Date
Page

Random Experiment:

An experiment whose outcome is unknown.

Ex: Tossing a coin, Waiting time in a queue

Sample space:

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

$\Omega = \{H, T\}$ - Tossing a coin
$\Omega = [0, \infty)$ - Waiting time in a queue
$\Omega = \{1, 2, 3, 4, 5, 6\}$ - Rolling a die

Event:

Subset of the sample space.

Sample Spaces
- Discrete - $\{H, T\}$, $\{1, 2, \dots 6\}$, $\mathbb{N}$
- Continuous - $[0, \infty)$, (a, b)

Case: discrete / finite / all outcomes are equally likely

Let A be an event.

$$p(A) = \frac{\#A}{\#\Omega}$$ Relative frequency assignment

Scanned by CamScanner

Case: discrete / finite / all outcomes are NOT equally likely.

→ Run the experiment n number of times
→ by f_A denotes the relative frequency

$$f_A = \frac{\text{no. of times event A occured in n runs}}{n}$$

$$p(A) = \lim_{n \to \infty} f_A$$

Observation: Only for certain key "events" one needs to assign probabilities.

For instance in case of rolling a die we need to assign only $p(1)$, $p(2)$, $p(3)$, $p(4)$ & $p(5)$. Then probabilities of all other events are automatically assigned

How does this automatic assignment happens??
- Because we intuitively assume certain laws of probility
- We will formalise these laws / these intutions into the axioms and propose an axiomatic definition of probability.

Why we need axiomatic definition?
- it works in all possible scenes.
- it matches with the ideas stated previously about probability assignment.

Scanned by CamScanner

Event : A subset of Ω.

Union / intersection of events is also an event.

Definition : σ - field

A collection of sets $\mathcal{F}$ is called σ - algebra if

i) $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$

ii) $A_1, A_2, A_3, \ldots \in \mathcal{F}$, then

$$\bigcup_{i=1}^{\infty} A_i \in \mathcal{F} \text{ and } \bigcap_{i=1}^{\infty} A_i \in \mathcal{F}$$

Ex: i) $\Omega = \{H, T\}$

$$\mathcal{F} = 2^{\Omega} = \{\phi, \{H\}, \{T\}, \{H, T\}\}$$

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

$$\mathcal{F} = 2^{\Omega} = \{\phi, \{1\}, \{2\}, \{3\}, \ldots \{1, 2, 3, 4, 5, 6\}\}$$

Definition: Let Ω be the sample space for a random experiment. Let $\mathcal{F}$ be the σ-algebra associated with Ω. Then a probability measure P is a function on $\mathcal{F}$ to $\mathbb{R}$ ($P : \mathcal{F} \to \mathbb{R}$) such that:

i) $P(A) \geq 0 \quad \forall A \in \mathcal{F}$

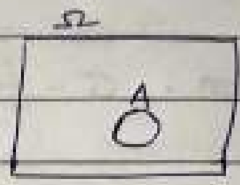
ii) $P(\Omega) = 1$

iii) If A_i are mutually disjoint for $i = 1, 2, \ldots$ ($A_i \cap A_j = \phi$ for $i \neq j$)

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

Properties :

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



$A \cup A^c = \Omega$

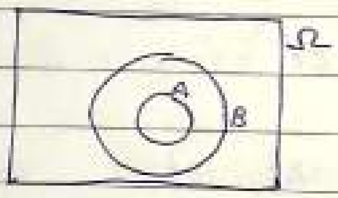
$$P(A \cup A^c) = P(A) + P(A^c) = P(\Omega)$$

because A & A^c are disjoint

ii) $P(\phi) = 0$

Take $A = \Omega$ in previous result.

iii) If $A \subset B$, then $P(A) \le P(B)$



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

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

$$P(B \cap A^c) \ge 0$$

$$P(B) \ge P(A)$$

Scanned by CamScanner

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

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

$\hookrightarrow$ disjoint union

$$P(A \cup B) = P(A \cap B^c) + P(A \cap B) + P(A^c \cap B) \quad \text{— (i)}$$

Now, notice

$$A = (A \cap B^c) \cup (A \cap B)$$
$$P(A) = P(A \cap B^c) + P(A \cap B)$$
$$P(A \cap B^c) = P(A) - P(A \cap B) \quad \text{— (ii)}$$

Similarly,

$$P(A^c \cap B) = P(B) - P(A \cap B) \quad \text{— (iii)}$$

From (i), (ii) & (iii)

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

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

Similarly, using induction we can prove

$$P(A_1 \cup A_2 \cap A_3 \ldots) = \ldots\ldots\ldots$$

Scanned by CamScanner

Examples:

i) Experiment: Toss 2 fair coins

$$\Omega = \{(HH), (H,T), (TH), (TT)\}$$

$$\mathcal{F} = 2^{\Omega}$$

Assume:

$p(HH) = \frac{1}{4}$ $\quad$ $p(HH) = p_1$

$p(HT) = \frac{1}{4}$ $\quad$ $p(HT) = p_2$

$p(TH) = \frac{1}{4}$ $\quad$ $p(TH) = p_3$

$p(TT) = \frac{1}{4}$ $\quad$ $p(TT) = 1 - p_1 - p_2 - p_3$

$0 \le p_1, p_2, p_3 < 1$

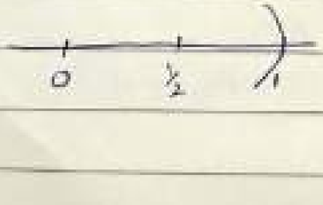
$0 \le p_4 < 1$

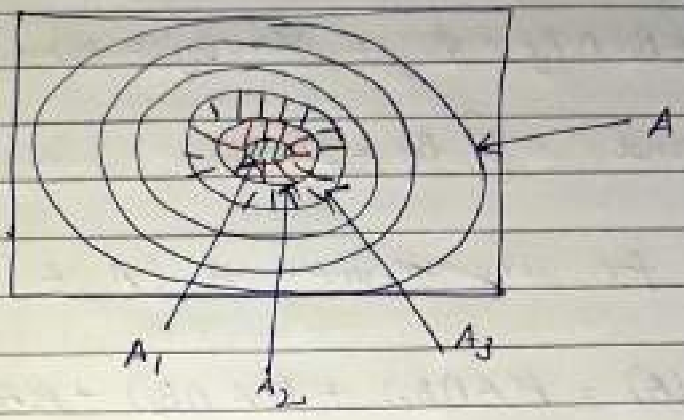
Theorem 1: Let $A_1 \subset A_2 \subset A_3 \subset \ldots\ldots$ (sequence of increasing sets)

and $A = \bigcup_{i=1}^{n} A_i$ then

$$P(A) = \lim_{n \to \infty} P(A_n)$$

Ex:





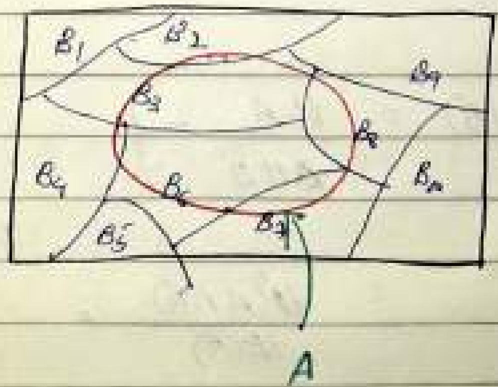
<u>Proof:</u>

$$B_1 = A_1$$
$$B_2 = A_2 \cap A_1^c$$
$$B_3 = A_3 \cap A_2^c$$
$$\vdots$$
$$B_n = A_n \cap A_{n-1}^c$$

$$\bigcup_{i=1}^{n} B_i = \bigcup_{i=1}^{n} A_i$$

$$A = \bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{\infty} B_i$$

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



Let $\{B_1, B_2 \dots B_n\}$ be a disjoint cover of Ω

$$\left(B_i \cap B_j = \phi \quad \forall \; i,j = 1,2, \ldots, n \right)$$
$$\text{and} \quad \bigcup_{i=1}^{n} B_i = \Omega$$

Then for any event A $(A \in \mathcal{F})$

$$P(A) = P(A \cap B_1) + P(A \cap B_2) + P(A \cap B_3) \cdots + P(A \cap B_n)$$

Conditional Probability:

Let $A, B \in \mathcal{F}$ such that $P(B) > 0$

Conditional Probability of A given B.

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

Ex. Rolling a fair dice

A: 2 appears on the top face of the die
B: even number appears on the top face of the die.

$$P(A|B) = \frac{\#A}{\#B} = \frac{1}{3}$$

$A = \{2\}$
$B = \{2, 4, 6\}$

$$= \frac{\#(A \cap B)}{\#(B)}$$

$$= \frac{\#(A \cap B) / \#(\Omega)}{\#(B) / \#(\Omega)}$$

← motivation for the formula of conditional probability

Scanned by CamScanner

$P(A \cap B) = P(A|B) \times P(B)$ .... (product rule of probability)
$\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad$ └ (i)

If

$P(A) > 0$

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

$$P(A \cap B) = P(B|A) \times P(A)$$
└ (ii)

From (1) & (2)

$$P(A|B) \times P(B) = P(B|A) \times P(A)$$

$$\boxed{\begin{aligned} P(A|B) &= \frac{P(B|A) \times P(A)}{P(B)} \\ P(B|A) &= \frac{P(A|B) \times P(B)}{P(A)} \end{aligned}}$$

Baye's Theorem

Consider $\{B_1, B_2, \ldots, B_n\}$ disjoint cover of Ω

$$\left[\begin{array}{ll} B_i \cap B_j = \phi & \text{for } i \neq j \\ \bigcup_{i=1}^{n} B_i = \Omega & \end{array} \right]$$

For any $A \in \mathcal{F}$

$A = (A \cap B_1) \cup (A \cap B_2) \cup (A \cap B_3) \ldots \cup (A \cap B_n)$

↑ disjoint union

Scanned by CamScanner

$$P(A) = P(A \cap B_1) + P(A \cap B_2) + \cdots\cdots + P(A \cap B_n)$$

$$= P(A|B_1) \cdot P(B_1) + P(A|B_2) \cdot P(B_2) + \cdots + P(A|B_n) \cdot P(B_n)$$

$$P(A) = \sum_{i=1}^{n} P(A|B_i) \cdot P(B_i) \quad \{\text{Total Law of Probability}\}$$

$$P(B_j|A) = \frac{P(A|B_j) \cdot P(B_j)}{\sum_{i=1}^{n} P(A|B_i) \cdot P(B_i)} \quad \{\text{Baye's Rule}\}$$

