 1 \le x \le 2 \\ 0 & o/w \end{cases}$

i) find c.

ii) Find the mean and median of X.

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

$\Rightarrow \int_{1}^{2} c x^2 \, dx = 1 \quad \Rightarrow \frac{c}{3}(8-1) = 1 \quad \Rightarrow c = 3/7$

(ii) $\text{mean} = E(X) = \int_{-\infty}^{\infty} x f(x)\, dx$

$= \int_{1}^{2} \frac{3}{7} x^3 \, dx = \frac{3}{7 \cdot 4}(16-1) = \frac{45}{28}$

Let m be the median.

$P(X \leq m) = \frac{1}{2}$

$\Rightarrow \int_{1}^{m} \frac{3}{7} x^3 \, dx = 1/2 \quad \Rightarrow \frac{1}{7}(m^3 - 1) = 1/2$

$\Rightarrow m^3 = 7/2 + 1 = 9/2 \quad \Rightarrow m = (9/2)^{1/3}$

- $P(a \leq X \leq b) = F_X(b) - F_X(a-)$

$P(X > b) = 1 - P(X \leq b) = 1 - F_X(b)$ CDF

$P(X \geq b) = 1 - P(X < b) = 1 - F_X(b-)$

(left limit of $f(x)$ at b)

- **Joint distribution of (X, Y)**

→ Suppose, both X and Y are discrete R.V.

Joint p.m.f of (X, Y):

$p(x, y) = P(X = x \text{ and } Y = y), \quad x \in S_X, \; y \in S_Y$

- $\sum_x \sum_y p(x,y) = 1$, $p(x,y) \geq 0$.

- $P(x \in A) = \sum_{x \in A} \sum_{y \in S_Y} p(x,y)$

- $P(X \in A \mid Y \in B) = \dfrac{P((X \in A) \cap (Y \in B))}{P(Y \in B)}$

$$= \frac{\sum_{x \in A} \sum_{y \in B} p(x,y)}{\sum_{x \in S_X} \sum_{y \in B} p(x,y)}$$

- $E(h(X,Y)) = \sum_{x \in S_X} \sum_{y \in S_Y} h(x,y)\, p(x,y)$

- $E(X) = \sum_{x \in S_X} \sum_{y \in S_Y} x\, p(x,y)$

- Eg: pmf of (X,Y)

$y \backslash x$	0	1	2
0	0.1	0	0.4
1	0.2	0.1	0.2

Calculate (i) $E(X^2)$

(ii) $P(X=2 \mid Y=1)$

(iii) $P(Y=0)$

i) $E(x^2) = \sum x^2 \cdot p(x,y) \Rightarrow 0 \times 0.3 + 1 \times (\underbrace{0 + 0.1}_{p(1,0)+p(1,1)})$
$+ 4(0.4 + 0.2) = 0.1 + 2.4$
$= 2.5$

(ii) $P(X=2 \mid Y=1) = \dfrac{P(X=2 \text{ and } Y=1)}{P(Y=1)} = \dfrac{0.2}{0.5} = 0.4$

(iii) $P(Y=0) = 0.5$

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

$\Rightarrow \int_{1}^{2} cx^2\,dx = 1 \quad \Rightarrow \frac{c}{3}(8-1) = 1 \quad \Rightarrow c = 3/7$

(ii) $\text{mean} = E(X) = \int_{-\infty}^{\infty} x f(x)\,dx$

$= \int_{1}^{2} \frac{3}{7} x^3\,dx = \frac{3}{7 \cdot 4}(16-1) = \frac{45}{28}$

Let m be the median.

$P(X \le m) = \frac{1}{2}$

$\Rightarrow \int_{1}^{m} \frac{3}{7} x^3\,dx = 1/2 \quad \Rightarrow \frac{1}{7}(m^3 - 1) = 1/2$

$\Rightarrow m^3 = 7/2 + 1 = 9/2 \quad \Rightarrow m = (9/2)^{1/3}$

- $P(a \le X \le b) = F_X(b) - F_X(a-)$

$P(X > b) = 1 - P(X \le b) = 1 - F_X(b)$

$P(X \ge b) = 1 - P(X < b) = 1 - F_X(b-)$

(left limit of $f(x)$ at b)

- **Joint distribution of (X, Y)**

→ Suppose, both X and Y are discrete R.V.

Joint p.m.f of (X,Y):

$p(x,y) = P(X = x \text{ and } Y = y)$, $x \in S_X$, $y \in S_Y$

- $\sum_{x} \sum_{y} p(x,y) = 1$, $p(x,y) \geq 0$.