Independence:

Conditional probability of A given B ($P(B) > 0$)

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

Ex: $\Omega = \{1, 2, 3, 4\}$ all the outcomes are equally likely

$A = \{1, 2\}$, $B = \{1, 3\}$, $C = \{1, 4\}$

$A_1 = \{1, 2, 3\}$ $A_2 = \{1\}$

$B_1 = \{1, 2, 3\}$ $B_2 = \{1, 2, 4\}$

$P(A_2 | A_1) = \frac{1}{3}$ $\quad P(A_2) = \frac{1}{4}$

$P(B_1 | B_2) = \frac{2}{3}$ $\quad P(B_1) = \frac{3}{4}$

$P(A) = 1/2$ $\quad P(A|B) = \frac{1}{2}$

$P(B) = 1/2$ $\quad P(B|A) = \frac{1}{2}$

Scanned by CamScanner

classmate
Date
Page

$$P(A \cap B) = P(A) \cdot P(B|A)$$

$$\boxed{P(A \cap B) = P(A) \cdot P(B)}$$

Pairwise independence of events

$$P(A \cap B) = P(A) \cdot P(B)$$

It is clear that A, B are pairwise independent and so are A, C and B, C in the previous examples.

$$\boxed{P(C \mid A \cap B) = \frac{P(C \cap A \cap B)}{P(A \cap B)} = 1}$$

↳ leads to the definition of mutual independence.

$$P(A \cap B \cap C) = P(A) \cdot P(B) \cdot P(C)$$

Definition:

Let A, B be two events from $(\Omega, \mathcal{F}, P)$. We call A & B pairwise independent if

$$P(A \cap B) = P(A) \cdot P(B)$$

Definition:

Let A, B, C be three events from $(\Omega, \mathcal{F}, P)$. We call A, B, C mutually independent if

$$P(A \cap B) = P(A) \cdot P(B) \; ; \; P(B \cap C) = P(B) \cdot P(C);$$
$$P(A \cap C) = P(A) \cdot P(C) \; ; \; P(A \cap B \cap C) = P(A) \cdot P(B) \cdot P(C)$$

Scanned by CamScanner

We can generalize the definition of mutual independence to the events $A_1, A_2, \ldots, A_n$ in an inductive way.

- Any subcollection of $A_1, \ldots, A_n$ containing at least 2 events and atmost $(n-1)$ events is mutually independent.

- $P(A_1 \cap A_2 \cap A_3 \cdots A_n) = P(A_1) \cdot P(A_2) \cdots P(A_n)$

* Setting up probability space in case of repeated random experiment (with only 2 possible outcomes)

Experiment: Tossing a coin

$$\Omega = \{0, 1\}$$

$$\mathcal{F} = 2^{\Omega} = \{\phi, \Omega, \{0\}, \{1\}\}$$

$$P(\{0\}) = 1-p \quad ; \quad P(\{1\}) = p$$

Set up probability space when one repeats R 'n' number of times with the assumption

For this compound experiment

$$\Omega = \{ y_1, y_2, \ldots, y_n \;/\; y_1, y_2, \ldots, y_n \in \{0, 1\}\}$$

Scanned by CamScanner

$\# \Omega = 2^n$

$\mathcal{F} = 2^\Omega$ power set of Ω

Probability assignment:

Probability of k successes in n repetitions of specified locations. Consider an element $\omega \in \Omega$

$$\omega = y_1 y_2 \cdots y_n$$

where $y_1, y_2, \ldots, y_k = 1$; $y_{k+1} = \cdots y_n = 0$

$$\Big(\omega = \underbrace{111\ldots\ldots1}_{k \text{ times}}\ \underbrace{000\ldots\ldots0}_{(n-k) \text{ times}}\Big)$$

A_i = success at i^{th} toss for $i = 1, 2, \ldots, n$

$$P(\omega) = P(y_1 \ldots y_n) = P(\underbrace{111\ldots1}_{k}\underbrace{000\ldots0}_{n-k})$$

mutual independence

$$= P(A_1 \cap A_2 \cap A_3 \cdots \cap A_{k+1}^c \cap A_{k+2}^c \cdots \cap A_n^c)$$

$$= P(A_1)P(A_2)\ldots P(A_k)P(A_{k+1}^c)\ldots P(A_n^c)$$

Now, if one is interested in computing the probability of the event that exactly k number of successes in the coin tosses.

$$= \binom{n}{k} p^k (1-p)^{n-k}$$

Scanned by CamScanner

n

Examples: A box contains 10 balls.

6 - red — 60%

4 - blue — 40%

Case 1: Pick a ball and replace it in the box and pick a ball again

$A_1 \rightarrow$ red ball is picked in the first trial.

$A_2 \rightarrow$ blue ball is picked in the second trial.

$$P(A_2 | A_1) = P(A_2) = \frac{4}{10} = 0.4$$

sampling with replacement

Case 2: Pick a ball; do not put it back in the box & pick a ball again.

(Sampling without replacement)

$A_1 \rightarrow$ red ball is picked in the first trial.

$A_2 \rightarrow$ blue ball is picked in the second trial

$$P(A_2 | A_1) = \frac{4}{9} \neq P(A_2) = \frac{6}{10} \times \frac{4}{9} + \frac{4}{10} \times \frac{3}{9} = \frac{36}{90} = \frac{2}{5}$$

Thus A_1 & A_2 are not independent.

Increase 'n' and keep the proportions of the red balls and blue balls constant.

[A_1 & A_2 start behaving more and more like independent event]

Uniform Probability Space:

Experiment: Picking up a number randomly from a given set.

Given set is finite and discrete.

$\{1, 2, \ldots, n\}$

Probability of two subsets of Ω with equal size be equal. (This is the uniform probability principle)
Here we associate size to cardinality of the set.
As a consequence we can see that the singleton sets are assigned probability $1/n$.

Case 2: The given set Ω is finite and an interval of $\mathbb{R}$.

In particular, take $\Omega = [0, 1]$

Principle of uniformity: Pick two subsets of Ω with equal 'size', then they should have same probability.

Take an interval $(a, b) \subseteq [0, 1]$
Size of (a, b) is associated with the length of (a, b) which is $b-a$. Probability assignment by obeying principle of uniformity.

Any two intervals of same length are assigned with probability

Scanned by CamScanner

The σ-algebra $\mathcal{F}$, here is the set subsets of $[0,1]$ which are generated as countable unions and intersection of intervals.
(Borel σ-algebra)

This is the continuous uniform probability space.

Remarks:

1. Note that any singleton set $\{c\} \subseteq [0,1]$ has probability zero
2. Consider the set of rational numbers in $[0,1]$ denoted as $Q \cap [0,1]$
Thus from (1) and the fact that $Q \cap [0,1]$ is countable

$$P(Q \cap [0,1]) = 0$$

Scanned by CamScanner

classmate
Date
Page

Question 1:

1. A machine contains 4 components in parallel with 0.1, 0.2, 0.3, 0.4 as their probabilities of failures respectively. The machine fails if all the components fail simultaneously. Note that the failure of machines is independent. Then what is the probability that the machine once started will not fail?

Let A_i: event that the component i fails.

This means $A_1 \cap A_2 \cap A_3 \cap A_4$ = machine fails

$$1 - P[A_1 \cap A_2 \cap A_3 \cap A_4] = P(A_1 \cap A_2 \cap A_3 \cap A_4)^c$$

$$P(\overline{A_1 \cap A_2 \cap A_3 \cap A_4}) = 1 - P(A_1)P(A_2) \cdot P(A_3) \cdot P(A_4)$$
$$= 1 - 0.1 \times 0.2 \times 0.3 \times 0.4 = 1 - 0.0024$$
$$= 0.9976$$

2. A rocket engine fails if one key component fails. The probability of failure of this key component is 5%. In order to increase the success probability of the rocket engine an assembly of this key components in parallel is proposed so that the engine fails if all these key components fail simultaneously. What is the minimum number of key components are required in the parallel assembly so that the engine has 99% of success probability?

$$1 - (0.05)^n > 0.99$$
$$1 - 0.99 \geq (0.05)^n$$
$$0.01 \geq (0.05)^n$$

Scanned by CamScanner

$$\log(0.01) \geq n \log(0.05)$$

3. A course on probability is a very famous course and students are allowed to register after the teacher's consent. It is observed that 20% times after obtaining the consent, student fails to register for the course. There are only 100 seats available in the class. If a teacher consulted 102 students, what is the probability that all the students will be accomodated in the class?

$$P(E) = 1 - \left((0.8)^{102} + {}^{102}C_{101} \times (0.8)^{101} \times 0.2 \right)$$

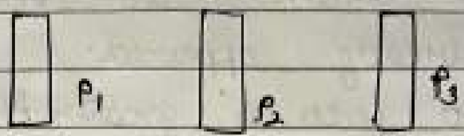
4. In a class of 100 students -

30 - MI
40 - IM
30 - AG

70% of MI students are dual degree
40% of IM students are dual degree
50% of AG students are dual degree.

Scanned by CamScanner

5. There 3 printers



Printer P_1 gets - 50% jobs, P_2 gets 30% jobs, P_3 gets 25% jobs

Printer P_1 has 0.15 as its probability of failure.
Printer P_2 has 0.1 as its probability of failure
Printer P_3 has 0.2 as its probability of failure

If a randomly selected printing job is a failure. what is the probability that it was printed by Printer 3?

F: failed job

$P(F|P_1) = 0.15$ $\quad$ $P(F|P_2) = 0.1$ $\quad$ $P(F|P_3) = 0.2$

$P(P_1) = 0.5$ $\quad$ $P(P_2) = 0.3$ $\quad$ $P(P_3) = 0.2$

Find $P(P_3|F)$

Applying Baye's theorem:

$$P(P_3|F) = \frac{P(F|P_3) \cdot P(P_3)}{P(F|P_1) \cdot P(P_1) + P(F|P_2) \cdot P(P_2) + P(F|P_3) \cdot P(P_3)}$$

$$P(P_3|F) = \frac{0.2 \times 0.2}{0.15 \times 0.5 + 0.1 \times 0.3 + 0.2 \times 0.2}$$

Scanned by CamScanner

6. 2 fair dice are thrown. What is the probability that you will obtain two sixes given that
i) one six has already appeared.
ii) sum of the two faces is greater than six.

i) $P(E) = \frac{1}{6}$

ii) $P(A|B)$

A : event two sixes

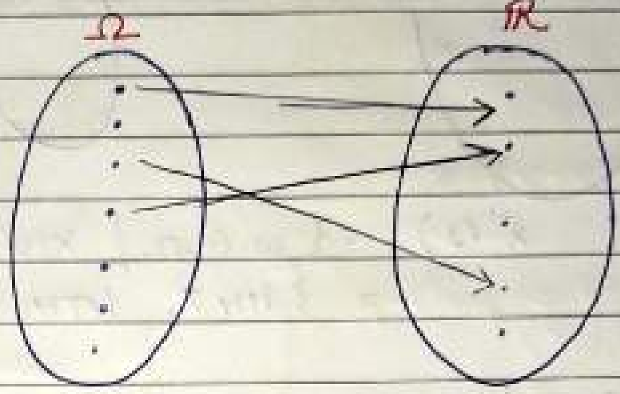
B : sum greater than six

$P(A|B) = \frac{1}{21}$

Scanned by CamScanner

Random Variable:

A random variable X is a function from Ω to $\mathbb{R}$ if $\{\omega : X(\omega) \in \mathbb{R}\} \in \mathcal{F}$



$\forall \ x \in \mathbb{R}, \{\omega : X(\omega) = x\} = X^{-1}(x)$
$\subseteq \Omega$ inverse image of X.

$\{\omega \,/\, X(\omega) = x\} \in \mathcal{F}$

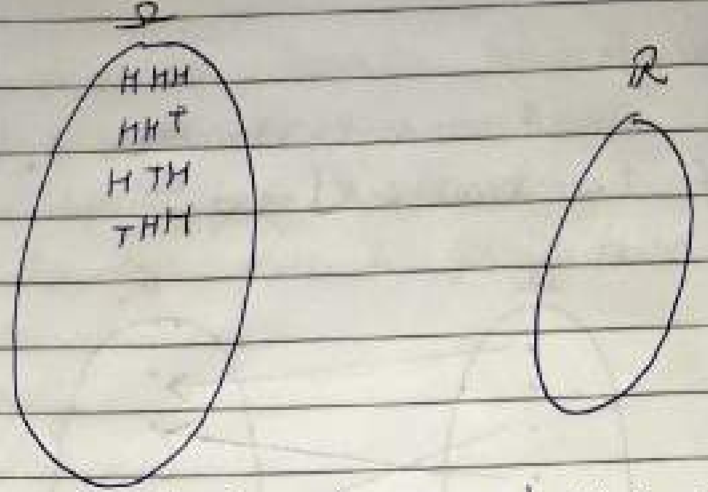
Experiment: 3 fair coins tossed independtly.

$\Omega = \{HHH, HHT, HTH, THH, TTH, THT, TTH, TTT\}$
$\mathcal{F} = 2^{\Omega}$

$P(HHH) = \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{8}$

X: no. of heads in 3 tosses.

Scanned by CamScanner



if $x = 2$ $\quad X^{-1}(2) = \{\omega \in \Omega \mid X(\omega) = 2\}$

$= \{HHT, HTH, THH\}$

$$P(X = 2) = P\{\omega \in \Omega \mid X(\omega) = 2\} = P\{HHT, HTH, THH\} = \frac{3}{8}$$

$x = 8.7$

$P(X = x) = P(X = 8.7) = 0$

* Range of X : $\{x \in \mathbb{R} \mid X(x) > 0\}$
* A random variable X is called a discrete random variable if range of x is finite or countably infinite.

In the previous example $R_x = \{0, 1, 2, 3\}$ and hence X is a discrete random variable.

Definition:

A real valued function $f(x) = P(X = x)$ is called as discrete density or probability mass function (pmf) of the discrete r.v. X.

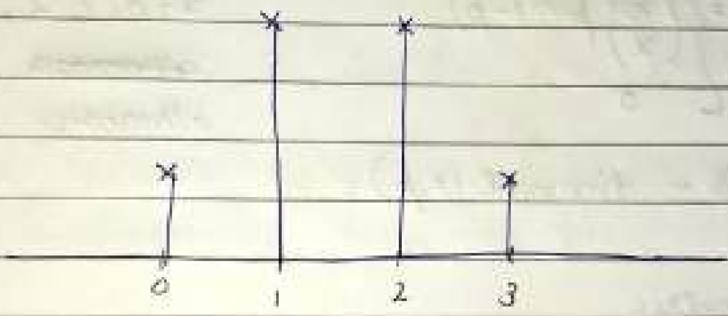
Scanned by CamScanner

$f(x) = \frac{1}{8}$ $\quad x = 0$

$f(x) = \frac{3}{8}$ $\quad x = 1$

$f(x) = \frac{3}{8}$ $\quad x = 2$

$f(x) = \frac{1}{8}$ $\quad x = 3$



x	0	1	2	3
$f(x)$	1/8	3/8	3/8	1/8

← tabular representation of pmf

Ex: Bernouli

Expt: Tossing a coin

$\Omega = \{H, T\}$ $\quad P(H) = p$; $p(T) = 1-p$

$X\{Heads\} = 1$ $\quad X\{Tails\} = 0$

$P(X=0) = 1-p$; $\quad p(X=1) = p$

$$f(x) = \left.\begin{cases} 1-p ; & x=0 \\ p ; & x=1 \\ 0 ; & \text{otherwise} \end{cases}\right\} \text{ p.m.f}$$

$X \sim$ Bernouli(p) → parameter

Ex: Binomial

Expt: n independent Bernouli trials are performed
X: number of success in n-trials
$R_x = \{0, 1, 2, \ldots, n\}$

$$f(x) = P(X = x)$$

$$f(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & x = 0, 1, 2, \ldots n \\ 0 & \text{otherwise} \end{cases}$$

$X \sim \text{Binomial}(n, p)$

Ex: Geometric
Expt: Perform independent Bernouli trials (p) until the first success

$$\Omega = \{ \underset{p}{S},\ \underset{p(1-p)}{FS},\ \underset{(1-p)^2 p}{FFS},\ \underset{(1-p)^3 p}{FFFS}, \ldots \}$$

X: The no. of failures before first success.

$R_x = \{0, 1, 2, \ldots\}$ discrete random variable

$$f(x) = \begin{cases} p(1-p)^x & x = 0, 1, 2, \ldots \\ = 0 & \text{otherwise} \end{cases}$$

* **Note:** that for any random experiment $P(\Omega)=1$ Let X be a random variable defined on Ω. Let R_x be the range of this random variable.

$$R_x = \{x_1, x_2, \ldots, x_n, \ldots\}$$

(discrete random variable case)

$$\sum_{x_i \in R_x} f(x_i) = \sum_{x_i \in R_x} P(X = x_i)$$

$$= \sum_{x_i \in R_x} P\{\omega : X(\omega) = x_i\}$$

$$= 1$$

Properties of pmf

i) $f(x) \geq 0 \quad \forall x \in \mathbb{R}$

ii) $\{x : f(x) > 0\}$ is a finite or countable set $= \{x_1, x_2, \ldots, x_j, \ldots\}$

iii) $\sum_{x_i} f(x_i) = 1$

Any function f which satisfies these three properties is called as discrete density or pmf.

Ex: Uniform density

$$f(x) = \begin{cases} \frac{1}{5} & x \in \{x_1, x_2, \ldots, x_5\} \\ 0 & \text{otherwise} \end{cases}$$

Scanned by CamScanner

Ex: Geometric pmf:

$$f(x) = p(1-p)^x \qquad x = 0, 1, 2, 3, \ldots$$
$$\quad 0 \qquad \text{otherwise}$$

Ex: Negative binomial distribution:

$$f(x) = p^{\alpha} \binom{\alpha+x-1}{x} (1-p)^x \qquad x = 0, 1, 2 \ldots$$
$$= 0 \qquad \text{otherwise}$$

(expt: performing independent Bernoulli(p) trials until α no. of successes are observed.

For $\alpha = 2$.

$\{HH, HTH, THH, \ldots\}$

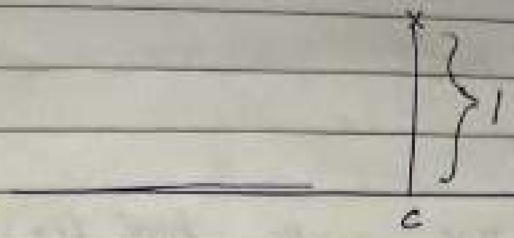
Ex: Poisson density(λ)

$$f(x) = \frac{e^{-\lambda} \cdot \lambda^x}{\lfloor x} \qquad x = 0, 1, 2, 3, \ldots$$
$$= 0 \qquad \text{otherwise}$$

Special examples of r.v.s

Ex: constant random variable

$$X(\omega) = c \qquad \forall\ \omega \in \Omega$$
$$P(X = c) = 1 \quad , \quad P(X \neq c) = 0$$



Ex: Indicator random variable. Let $A \in \mathcal{F}$

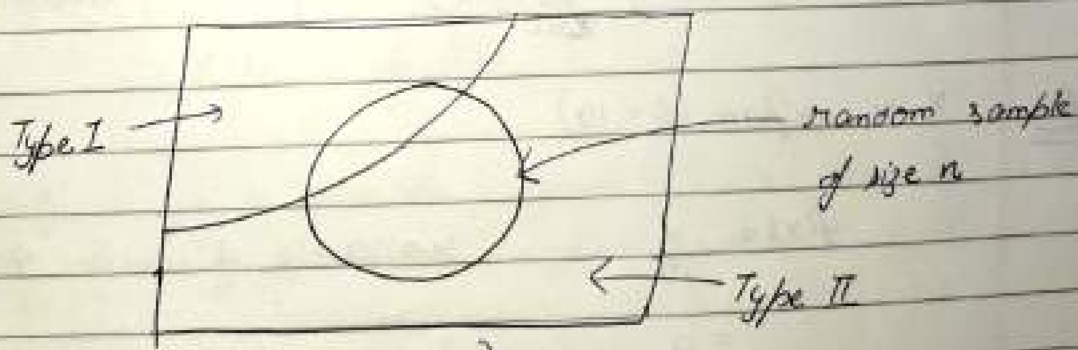
$$X_A(\omega) = 1 \quad \forall\, \omega \in A$$
$$= 0 \quad \forall\, \omega \in A^c$$

Ex: Hypergeometric distribution:

Population of r objects
Type I objects r_1
Type II objects $r - r_1 = r_2$

Let a sample of size n is chosen from this population $(n \leq r)$

X : number of objects of Type I in the random sample.



$$P(X=x) = \frac{\binom{r_1}{x}\binom{r-r_1}{n-x}}{\binom{r}{n}} \qquad x = 0, 1, 2, \ldots, n$$

$$= 0 \qquad \text{otherwise}$$

Scanned by CamScanner

Computations with pmf:

$(\Omega, \mathcal{F}, P)$

X: discrete r.v. with pmf $f(x)$

Event: $x \leq t$ for some real no. t

$= \{\omega : X(\omega) \leq t\}$

$= \bigcup_{i=-\infty}^{t} \{\omega : X(\omega) = i\}$

$$P(X \leq t) = P\left(\bigcup_{i=-\infty}^{t} \{\omega : X(\omega) = i\}\right) = \sum_{i=-\infty}^{[t]} P(X = i)$$

Cumulative distribution function:

Discrete random variable x with pmf $f(x)$.