- $P(x \in A) = \sum_{x \in A} \sum_{y \in S_Y} p(x,y)$

- $P(X \in A \mid Y \in B) = \dfrac{P((X \in A) \cap (Y \in B))}{P(Y \in B)}$

$$= \frac{\sum_{x \in A} \sum_{y \in B} p(x,y)}{\sum_{x \in S_X} \sum_{y \in B} p(x,y)}$$

- $E(h(X,Y)) = \sum_{x \in S_X} \sum_{y \in S_Y} h(x,y)\, p(x,y)$

- $E(X) = \sum_{x \in S_X} \sum_{y \in S_Y} x\, p(x,y)$

- Eg: pmf of (X,Y)

$y \backslash x$	0	1	2
0	0.1	0	0.4
1	0.2	0.1	0.2

Calculate (i) $E(X^2)$

(ii) $P(X=2 \mid Y=1)$

(iii) $P(Y=0)$

i)

(ii) $P(X=2 \mid Y=1) = \dfrac{P(X=2 \text{ and } Y=1)}{P(Y=1)} = \dfrac{0.2}{0.5} = 0.4$

(iii) $P(Y=0) = 0.5$

→ Suppose both X and Y are continuous

Joint pdf of (X, Y) : $f(x,y)$

$$P(X \in A \text{ and } Y \in B) = \int_{x \in A} \int_{y \in B} f(x,y)\, dy\, dx.$$

$$P(X \in A \mid Y \in B) = \frac{P(X \in A \text{ and } Y \in B)}{P(Y \in B)}$$

$$= \frac{\int_{x \in A} \int_{y \in B} f(x,y)\, dy\, dx}{\int_{x \in S_X} \int_{y \in B} f(x,y)\, dy\, dx}$$

- $$E(h(X,Y)) = \cancel{\int_{x \in S_X} \int_{y \in S_Y} x f(x,y)\, dy\, dx} \int_{x \in S_X} \int_{y \in S_Y} h(x,y) f(x,y)\, dy\, dx$$

- $$E(X) = \int_{x \in S_X} \int_{y \in S_Y} x f(x,y)\, dy\, dx$$

- Joint pdf of (X, Y) :

$$f(x,y) = \begin{cases} c(x+y^2), & 0 \le x \le 1 \text{ and } 0 \le y \le 1 \\ 0, & \text{o/w} \end{cases}$$

Find c. Calculate $P(X \le 0.5 \mid Y \le 0.6)$

— $$\int_{x=0}^{1} \int_{y=0}^{1} c(x+y^2) = 1$$

$$\Rightarrow \int_0^1 c\left(x + \frac{1}{3}\right) = 1 \quad \Rightarrow c\left(\frac{x^2}{2} + \frac{x}{3}\right)\Bigg|_{x=0}^{1} = 1$$

$$\Rightarrow c \cdot \left(\tfrac{1}{2} + \tfrac{1}{3}\right) = 1 \quad \Rightarrow c = 6/5$$

$$P(X \le 0.5 \mid Y \le 0.6) = \frac{P(X \le 0.5 \text{ and } Y \le 0.6)}{P(Y \le 0.6)}$$

$$= \frac{\int_{x=0}^{1/2} \int_{y=0}^{3/5} \frac{6}{5}(x+y^2)\,dy\,dx}{\int_{x=0}^{1} \int_{y=0}^{3/5} \frac{6}{5}(x+y^2)\,dy\,dx}$$

$$= \frac{\int_{x=0}^{1/2} \frac{3x}{5}}{\qquad}$$

?!

- Calculate $P(X \le 0.5 \mid Y = 0.6)$

Calculate pdf of X, give $Y = 0.6$

$$\frac{\frac{6}{5}(X + 0.6^2)}{\int_0^1 6/5\,(t + 0.6^2)\,dt}$$

To make it a pdf we divide by this

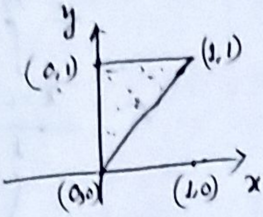
- Conditional distribution of X, given Y

$$\cancel{f_{X}}\ f_{X|Y} = (x \mid y^*)$$

$$= \frac{f(x, y^*)}{\int_{x \in S_X} f(x, y^*)\,dx}$$

$$f(x,y) = \begin{cases} cxy^2 & 0 \le x \le y < 1 \\ 0 & o/w \end{cases}$$

Find c, and $P(X \le 0.5)$.



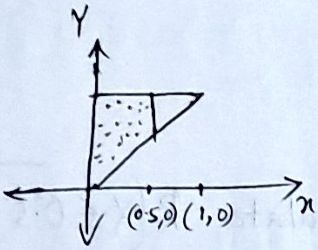
$$\int_{x=0}^{1}\int_{y=x}^{1} cxy^2 = 1 \; dy\,dx$$

$$\Rightarrow \frac{c}{3}\int_0^1 x(1-x^3) = 1 \Rightarrow \frac{c}{3}\int_0^1 (x - x^4)\,dx = 1$$

$$\Rightarrow \frac{c}{3}\left(\frac{x^2}{2} - \frac{x^5}{5}\right)\Big|_0^1 = 1$$

$$\Rightarrow \frac{c}{3}\left(\tfrac{1}{2} - \tfrac{1}{5}\right) = 1 \Rightarrow c = 10$$

$P(X \le 0.5)$



$$= \int_{x=0}^{1/2}\int_{y=x}^{1} 10xy^2\,dy\,dx$$

$$= \frac{10}{3}\int_0^{1/2} (x - x^4)\,dx$$

$$= \frac{10}{3}\left(\frac{x^2}{2} - \frac{x^5}{5}\right)\Big|_0^{1/2} = \frac{10}{3}\left(\frac{1}{8} - \frac{1}{32\times 5}\right)$$

$$= \frac{10}{24}\left(1 - \frac{1}{20}\right) = \frac{\cancel{10}\times 19}{24\times \cancel{20}_2} = \frac{19}{48}$$