For every real number $t \in \mathbb{R}$, define

$$\left.\begin{aligned} F(t) &= P(X \leq t) \\ &= \sum_{x \leq t} f(x) \end{aligned}\right\} \text{cumulative distribution function}$$

Ex: $X \sim$ uniform $(s = 10)$

$f(x) = \frac{1}{10}$ $\quad x = 0, 1, 2, 3, \ldots, 9$

$= 0$ $\quad$ otherwise

Clearly for $t < 0$, $F(t) = 0$

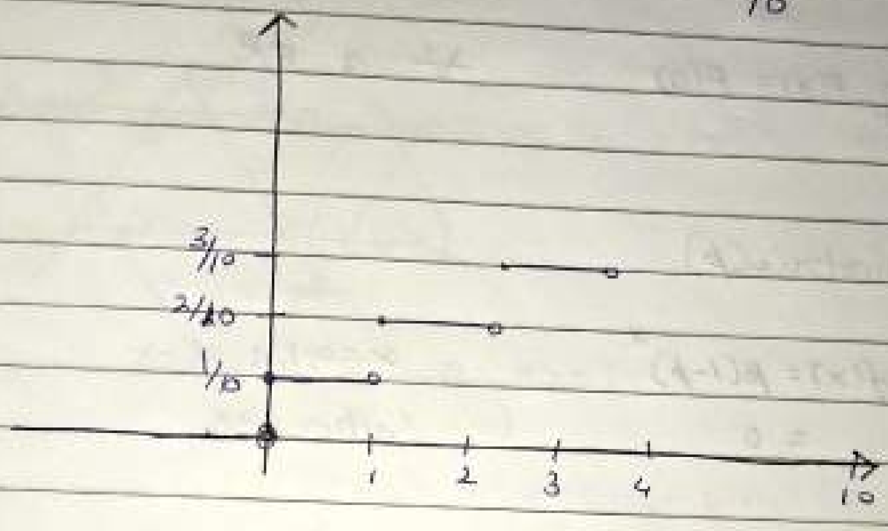
$F(0) = f(0) = \frac{1}{10}$

Scanned by CamScanner

for any $0 \le t < 1 \Rightarrow F(t) = \frac{1}{10}$

$$F(1) = f(0) + f(1) = \frac{2}{10}$$

for any $1 \le t < 2$; $F(t) = \frac{2}{10}$



CDF: {Cumulative Distribution Function}

Let X be discrete r.v. with pmf $f(x)$. For any $t \in \mathbb{R}$ we define

$$F(t) = \sum_{x \le t} f(x)$$

Clearly, F is a function from $\mathbb{R}$ to $\mathbb{R}$. • F is a non-decreasing function. for $t_1 < t_2$

$$\Rightarrow F(t_1) \le F(t_2)$$

Scanned by CamScanner

- The graph of F is like a staircase with jumps occuring exactly at the points which are in the range of X.

- Right continuity of F

$$\lim_{x \to a^+} F(x) = F(a) \qquad \forall\ a \in \mathbb{R}$$

Ex: $X \sim$ geometric(p)

$$f(x) = p(1-p)^x \qquad x = 0, 1, 2, 3, \dots$$
$$= 0 \qquad \text{otherwise}$$

$$F(t) = 0 \qquad \forall\ t < 0$$

$$F(t) = \sum_{x=0}^{[t]} p(1-p)^x \qquad t \geq 0$$

$$= p\left[\frac{1-(1-p)^{[t]+1}}{1-(1-p)}\right] = \cancel{p}\left[\frac{1-(1-p)^{[t]+1}}{\cancel{1}-\cancel{1}+\cancel{p}}\right]$$
$$= 1-(1-p)^{[t]+1}$$

For this geometric r.v.

$$P(X \geq x) = 1 - P(X < x)$$
$$= 1 - P(X \leq x-1)$$
$$= 1 - F(x-1)$$
$$= 1 - [1-(1-p)^{x-1+1}]$$

$$P(X \geq x) = (1-p)^x$$

Events

$X < x$ and $X \geq x$

$X < x \cup X \geq x = \Omega$

disjoint

$\{\omega : X(\omega) < x\}$

$\{\omega : X(\omega) \geq x\}$

Scanned by CamScanner

Prove:

$$P(X > n+m / X > n) = P(X > m)$$

$\longrightarrow$ Memoryless property

~~For this geometric~~

$$\cancel{P(X \geq x) = 1 - P(X < x)}$$

$$\cancel{1 - P(X}$$

$$\frac{P(X > n+m \cap X > n)}{P(X > n)}$$

$$= \frac{P(X > (n+m))}{P(X > n)}$$

H, TH, TTH, TTTH, . . .

CD.F. $\quad F(x)$

X is a random variable $(\Omega, \mathcal{F}, P)$
Probability of the event $X \leq x$

$$F(x) = P(X \leq x) \quad \forall \; x \in \mathbb{R}$$

This function F is called as cumulative distribution function $F(x)$ of random variable X. $(F_X(x))$

Scanned by CamScanner

Properties of CDF:

1. $0 \leq F_X(x) \leq 1 \quad \forall\ x \in \mathbb{R}$

2. $\lim_{x \to +\infty} F_X(x) = 1$

3. $\lim_{x \to -\infty} F_X(x) = 0$

4. The function $F_X(x)$ is non-decreasing for $x_1 \leq x_2$
$\Rightarrow$

5. Right continuity

$$\lim_{x \to a^+} F_X(x) = F_X(a) \quad \forall\ a \in \mathbb{R}$$

for every $a \in \mathbb{R}$ and $\delta > 0$,

$$\lim_{\delta \to 0} \left(F_X(a+\delta) - F_X(a) \right) = 0$$

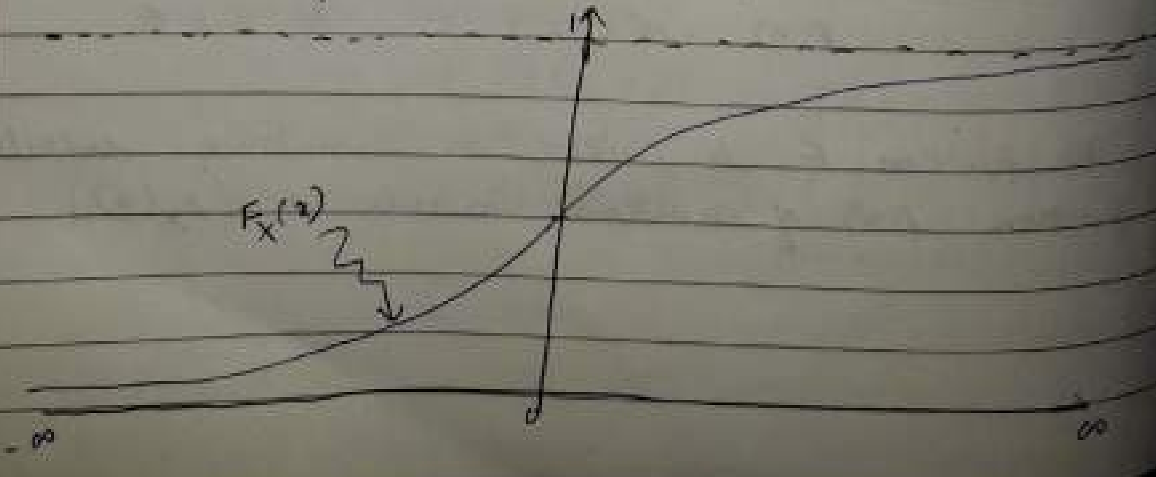
Lebesgue Decomposition Theorem:

If a function $F_X(x)$ satisfies the properties stated above, then $F_X(x)$ can be represented as sum of two functions say $G_X(x)$ and $H_X(x)$ as:

$$F_X(x) = G_X(x) + H_X(x)$$

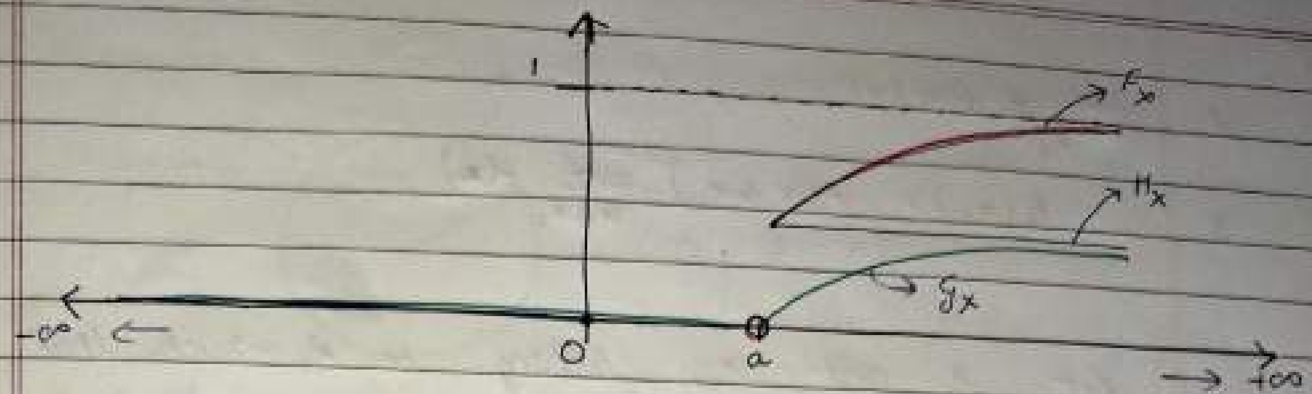
where $G_X(x)$ is continuous and $H_X(x)$ is a right continuous step function with jumps coinciding with those of $F_X(x)$ and $H_X(-\infty) = 0$.