LEC-06

09/08/24

Generating Function

(Moment Generating Function / MGF)

I. rth order raw moment is $E(X^r)$

II. rth order central moment is $E.(X - E(X))^r$

MGF: Suppose X is a random variable. Then MGF of X is defined as

$$M_X(t) = E(e^{tX}) \quad \cancel{t \in \mathbb{R}}$$

— We say that MGF of X exist if $M_X(t)$ is finite for all $t \in (-a, a)$, (a is some positive ~~constant)~~ for some a.

- Lemma: If the expectation does not exists in an open nbd of 0, then we say that MGF doesn't exist.

(neighbourhood)

- Lemma: It generates all the moments of X

$$M_X(t) = E(e^{tX})$$

$$= E\left[1 + tX + \frac{t^2X^2}{2!} + \frac{t^3X^3}{3!} + \cdots\right]$$

$$= 1 + tE(X) + \frac{t^2}{2!}E(X^2) + \cdots$$

$$\therefore E(X^r) = \frac{d^r}{dt^r}(M_X(t)).$$

- MGF uniquely determines the distribution, if exists.

- Eg: $X \sim f(x) = \frac{1}{2}e^{-|x|}$, $x \in \mathbb{R}$. Find the MGF of X.

Soln $M_X(t) = E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} \cdot \frac{1}{2} e^{-|x|} \, dx$

$$\cancel{= \frac{1}{2}\int_{-\infty}^{0} e^{tx} e^{x}\, dx + \frac{1}{2}\int_{0}^{\infty} e^{tx} e^{-x}\, dx}$$

$$\cancel{= \frac{1}{2(t+1)} e^{(t+1)x}\Big|_{-\infty}^{0} + \frac{1}{2}\, \frac{e^{t(x-1)}}{t}\Big|_{0}^{\infty}}$$

$$\cancel{= \frac{1}{2t} + \frac{1}{2t}} = \frac{1}{2}\left[\cancel{\frac{e^{(t+1)x}}{(t+1)}\Big|_{-\infty}^{0} + \frac{e^{(t-1)x}}{(t-1)}\Big|_{0}^{\infty}}\right]$$

$$\cancel{= \frac{1}{2}\cdot \frac{e^{\infty}}{1}\, t}$$

$$= \frac{1}{2}\left[\int_{-\infty}^{0} e^{(t+1)x}\, dx + \int_{0}^{\infty} e^{(t-1)x}\, dx\right]$$

$$= \frac{1}{2}\left[\int_{0}^{\infty} e^{-(t+1)x}\, dx + \int_{0}^{\infty} e^{(t-1)x}\, dx\right]$$

The value of the integration will be finite if

$$-(t+1) < 0 \quad \text{and} \quad (t-1) < 0$$

$$\Rightarrow \quad -1 < t < 1$$

Homework: Compute $E(X)$ and $var(X)$ from $M_X(t)$.

- Theorem: Let X and Y be two independent random variable with MGF $M_X(t)$ and $M_Y(t)$, respectively. Define $Z = X+Y$. Then $M_Z(t) = M_X(t) \cdot M_Y(t)$.

$\rightarrow Z = X_1 + X_2 + \cdots + X_n$

$$M_Z(t) = \prod_{i=1}^{n} M_{X_i}(t)$$

- Some Standard distributions:

I. Bernoulli's R.V : If a random variable has two possible outcomes (say 0 and 1), then this random variable is known as Bernoulli's RV.

$X \sim$ Bernoulli's RV

If $P(X=x) = \begin{cases} (1-p) & \text{if } x=0 \\ p & \text{if } x=1 \\ 0 & \text{o/w} \end{cases}$

$= p^x(1-p)^{1-x} \quad , x \in \{0,1\}$

$E(X) = p$

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

$M_X(t) = (1-p) + pe^t$

$M_X(t) = E(e^{tX})$

$= e^{t.0}\cdot(1-p) + e^{t.1}\cdot p$

$= (1-p) + pe^t.$

Suppose we perform n independent Bernoulli trials (with outcome 0 and 1).

$\binom{n}{r} \leftarrow$ exactly r many success

Binomial Distribution:

$Y_1, Y_2, \ldots, Y_n \in$ Bernoulli (p)

$X = \sum_{i=1}^{n} Y_i \quad ,$ # of success in n trials.

$X \sim$ Binomial Random variable [Binomial (n, p)]

$P(X=x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & x = 0,1,2,\ldots,n \\ 0 & \text{o/w.} \end{cases}$

— proposition: $E(X) = np$

$X = \sum Y_i \Rightarrow E(X) = \sum_{i=1}^{n} E(Y_i) = np$

$$= \frac{1}{2}\int_{-\infty}^{0} e^{tx} e^{x} dx + \frac{1}{2}\int_{0}^{\infty} e^{tx} e^{-x} dx$$

$$= \frac{1}{2(t+1)} e^{(t+1)x}\Big|_{-\infty}^{0} + \frac{1}{2} \frac{e^{t(x-1)}}{t}\Big|_{0}^{\infty}$$

$$= \cancel{\frac{1}{2t}} + \cancel{\frac{1}{2t}} = \frac{1}{2}\left[\frac{e^{(t+1)x}}{(t+1)}\Big|_{-\infty}^{0} + \frac{e^{(t-1)x}}{(t-1)}\Big|_{0}^{\infty}\right]$$

$$\cancel{= \frac{1}{2} \cdot \frac{e^{\infty}}{1} +}$$

$$= \frac{1}{2}\left[\int_{-\infty}^{0} e^{(t+1)x} dx + \int_{0}^{\infty} e^{(t-1)x} dx\right]$$

$$= \frac{1}{2}\left[\int_{0}^{\infty} e^{-(t+1)x} dx + \int_{0}^{\infty} e^{(t-1)x} dx\right]$$

The value of the integration will be finite if

$$-(t+1) < 0 \quad \text{and} \quad (t-1) < 0$$

$$\Rightarrow -1 < t < 1$$

<u>Homework</u>: Compute $E(X)$ and $var(X)$ from $M_X(t)$.

- Theorem: Let X and Y be two independent random variable with MGF $M_X(t)$ and $M_Y(t)$, respectively. Define $Z = X+Y$. Then $M_Z(t) = M_X(t) \cdot M_Y(t)$.

$\rightarrow Z = X_1 + X_2 + \cdots + X_n$

$$M_Z(t) = \prod_{i=1}^{n} M_{X_i}(t)$$

- Some standard distributions:

I. Bernoulli's R.V. : If a random variable has two possible outcomes (say 0 and 1), then this random variable is known as Bernoulli's RV.

$X \sim$ Bernoulli's RV

$$\text{If } P(X=x) = \begin{cases} (1-p) & \text{if } x=0 \\ p & \text{if } x=1 \\ 0 & \text{o/w} \end{cases}$$

$$= p^x (1-p)^{1-x}, \quad x \in \{0,1\}$$

$E(X) = p$

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

$M_X(t) = (1-p) + pe^t.$

$M_X(t) = E(e^{tX})$

$= e^{t.0} \cdot (1-p) + e^{t.1} \cdot p$

$= (1-p) + pe^t.$

Suppose we perform n independent Bernoulli trials (with outcome 0 and 1).

$\binom{n}{r} \leftarrow$ exactly r many success

Binomial Distribution:

$Y_1, Y_2, \ldots, Y_n \in$ Bernoulli (p)

$X = \sum_{i=1}^{n} Y_i$, # of success in n trials.

$X \sim$ Binomial Random variable [Binomial (n, p)]

$$P(X=x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & x = 0,1,2,\ldots,n \\ 0 & \text{o/w.} \end{cases}$$

PMF of Bin. Dist^n: $(p+q)^n$

— proposition: $E(X) = np$

$X = \sum Y_i \Rightarrow E(X) = \sum_{i=1}^{n} E(Y_i) = np$

$$V(X) = \sum_{i=1}^{n} V(Y_i) = np(1-p)$$

$$M_X(t) = E(e^{tX})$$

$$= \sum_{i=0}^{n} e^{ti} \binom{n}{i} p^i (1-p)^{n-i}$$

$$= (pe^t + (1-p))^n$$

$$M_X(t) = \prod_{i=1}^{n} M_{Y_i}(t)$$

$$= (1-p+pe^t)^n$$

$$\sum_{i=0}^{n} {}^nC_i (e^t \cdot p)^i (1-p)^{n-i}$$

$p \to e^t p$, $q \to (1-p)$.

Just like PMF of binomial distr.: $(e^t p + q)^n$ where $q = (1-p)$.

Example: A and B play a game in which A's choice of winning is $2/3$. In a series of 8 games what is the probability that A will win at least 6 games.

— Let $X \sim$ # of times A wins.

Then the required probability $= P(X \geq 6)$

$= P(X=6) + P(X=7) + P(X=8)$

Example: Two dice are tossed together a number of times in succession. Find the minimum number of throws to ensure that the probability of obtaining 6 on both dice at least once is greater than 0.5.

ans $= 25$

— Let $X = \#$ times $(6,6)$

$$P(X<1) = P(X=0)$$

$$= \binom{n}{0}\left(\frac{1}{36}\right)^0 \left(\frac{35}{36}\right)^n < 0.5$$

- Poisson Distribution : $X \sim \text{Poisson}(\lambda)$

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

$E(X) = \lambda$

$V(X) = \lambda$

$M_X(t) = e^{\lambda(e^t-1)}$.

- Theorem : $X \sim \text{Binomial}(n, p)$. Suppose $n \to \infty$, $p \to 0$, and $\boxed{np = \lambda}$.