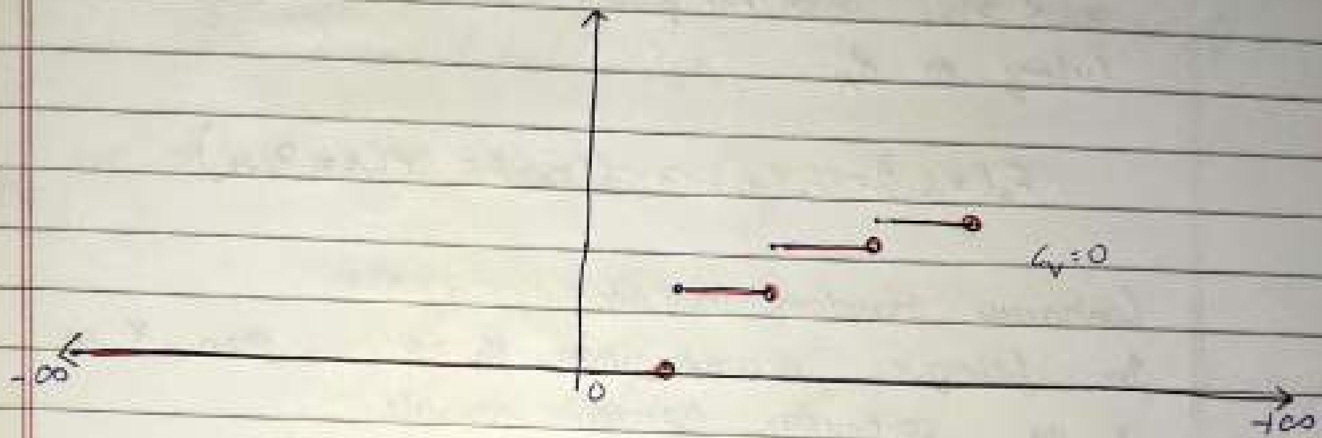
Ex:



Scanned by CamScanner

classmate
Date
Page





In Lebesgue decomposition if $G_x(x) = 0$ (indentically), then the random variable X is called as discrete random variable.

If $H_x(x) = 0$ (identically), then X is called as continuous random variable.

The situation where neither of these functions are initially equal to zero, we call X a ~~var~~ mixed random variable.

Definitions/ Observations:

If X is a discrete random variable

1. $P(x_i) = f(x_i) = P(X = x_i) > 0 \quad \forall \; x_i \in R_x$

Scanned by CamScanner

2. $\sum_{x_i} f(x_i) = 1$

3. $F_X(x_i) = P(X \le x_i) = \sum_{x \le x_i} f(x)$

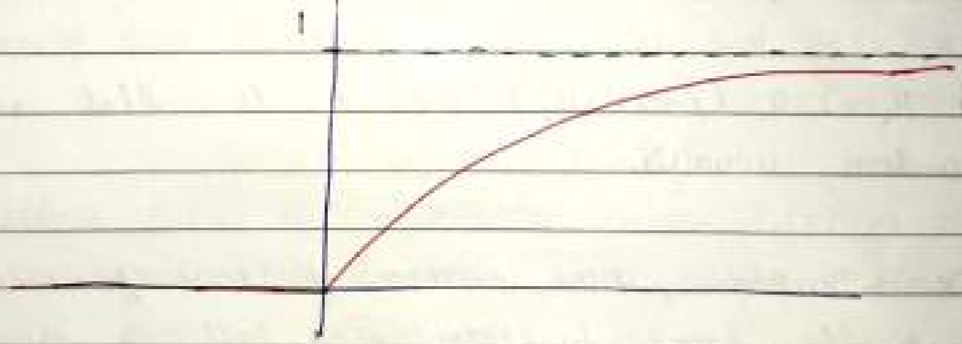
Let x_i and x_{i+1} belong to R_X such that $x_i < x_{i+1}$ and no point between x_i & x_{i+1} belong to R_X.

$$F_X(x_{i+1}) - F_X(x_i) = f(x_{i+1}) = P(X = x_{i+1})$$

Continuous random variable:

In Lebesgue decomposition $H_X = 0$, then X is x as continuous random variable.

$$P(X=x) = \lim_{\delta \to 0} [F_X(x+\delta) - F_X(x)]$$

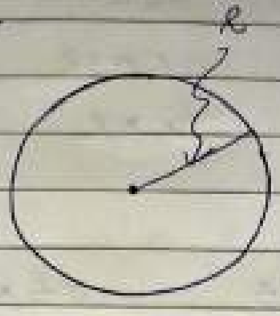


define $f_X(x) = \frac{d}{dx} F_X(x)$

(always exists except for a "few" points.) in the above example at $x=0$, $f_X(0)$ does not exist

$f_X(x)$: probability density function

Example:



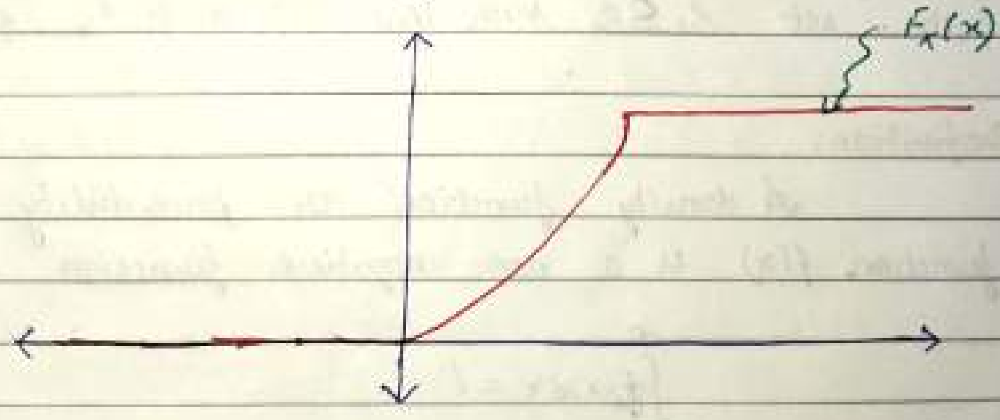
Expt: Throw dart on this circular board.
$(\Omega, \mathcal{F}, P)$
uniform probability space

X: distance of the dart from the center of the board (bull's eye)

Let $F_X(x)$ denote the CDF of X.

$$F_X(x) = \text{Prob}(X \le x)$$
$$= \frac{\pi x^2}{\pi R^2} = \frac{x^2}{R^2}$$ uniformity principle

$$F_X(x) = \begin{cases} 0 & x < 0 \\ x^2/R^2 & 0 \le x \le R \\ 1 & x > R \end{cases}$$



The probability density function

$$f_X(x) = \frac{d}{dx} F_X(x)$$

Scanned by CamScanner

$$= \begin{cases} 0 & x < 0 \\ \frac{2x}{R^2} & 0 \le x \le R \\ 0 & x > R \end{cases}$$

$$f_X(x) = \begin{cases} \frac{2x}{R^2} & 0 \le x \le R \\ 0 & \text{otherwise} \end{cases}$$

$$P(a \le x \le b) = F_X(b) - F_X(a) = \int_a^b f_X(x)\,dx$$

$$= \int_{-\infty}^{b} f_X(x)\,dx - \int_{-\infty}^{a} f_X(x)\,dx$$

$$= \int_a^b f_X(x)\,dx$$

Let X be a continuous random variable with CDF $F_X(x)$ and $f_X(x) = \frac{d}{dx} F_X(x)$ be its probability density function (pdf). Then the range of X is the set $R_X \subseteq R$ such that $\forall\ x \in R_X,\ f(x) > 0$.

Definition:

A density function or probability density function $f(x)$ is a non-negative function such that

$$\int_{-\infty}^{\infty} f_X(x)\,dx = 1$$

Then obviously,

$$F_X(x) = \int_{-\infty}^{x} f_X(x)\,dx$$

satisfies all the properties of CDF.

Scanned by CamScanner

$$P(a \le X \le b) = \int_a^b f_X(x)\,dx = F_X(b) - F_X(a)$$

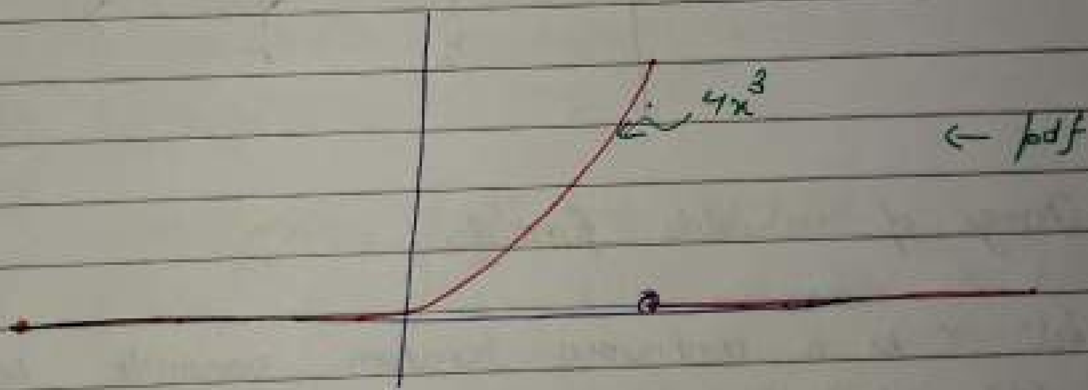
Example:

i) for what value k, $f(x) = kx^3$

$$\int_0^1 kx^3\,dx = 1$$

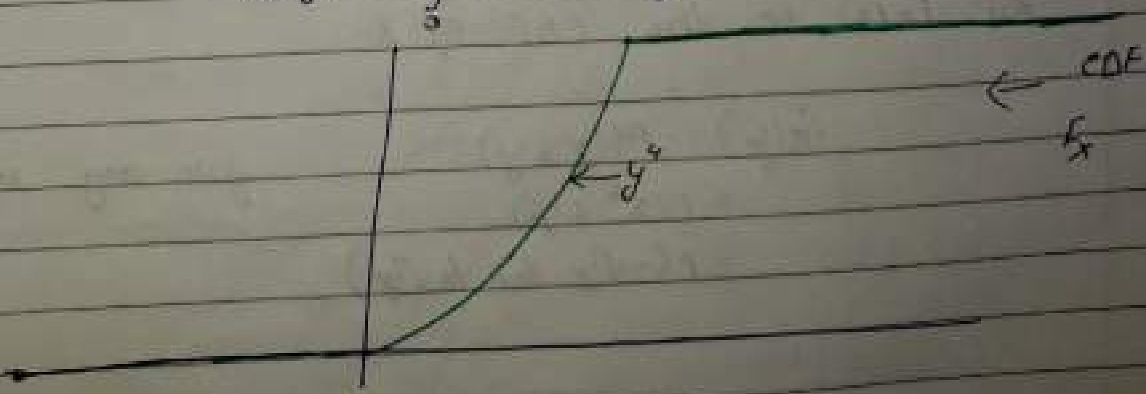
$$\left|\frac{kx^4}{4}\right|_0^1 = 1$$

$$\boxed{k=4}$$



For $y \in (0,1)$

$$F_X(y) = \int_0^y 4x^3\,dx = y^4$$

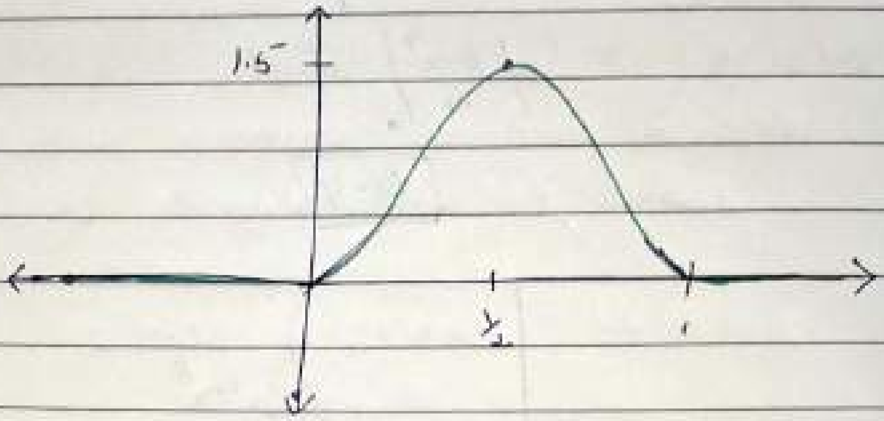


Scanned by CamScanner

Ex:

$$f(x) = \begin{cases} 6x(1-x) & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

$$\int_0^1 x - x^2 \, dx = \left. \frac{x^2}{2} - \frac{x^3}{3} \right|_0^1 = \frac{1}{6}$$



Change of variable formula

Let X be a continuous random variable with pdf $f_X(x)$. Find the density of the random variable $Y = X^2$.

Let $G_Y(y)$ be the CDF of Y.
Let $F_X(x)$ be the CDF of X.

$$\begin{aligned} G(y) &= P(Y \le y) \\ &= P(X^2 \le y) \\ &= P(-\sqrt{y} \le X \le \sqrt{y}) \end{aligned}$$

for any real no. $y \in \mathbb{R}$

Scanned by CamScanner

$$G_Y(y) = F_X(\sqrt{y}) - F_X(-\sqrt{y})$$

In order to find density of Y, say g(y), we take derivative of $G_Y(y)$.

$$g(y) = \frac{d}{dy} G_Y(y)$$

$$= \frac{d}{dy}\left[F_X(\sqrt{y}) - F_X(-\sqrt{y}) \right]$$

$$= \frac{1}{2\sqrt{y}}\left[F_X'(\sqrt{y}) + F_X'(-\sqrt{y}) \right]$$

$$g(y) = \frac{1}{2\sqrt{y}}\left[f_X(\sqrt{y}) + f_X(-\sqrt{y}) \right]$$

Ex:

$$f_X(x) = \begin{cases} \frac{2x}{R^2} & 0 \le x \le R \\ 0 & \text{otherwise} \end{cases}$$

Find the density of $Y = X^2$.

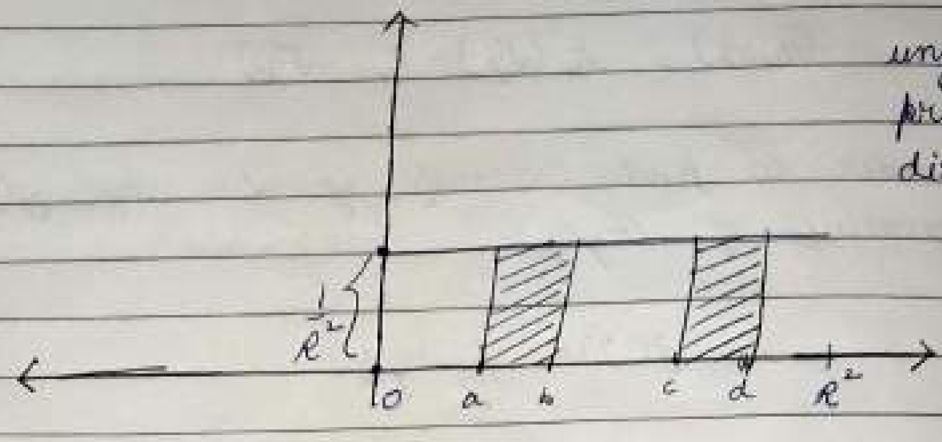
Let g(y) be the density of Y.

$$g(y) = \frac{1}{2\sqrt{y}}\left[f_X(\sqrt{y}) + f_X(-\sqrt{y}) \right]$$

$$= \frac{1}{2\sqrt{y}}\left[\frac{2\sqrt{y}}{R^2} + 0 \right] \qquad 0 \le y \le R^2$$

$$g(y) = \begin{cases} \frac{1}{R^2} & 0 \le y \le R^2 \\ = 0 & \text{otherwise} \end{cases}$$

Scanned by CamScanner



uniform probability distribution

Continuous uniform random variable.

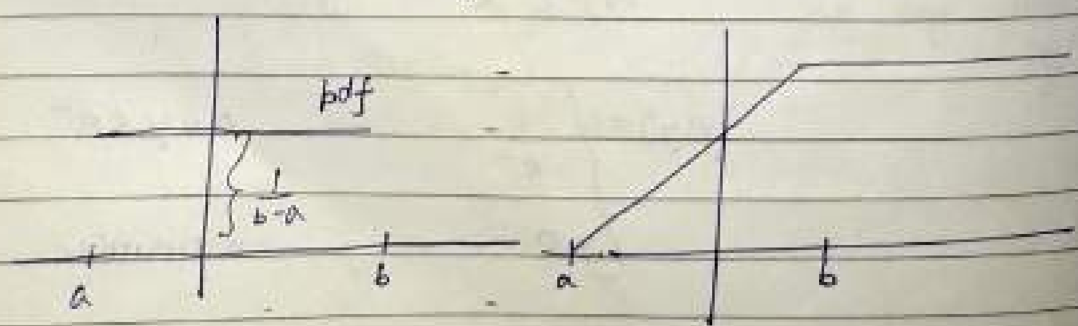
Let X be a continuous random variable with the probability distribution function defined as

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

for any two $a, b \in \mathbb{R}$ such that $a < b$, then X is said to follow uniform density with parameters a & b. $X \sim U(a, b)$

$$F_X(x) = \int_{-\infty}^{x} f_X(x)\, dx = \frac{x-a}{b-a} \qquad a < x \le b$$

$$F_X(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \le x \le b \\ 1 & x \ge b \end{cases}$$



classmate
Date
Page

Ex: Let $X \sim U(0,1)$

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

$$F_X(x) = \begin{cases} 0 & x < 0 \\ x & 0 \le x \le 1 \\ 1 & x \ge 1 \end{cases}$$

Consider the transformation

$$Y = \frac{-1}{\lambda} \log(1-X) \quad \text{for} \quad X > 0$$

Let $G_Y(y)$ be the CDF of Y and $g(y)$ be the pdf of Y

$$G_Y(y) = P(Y \le y) \qquad \forall\, y \in \mathbb{R}$$

$$P\left(\frac{-1}{\lambda} \log(1-X) \le y\right)$$

$$= P\left(\log(1-X) \ge -\lambda y\right)$$

$$= P\left(1-X \ge e^{-\lambda y}\right)$$

$$= P\left(X \le 1-e^{-\lambda y}\right)$$

$$G_Y(y) = 1 - e^{-\lambda y}$$

$$g(y) = \frac{d}{dy} G_Y(y) = \frac{d}{dy}\left(1 - e^{-\lambda y}\right)$$

$$\boxed{g(y) = \begin{cases} \lambda e^{-\lambda y} & y > 0 \\ 0 & y \le 0 \end{cases}}$$

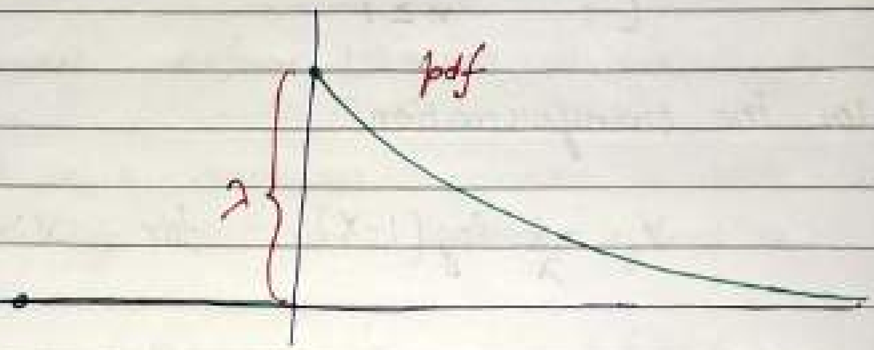
Scanned by CamScanner

This is an exponential density with parameter $\lambda (>0)$

Exercise:

Let $X \sim \exp(\lambda)$

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x > 0 \\ 0 & \text{otherwise} \end{cases}$$



$$F_X(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-\lambda x} & x \geq 0 \end{cases}$$