Then $\lim\limits_{n\to\infty} P(X = x) \longrightarrow e^{-\lambda} \dfrac{\lambda^x}{x!}$

$$\lim_{n\to\infty} P(X=x) = \lim_{n\to\infty} \binom{n}{x} p^x (1-p)^{n-x}$$

$$= \lim_{n\to\infty} \frac{n(n-1)\cdots(n-x+1)}{x!} p^x (1-p)^{n-x}$$

$$= \lim_{n\to\infty} \frac{n(n-1)\cdots(n-x+1)}{x!} \left(\frac{\lambda}{n}\right)^x \left(1-\frac{\lambda}{n}\right)^{n-x}$$

$$= \lim_{n\to\infty} \frac{n(n-1)\cdots(n-x+1)}{x!} \cdot \left(\frac{\lambda/n}{1-\lambda/n}\right)^x \cdot \underbrace{\lim_{n\to\infty} \left(1-\lambda/n\right)^n}$$

$$e^{\lim_{n\to\infty}\left(1-\frac{\lambda}{n}-1\right)n}$$

$$e^{\lim_{n\to\infty} -\lambda}$$

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

$$= \frac{\lambda^x \cdot e^{-\lambda}}{x!} \lim_{n\to\infty} \frac{n(n-1)\cdots(n-x+1)}{(n-\lambda)^x}$$

$\longrightarrow 1$ as $n \to \infty$

$$= e^{-\lambda} \cdot \frac{\lambda^x}{x!}$$

Eg ~~An~~ individual suffers an ~~advance~~ adverse reaction from a particular drug is known to be 0.001. Determine the probability that out of 2000 individual

i) Exactly three will suffer $X \sim Bin(2000, 0.\ldots)$

ii) More than two will suffer.

— $P(X=3) = e^{-\lambda} \cdot \frac{\lambda^3}{3!}$ $\lambda = 2000 \times 0.001.$

- Continuous distribution

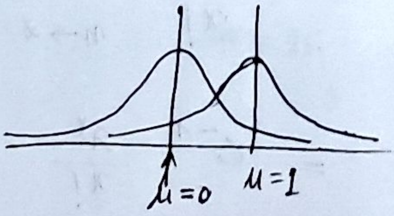
i) Normal / Gaussian Distribution

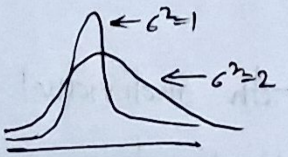
A r.v $X \sim$ Gaussian dist

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2\sigma^2} \cdot (x-\mu)^2} \quad ; \quad x \in \mathbb{R}, \ \mu \in \mathbb{R}, \ \sigma^2 > 0$$

If $\mu = 0$, $\sigma^2 = 1$
↳ standard normal.

$$\boxed{\begin{aligned} E(X) &= \mu \\ V(X) &= \sigma^2 \\ M_X(t) &= e^{\mu t + \frac{1}{2}\sigma^2 t^2} \end{aligned}}$$





Properties : $Z \sim N(0,1)$

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \quad , z \in \mathbb{R}$$

CDF of $N(0,1) := \Phi$

$\Phi(z) = P(Z \le z)$

I. $\lim_{x \to \infty} \Phi(x) = 1$

II. $\lim_{x \to -\infty} \Phi(x) = 0$

III. ~~lim~~ $\Phi(0) = 1/2$

IV. $\Phi(-x) = 1 - \Phi(x)$

proof of IV.

$$1 - \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-x^2/2} dx$$

$$= \frac{1}{\sqrt{2\pi}} \int_{x}^{\infty} e^{-x^2/2} dx$$, Let $y = -x \rightarrow y^2 = x^2$. $dy = -dx$

$$= \frac{1}{\sqrt{2\pi}} \int_{-x}^{-\infty} e^{-y^2/2} (-dy) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-x} e^{-y^2/2} dy = \Phi(-x)$$

$$\phi(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

$\mu = 0, \sigma^2 = 1$

| LEC-07 |

13/08/24

$N(\mu, \sigma^2) \rightarrow$ Normal distr.

- pdf $:= \phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}, \quad -\infty < x < \infty$

$N(0,1)$

- cdf $\quad \Phi(x) = \int_{-\infty}^{x} \phi(t)\, dt$

- pdf of $N(\mu, \sigma^2)$, $\quad f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty$.

- Property:

1. If $X \sim N(\mu, \sigma^2)$, then $\frac{X-\mu}{\sigma} \sim N(0,1)$

soln $\quad P\left(\frac{X-\mu}{\sigma} \le t\right) = P(X \le (\mu + \sigma t))$

$$= \int_{-\infty}^{\mu+\sigma t} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2\sigma^2}(w-\mu)^2} dw$$

Take $y = \frac{w-\mu}{\sigma}$

$$= \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-y^2/2} \cdot \sigma\, dy$$

$$= \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} \, dy = \Phi(t).$$

This is true for any $t \in (-\infty, \infty)$

$\therefore$ The cdf of $\left(\frac{X-\mu}{\sigma}\right)$ is same as Φ.

$\therefore \frac{X-\mu}{\sigma} \sim N(0,1)$.

- MGF of $N(0,1)$

$$M_Z(t) = E\left(e^{tZ}\right) \quad \text{where } Z \sim N(0,1)$$

$$= \dots = e^{\frac{t^2}{2}}$$

proof:
$$E\left(e^{tZ}\right) = \int_{-\infty}^{\infty} e^{tw} \frac{1}{\sqrt{2\pi}} e^{-\frac{w^2}{2}} \, dw$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{(w-t)^2}{2}} \cdot e^{+\frac{t^2}{2}} \, dw$$

$$= e^{\frac{1}{2}t^2} \underbrace{\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} \, dy}_{\text{from } -\infty \to \infty \text{ the area under the curve } = 1.} \qquad \begin{aligned} y &= w-t \\ dy &= dw \end{aligned}$$

$$= e^{t^2/2}.$$

- Find the mgf of $N(\mu, \sigma^2)$.

$$M_X(t) = E\left(e^{tX}\right) \quad \text{, where } X \sim N(\mu, \sigma^2)$$

$$= \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \, dx$$

$$E(e^{wx}) = e^{\mu w + \frac{1}{2}\sigma^2 w^2}$$

Property 2. If $X \sim N(\mu, \sigma^2)$, then $(aX+b) \sim N(a\mu+b, a^2\sigma^2)$

proof: Use mgf. [If $M_X(t) = M_Y(t)$ for all t in a small open neighbourhood containing 0, then $F_X \equiv F_Y$]

$$M_{aX+b}(t) = E\left(e^{taX+tb}\right)$$

$$= e^{tb} \cdot e^{\mu ta + \frac{1}{2}\sigma^2 t^2 a^2}$$

$$= e^{(a\mu+b)t + \frac{1}{2}\sigma^2 t^2 a^2} = \text{MGF of } N(a\mu+b, a^2\sigma^2)$$

Property 3: If $X \sim N(\mu_1, \sigma_1^2)$, $Y \sim N(\mu_2, \sigma_2^2)$, X and Y are independent, then $(a_1X + a_2Y) \sim N(a_1\mu_1 + a_2\mu_2, a_1^2\sigma_1^2 + a_2^2\sigma_2^2)$

proof: using mgf.

$$= \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy = \Phi(t).$$

This is true for any $t \in (-\infty, \infty)$

$\therefore$ The cdf of $\left(\frac{X-\mu}{\sigma}\right)$ is same as Φ.

$$\therefore \frac{X-\mu}{\sigma} \sim N(0,1).$$

- MGF of $N(0,1)$

$$M_Z(t) = E(e^{tZ}) \qquad \text{where } Z \sim N(0,1)$$

$$= \dots = e^{\frac{t^2}{2}}$$

proof: $$E(e^{tZ}) = \int_{-\infty}^{\infty} e^{tw} \frac{1}{\sqrt{2\pi}} e^{-\frac{w^2}{2}} dw$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{(w-t)^2}{2}} \cdot e^{+\frac{t^2}{2}} dw$$

$$= e^{\frac{1}{2}t^2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy. \qquad y = w-t, \quad dy = dw$$

$$= e^{t^2/2}.$$

- Find the mgf of $N(\mu, \sigma^2)$.

$$M_X(t) = E(e^{tX}) \quad \text{, where } X \sim N(\mu, \sigma^2)$$

$$= \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

Pt r1 wordfle

$E(e^{wx}) = e^{\mu w + \frac{1}{2}\sigma^2 w^2}$

Property 2. If $X \sim N(\mu, \sigma^2)$, then $(aX+b) \sim N(a\mu+b, a^2\sigma^2)$

proof: Use mgf. [If $M_X(t) = M_Y(t)$ for all t in a small open neighbourhood containing 0, then $F_X \equiv F_Y$]

$$M_{aX+b}(t) = E\left(e^{taX+tb}\right)$$

$$= e^{tb} \cdot e^{\mu ta + \frac{1}{2}\sigma^2 t^2 a^2}$$

$$= e^{(a\mu+b)t + \frac{1}{2}\sigma^2 t^2 a^2} = \text{MGF of } N(a\mu+b, a^2\sigma^2)$$

Property 3: If $X \sim N(\mu_1, \sigma_1^2)$, $Y \sim N(\mu_2, \sigma_2^2)$, X and Y are independent, then $(a_1X + a_2Y) \sim N(a_1\mu_1 + a_2\mu_2, a_1^2\sigma_1^2 + a_2^2\sigma_2^2)$

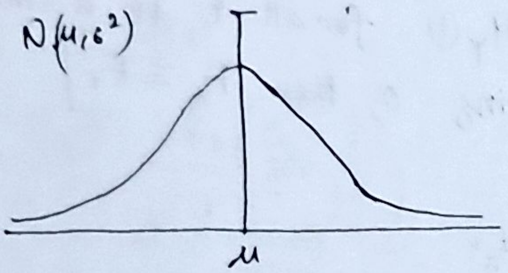
proof: using mgf.

— $X \sim N(\mu, \sigma^2)$

$E(X) = \mu$

$E(X^2) = \mu^2 + \sigma^2$

$Var(X) = (\mu^2 + \sigma^2) - \mu^2$
$= \sigma^2$



μ is mean, median, mode

---

1. Suppose $X \sim N(5, 8^2)$. Calculate $P(X \le 7)$.

cdf of $N(0,1)$ Φ

- $P(X \le 7) = P\left(\frac{X-5}{8} \le \frac{7-5}{8}\right)$

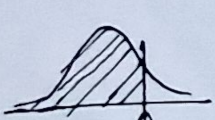
$= P\left(Z \le \frac{1}{4}\right)$ $\quad Z = \frac{X-5}{8} \sim N(0,1)$.