For any real no.s $x, y > 0$

$$P(X > x+y) = 1 - F_X(x+y)$$
$$= 1 - (1 - e^{-\lambda(x+y)})$$
$$= e^{-\lambda(x+y)} = e^{-\lambda x} \cdot e^{-\lambda y}$$

$$P(X > x+y) = e^{-\lambda x} \cdot e^{-\lambda y}$$

$$P(X > x+y) = P(X > x) \cdot P(X > y)$$

$$\frac{P(X > x+y)}{P(X > y)} = P(X > x)$$

Scanned by CamScanner

(20)

$$\boxed{P(X > x+y \mid X > y) = P(X > x)}$$

Memoryless property of exponential density

<u>Theorem:</u> Let ϕ be a differential function which is strictly increasing or strictly decreasing on an interval I. Let $\phi(I)$ denote the range of ϕ and ϕ^{-1} be the inverse of ϕ on I. Let X be a continuous random variable having density $f_X(x)$ such that $f_X(x) \neq 0$ on I.

then $Y = \phi(X)$ whose density is given by

$$g(y) = f(\phi^{-1}(y)) \left| \frac{d}{dy} \phi^{-1}(y) \right| \qquad y \in \phi(I)$$

<u>Proof:</u>

Let $G(y)$ be the CDF of y.

$$\left.\begin{aligned} G(y) &= P(Y \le y) \\ &= P(\phi(X) \le y) \\ &= P(X \le \phi^{-1}(y)) \end{aligned}\right\} \phi \text{ is increasing}$$

$$G(y) = F_X(\phi^{-1}(y))$$

$$\Rightarrow \quad g(y) = F_X'(\phi^{-1}(y)) \frac{d}{dy} \phi^{-1}(y)$$

$$\boxed{g(y) = f_X(\phi^{-1}(y)) \frac{d}{dy} (\phi^{-1}(y))}$$

Scanned by CamScanner

ϕ is a decreasing function:

$$\left.\begin{aligned} G(y) &= P(Y \le y) \\ &= P(\phi(X) \le y) \\ &= P(X \ge \phi^{-1}(y)) \end{aligned}\right\} \quad \phi \downarrow$$

$$G(y) = 1 - F_X(\phi^{-1}(y))$$

$$\boxed{g(y) = -f_X(\phi^{-1}(y)) \cdot \frac{d}{dy}(\phi^{-1}(y))}$$

Ex:

Let X be a continuous random variable with density f. Let $a, b \in \mathbb{R}$ and $b \neq 0$. Then define $Y = a + bX$. then what is the density of Y.

$$g(y) = f(\phi^{-1}(y)) \left| \frac{d}{dy} \phi^{-1}(y) \right|$$

$$\phi(x) = a + bx = y$$

$$x = \frac{y-a}{b} = \phi^{-1}(y)$$

$$g(y) = f\left(\frac{y-a}{b}\right) \cdot \frac{1}{|b|}$$

Ex:

$$f_X(x) = \begin{cases} \frac{2x}{R^2} & 0 < x < R \\ 0 & \text{otherwise} \end{cases}$$

$$Y = \frac{X}{R} \qquad 0 < y < 1 = R_Y$$

$$g(y) = \frac{1}{|b|} f\left(\frac{y-a}{b}\right) \quad \text{Here} \quad a=0,\ b=\frac{1}{R}$$

$$= \frac{1}{1/R} \cdot 2\left(\frac{\frac{y}{1/R}}{R^2}\right) = 2y$$

$$g(y) = \begin{cases} 2y & 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$$

Symmetric densities:

f is symmetric density if $f(x) = f(-x)$ $\forall$ $x \in \mathbb{R}$

$U(-a, a)$ is symmetric.

A random variable is symmetric if its pdf $f_X(x)$ is a symmetric density.

Ex: If X is a symmetric random variable with CDF $F_X(x)$. Then $F_X(0) = 1/2$.

$$F_X(-x) = \int_{-\infty}^{-x} f(y)\,dy = \int_{x}^{\infty} f(-y)\,dy$$

$$= \int_{x}^{\infty} f(y)\,dy$$

$$= 1 - F_X(x)$$

$$F_X(-x) = 1 - F_X(x) \quad \forall\ x \in \mathbb{R}$$

Scanned by CamScanner

Ex:

$$g(x) = \frac{1}{1+x^2} \qquad -\infty < x < \infty$$

Is $g(x)$ a pdf ??

$$\int_{-\infty}^{\infty} \frac{1}{1+x^2}\, dx = \tan^{-1} x \Big|_{-\infty}^{\infty} = \frac{\pi}{2} - \left(-\frac{\pi}{2}\right) = \pi$$

$$\boxed{g(x) = \frac{1}{\pi(1+x^2)} \qquad -\infty < x < \infty}$$

is a density   Cauchy density

Since $g(x) = g(-x)$; cauchy density is symmetric

Ex: $g(x) = e^{-x^2/2} \qquad -\infty < x < \infty$

$$c = \int_{-\infty}^{\infty} e^{-x^2/2}\, dx$$

$$c^2 = \int_{-\infty}^{\infty} e^{-x^2/2}\, dx \cdot \int_{-\infty}^{\infty} e^{-y^2/2}\, dy$$

$$c^2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(x^2+y^2)/2}\, dy\, dx$$

Polar co-ordinate substitution:

$$x = r\cos\theta \qquad y = r\sin\theta$$

$$c^2 = \int_0^{\infty}\int_0^{2\pi} r \cdot e^{-r^2/2}\, d\theta\, dr$$

Scanned by CamScanner

$$c^2 = 2\pi$$
$$c = \sqrt{2\pi}$$

$$\int_{-\infty}^{\infty} e^{-x^2/2}\, dx = \sqrt{2\pi}$$

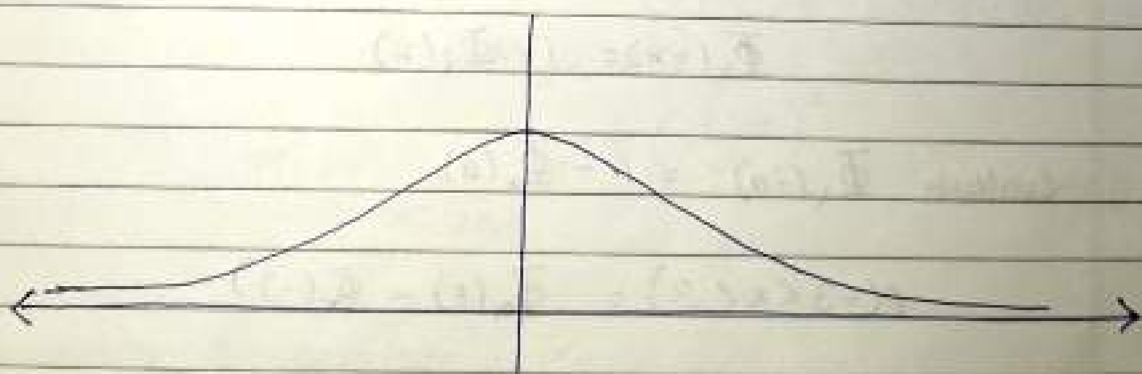
Define:

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \qquad -\infty < x < \infty$$

Clearly this $\phi(x)$ is a pdf. this density is called as standard normal density.

The random variable corresponding to this density, X, is called as standard normal random variable.

Standard Normal Density $\phi(x)$ is clearly symmetric



$$R_X = \mathbb{R} \quad \text{as} \quad \phi(x) > 0 \quad \forall \quad x \in \mathbb{R}$$

Gaussian random Distribution.

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

Scanned by CamScanner

Let Φ_x denote the CDF of std normal random variable.

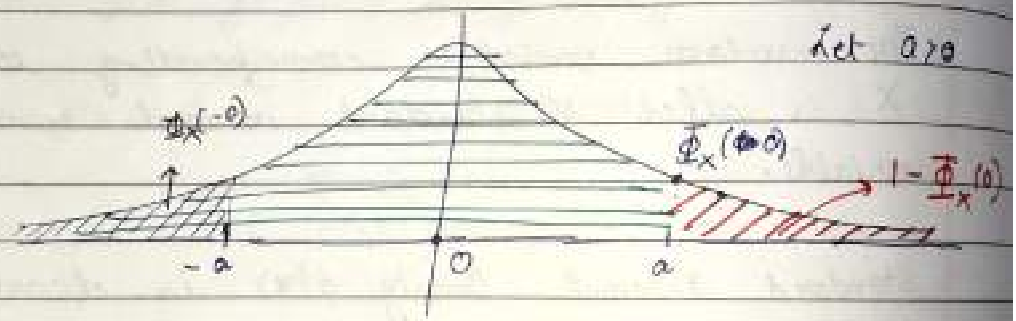
$$P(a \le X \le b) = \Phi_x(b) - \Phi_x(a)$$

$$= \int_a^b \phi(x)\,dx = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\,dx$$

It seems it is enough to know $\Phi_x(a)\ \forall a \in \mathbb{R}$

$$\Phi_x(x) = \int_{-\infty}^{x} \phi(x)\,dx$$

Let $a>0$



$$\Phi_x(-x) = 1 - \Phi_x(x)$$

Evaluate $\Phi_x(-a) := 1 - \Phi_x(a)$

$$P(-3 < X < 3) = \Phi_x(3) - \Phi_x(-3)$$

$$= \Phi_x(3) - [1 - \Phi_x(3)]$$

$$= 2\Phi_x(3) - 1 = 2(0.9986) - 1$$

$$P(-3 < X < 3) = 0.9973$$

Scanned by CamScanner

$$P(-0.5 < X < -0.25) = \Phi_X(0.25) - \Phi_X(-0.5)$$
$$=$$
$$= \Phi_X(0.25) - 1 + \Phi_X(0.5)$$
$$= -\Phi_X(1/4) + \Phi_X(1/2)$$
$$= 0.0928$$

Let X be a standard normal random variable. Define $Y = \mu + \sigma X$ where $\sigma > 0$. Find density of Y.