$= \Phi(0.25)$

- $P(X \geq 7) = 1 - \Phi(0.25)$

- $P(3 \le X \le 6) = \Phi\left(\frac{6-5}{8}\right) - \Phi\left(\frac{3-5}{8}\right)$.

Find q-quantile of $N(0,1)$.



$\Phi(m_q) = q$, $\quad m_q = \Phi^{-1}(q)$.

Find q-quantile of $N(\mu, \sigma^2)$.

$P(X \le m_q) = q \Rightarrow P\left(Z \le \frac{m_q - \mu}{\sigma}\right) = q$

$$\Rightarrow \Phi\left(\frac{m_q - \mu}{6}\right) = q$$

$$\Rightarrow m_q = 6\,\Phi^{-1}(q) + \mu.$$

- Linear Combination Random Variables.

$a_1 X_1 + a_2 X_2 + \cdots + a_n X_n.$

$$E(a_1X_1 + a_2X_2 + \cdots + a_nX_n) = \sum_{i=1}^{n} a_i E(X_i)$$

$$V(a_1X_1 + a_2X_2 + \cdots + a_nX_n) = \sum_{i=1}^{n} a_i^2 V(X_i) \quad \text{if } X_i\text{'s are independent}$$

- Example: Suppose, the pmf of X is as follows.

$$P(X=-1) = \tfrac{1}{3}, \quad P(X=0) = \tfrac{1}{3}, \quad P(X=1) = \tfrac{1}{3}$$

Suppose X_1, X_2, X_3 are independent and each has the same distribution of X. Find the distribution of $(X_1 + X_2 + X_3)$

Soln: Take $T = X_1 + X_2 + X_3$

Let's try to find the pdf of T.

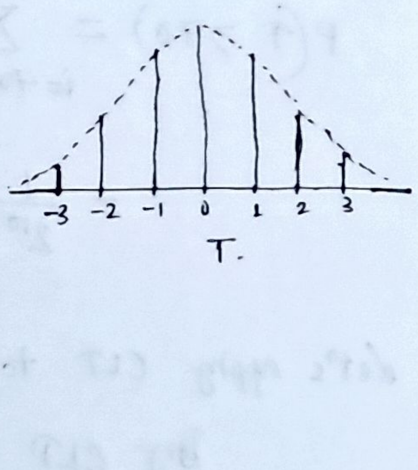
$P(T=-3) = P(X_1=-1, X_2=-1, X_3=-1) = \left(\frac{1}{3}\right)^3$

$P(T=-2) = 3\cdot\left(\frac{1}{3}\right)^3$ (since there are 3 different ways)

$P(T=-1) = 6\cdot\left(\frac{1}{3}\right)^3$ $\qquad P(T=2) = 3\left(\frac{1}{3}\right)^3$

$P(T=0) = 7\left(\frac{1}{3}\right)^3$ $\qquad P(T=3) = \left(\frac{1}{3}\right)^3$

$P(T=1) = 6\cdot\left(\frac{1}{3}\right)^3$

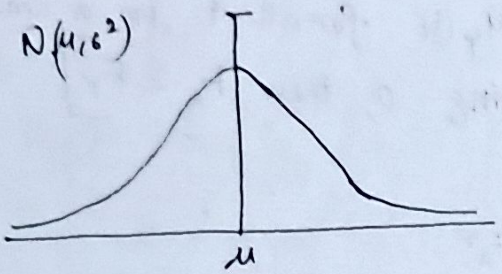


$X \sim N(\mu, \sigma^2)$

$E(X) = \mu$

$E(X^2) = \mu^2 + \sigma^2$

$\mathrm{var}(X) = (\mu^2 + \sigma^2) - \mu^2$
$= \sigma^2$



μ is mean, median, mode

---

1. Suppose $X \sim N(5, 8^2)$. Calculate $P(X \le 7)$.

cdf of $N(0,1)$
Φ

- $P(X \le 7) = P\left(\frac{X-5}{8} \le \frac{7-5}{8}\right)$

$= P\left(Z \le \frac{1}{4}\right)$ $\qquad Z = \frac{X-5}{8} \sim N(0,1)$.

$= \Phi(0.25)$

- $P(X \geq 7) = 1 - \Phi(0.25)$

- $P(3 \le X \le 6) = \Phi\left(\frac{6-5}{8}\right) - \Phi\left(\frac{3-5}{8}\right)$.

Find q-quantile of $N(0,1)$.

$\Phi(m_q) = q$, $\quad m_q = \Phi^{-1}(q)$.

Find q-quantile of $N(\mu, \sigma^2)$.

$P(X \le m_q) = q \Rightarrow P\left(Z \le \frac{m_q - \mu}{\sigma}\right) = q$

Needs Revision.

$\Rightarrow \Phi\left(\frac{m_q - \mu}{\sigma}\right) = q$

$\Rightarrow m_q = \sigma\,\Phi^{-1}(q) + \mu.$

- **Linear Combination Random Variables.**

$a_1X_1 + a_2X_2 + \cdots + a_nX_n.$

$$E(a_1X_1 + a_2X_2 + \cdots + a_nX_n) = \sum_{i=1}^{n} a_i E(X_i)$$

$$V(a_1X_1 + a_2X_2 + \cdots + a_nX_n) = \sum_{i=1}^{n} a_i^2 V(X_i) \quad \text{if } X_i\text{'s are independent}$$

- Example: Suppose, the pmf of X is as follows.

$P(X=-1) = \frac{1}{3}, \quad P(X=0) = \frac{1}{3}, \quad P(X=1) = \frac{1}{3}$

Suppose X_1, X_2, X_3 are independent and each has the same distribution of X. Find the distribution of $(X_1 + X_2 + X_3)$

Soln: Take $T = X_1 + X_2 + X_3$

Let's try to find the pdf of T.

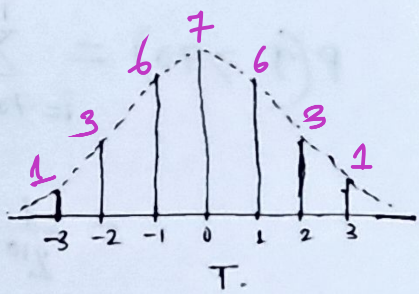
$P(T=-3) = P(X_1=-1, X_2=-1, X_3=-1) = \left(\frac{1}{3}\right)^3$

$P(T=-2) = 3\cdot\left(\frac{1}{3}\right)^3$ (since there are 3 different ways)

$P(T=-1) = 6\cdot\left(\frac{1}{3}\right)^3 \qquad P(T=2) = 3\left(\frac{1}{3}\right)^3$

$P(T=0) = 7\left(\frac{1}{3}\right)^3 \qquad P(T=3) = \left(\frac{1}{3}\right)^3$

$P(T=1) = 6\cdot\left(\frac{1}{3}\right)^3$



NOTE If $x_1, x_2, \dots, x_n$ is a random sample from a distribution, with (X_i's are iid)

mean μ and variance σ^2 (finite);

if n is large, then the approximate distribution of $T = \sum_{i=1}^{n} X_i$ is $N(n\mu, n\sigma^2)$. (finite)

$E(X_i) = \mu$
$V(X_i) = \sigma^2$

So the approx distribution of $\bar{X} = \frac{1}{n} \sum X_i$ is $N\left(\mu, \frac{\sigma^2}{n}\right)$

This is called Central Limit Theorem (CLT).

Ordinary ($n > 30$) is considered large enough.

But if the original dist is Normal, then n is much lesser than 30.

• Eg ~~$X_i = \dots$~~ $X_i = \begin{cases} 0 & \text{prob } (1-p) \\ 1 & \text{prob } p \end{cases}$

X_i's are independent $\therefore T = \sum_{i=1}^{n} X_i$

$T = \sum_{i=1}^{n} X_i \sim \text{Binomial}(n, p)$. (exact distribution)

• Prob: Suppose you toss a fair coin 100 times independently. What is the probability that you get at least 70 head counts?

soln: Suppose $p = P(\text{Head}) = 0.5$.

$$P(T \geq 70) = \sum_{i=70}^{100} \binom{100}{i} (0.5)^i (0.5)^{100-i}$$

$$= \frac{1}{2^{100}} \sum_{i=70}^{100} \binom{100}{i}.$$

Let's apply CLT to approximate the same probability.

By CLT $T \sim N(np, np(1-p))$

i.e, $N\left(100 \times \frac{1}{2}, 100 \times \frac{1}{2} \times \frac{1}{2}\right)$, i.e. $N(50, 25) = N(50, 5^2)$

$$P(T \geq 70)$$
$$= 1 - P(T \leq 69.5)$$
$$= 1 - \Phi\left(\frac{69.5 - 50}{5}\right)$$
$$= 1 - \Phi(3.9\ldots)$$
$$\approx 0.$$

$$\boxed{\begin{aligned} P(T \leq a) &\approx P(T \leq a + 0.5) \\ P(T \geq a) &\approx P(T \geq a - 0.5) \end{aligned}}$$

$$1 - P\left(\frac{T-\mu}{\sigma} \leq \frac{69.5-\mu}{\sigma}\right)$$
$$= 1 - P\left(Z \leq \frac{69.5 - 50}{\sigma}\right)$$
$$\stackrel{d}{=} 1 - \Phi\left(Z \leq \frac{69.5-50}{5}\right)$$

Example 3.5.6

A soft-drink vending machine is set so that the amount of drink dispensed is a random variable with a mean of 8 ounces and a standard deviation of 0.4 ounces. What is the approximate probability that the average of 36 randomly chosen fills exceed 8.1 ounces?

Solution

From the CLT, $((\overline{X} - 8)/(0.4/\sqrt{36})) \sim N(0, 1)$. *Hence, from the normal table,*

$$P\{\overline{X} > 8.1\} = P\left\{Z > \frac{8.1 - 8.0}{\frac{0.4}{\sqrt{36}}}\right\}$$
$$= p\{Z > 1.5\} = 0.0668.$$