Let $g(y)$ be the density

$$g(y) = \frac{1}{\sigma}\phi\left(\frac{y-\mu}{\sigma}\right)$$
$$= \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y-\mu)^2}{2\sigma^2}} \qquad -\infty < y < \infty$$

This Y is said to follow normal density with two parameters μ & σ^2

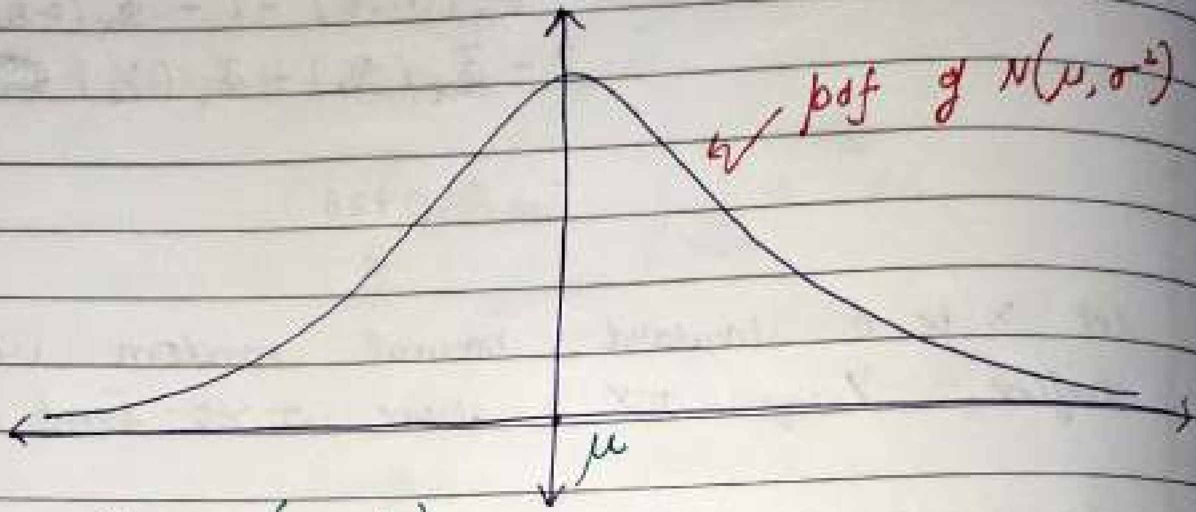
$$Y \sim N(\mu, \sigma^2)$$
$$g(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2} \qquad -\infty < y < \infty$$

Probability Computation

$$P(a < Y < b) = P(a < \mu + \sigma X < b)$$
$$= P(a - \mu < \sigma X < b - \mu)$$
$$= P\left(\frac{a-\mu}{\sigma} < X < \frac{b-\mu}{\sigma}\right)$$

Scanned by CamScanner

$$= \Phi_X\left(\frac{b-\mu}{\sigma}\right) - \Phi_X\left(\frac{a-\mu}{\sigma}\right)$$



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

$P(\mu - 3\sigma < Y < \mu + 3\sigma)$

$= P(\mu - 3\sigma - \mu < Y - \mu < \mu + 3\sigma - \mu)$

$= P\left(\frac{-3\sigma}{\sigma} < \frac{Y-\mu}{\sigma} < \frac{3\sigma}{\sigma}\right)$

$= P(-3 < Y < 3) = \Phi_X(3) - \Phi_X(-3) = 0.9970$

Scanned by CamScanner

classmate
Date
Page

Gamma Densities:

Let $X \sim N(0, \sigma^2)$
$Y = X^2$

Find the density of Y

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}} \qquad -\infty < x < \infty$$

Let $g(y)$ be the density of Y

$$g(y) = \frac{1}{2\sqrt{y}} \left(f(\sqrt{y}) + f(-\sqrt{y}) \right) \qquad y > 0$$

$$g(y) = \frac{1}{\sigma\sqrt{2\pi y}} e^{-y/2\sigma^2} \qquad y > 0$$

Note: $g(y)$ is a Gamma density.

Gamma Functions:

$$\int_0^\infty x^{\alpha-1} e^{-\lambda x}\, dx = \frac{\Gamma(\alpha)}{\lambda^\alpha}$$

Gamma pdf:

$$\Gamma(x = \alpha, \lambda) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} & x > 0 \\ 0 & x \le 0 \end{cases}$$

parameter of this density are α, λ.

Scanned by CamScanner

take $\alpha = \frac{1}{2}$, $\lambda = \frac{1}{2\sigma^2}$

Expectation of a random variable:

$$\begin{array}{ccc} & \text{r.v.s} & \\ \swarrow & & \searrow \\ \text{discrete} & & \text{continuous} \end{array}$$

- R_x countably finite/infinite
- p.m.f.

- p.d.f.

Definition: Let X be a discrete random variable with R_x as its image

$$E(X) = \sum_{x \in R_x} x\, P(X = x)$$

{provided that the sum is convergent}

Ex: Let X be a discrete uniform random variable on $\{x_1, x_2, \ldots, x_n\} = R_x$

$$p(X = x_i) = \frac{1}{n} \quad \text{for } x_i \in R_x$$

$$= 0 \quad \text{otherwise}$$

$$E(X) = \sum x_i\, p(X = x_i) = \sum_{x_i} x_i\, p(x_i) = \frac{\sum_{i=1}^{n} x_i}{n}$$

Scanned by CamScanner

In case of uniform discrete random variable $E(X)$ is nothing but the A.M. of R_X.

In general, we can ~~reason~~ understand $E(X)$ as weighted average.

Ex: $X \sim \text{Bernoulli}(p)$

x	0	1
$p(x)$	$1-p$	p

← pmf of X in tabular form

$$E(X) = 0\cdot(1-p) + 1\cdot p = p$$

Ex: $X \sim \text{Binomial}(n, p)$

$$p(X=x) = \binom{n}{x} p^x (1-p)^{n-x} \qquad x = 0, 1, 2, \ldots n$$

$$= 0 \qquad \text{otherwise}$$

$$E(X) = \sum_{j=0}^{n} j \binom{n}{j} p^j (1-p)^{n-j}$$

notice:

$$j\binom{n}{j} = \frac{j \times n!}{j! \times (n-j)!} = n\binom{n-1}{j-1}$$

$$E(X) = np$$

Ex: $X \sim \text{poisson}(\lambda)$

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

$$= 0 \qquad \text{otherwise}$$

$$E(X) = \sum_{j=0}^{\infty} j \frac{e^{-\lambda}\lambda^{j}}{j!}$$

$$= \lambda$$

Ex: $X \sim \text{Geometric}(p)$

$$p(X=x) = (1-p)^{x} p \qquad x = 0, 1, 2, 3, 4, \ldots$$

$$= 0 \qquad \text{otherwise}$$

$$E(X) = \sum_{j=0}^{\infty} j\, p(1-p)^{j}$$

$$= p(1-p) \sum_{j=0}^{\infty} j (1-p)^{j-1}$$

Interchanging Σ and $\frac{d}{dp}$, because the series is absolutely summable (accept)

$$E(X) = -p(1-p) \frac{d}{dp} \sum_{j=0}^{\infty} (1-p)^{j}$$

$$E(X) = \frac{1-p}{p}$$

Scanned by CamScanner

Example:

Let X be a discrete r.v. with the p.m.f.

$$f(x) = \begin{cases} \frac{1}{x(1+x)} & x = 1, 2, \ldots \\ 0 & \text{otherwise} \end{cases}$$

clearly, $f(x) \geq 0 \quad \forall \; x \in \mathbb{R}$

and $\sum_{x=1}^{\infty} f(x) = \sum_{x=1}^{\infty} \frac{1}{x(1+x)} = \sum_{x=1}^{\infty} \left[\frac{1}{x} - \frac{1}{x+1}\right]$

$$E(X) = \sum_{x=1}^{\infty} x \cdot f(x) = \sum_{x=1}^{\infty} x \cdot \frac{1}{x(1+x)} = \sum_{x=1}^{\infty} \frac{1}{1+x}$$

does not exist

~~Properties of expectation~~ Expectation of function of a discrete random variable:

1. Let $Z = \phi(X)$ be a function of the discrete random variable X.

$$E(Z) = E(\phi(x)) = \sum_{x} \phi(x) \cdot P(X = x)$$

provided that the expectation exists.

Ex, X a discrete r.v. with pmf

X	-2	-1	0	1	2
p(X)	1/5	1/5	1/5	1/5	1/5

$E(X)$, $E(X^2)$
$\quad 0 \qquad\quad 2$

Properties of expectation:

Let X and Y be the discrete random variable with finite expectations.

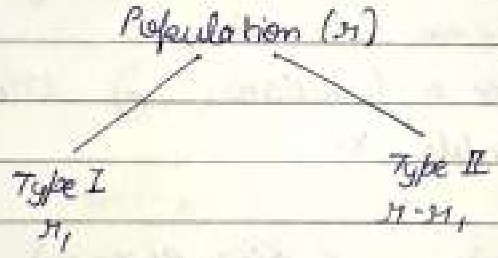
1) For an

3) $E(X+Y) = E(X) + E(Y)$

4) If $P(X \geq Y) = 1$, then $E(X) \geq E(Y)$ moreover, $E(X) = E(Y)$ if & only if $P(X = Y) = 1$.

5) $|E(X)| \leq E(|X|)$

Exercise: Hypergeometric



A sample of size n is drawn from the population.

S_n = number of objects of Type I in this sample.

Define: X_i : i^{th} indicator random variable indicating whether the i^{th} object in the sample is of Type I.

$S_n = X_1 + X_2 + \cdots + X_n$; $E(X_i) = \frac{r_1}{r}$

$E(S_n) = n \cdot \frac{r_1}{r}$

Moments: (For a fixed $r \in \{1, 2, 3, \ldots\}$)

$E(X^r) = \sum_x x^r P(X=x)$ for $r = 1, 2, \ldots$

These are called as raw moments of X. $r=1$ is $E(X)$ also called as mean of X.

$E(X-a)^r = \sum_x (x-a)^r P(X=x)$

for some fixed $a \in \mathbb{R}$

Central moments of X.

Take $a=0$, central moments are same as raw moments.

In particular take $a = \mu = E(X)$

For $r=2$

$E(X-\mu)^2 = E(X - E(X))^2$ = variance of X.

The positive square root of the variance of X is called as standard deviation of X.

$E(X-\mu)^2 = \sigma^2$ ← notation

Interpretations of $\sigma^2 = Var(X)$

measuring squared variability of X around a

$$= E(X-a)^2$$

Interested in minimising $E(X-a)^2$ w.r.t. a

(Approximating a random variable by a constant a and the error in the approximation is $E(X-a)^2$)

$$E(X-a)^2 = E(X^2 - 2aX + a^2)$$
$$= E(X^2) - 2aE(X) + a^2$$

differentiate w.r.t a

$$-2E(X) + 2a = 0$$
$$a = E(X) = \mu$$

Another interpretation of variance:

Given a random variable X and a number $a \in \mathbb{R}$.

$$(X-a)^2 = [(X-\mu) + (\mu - a)]^2 \quad \text{when} \quad \mu = E(X)$$

$$= (X-\mu)^2 + (\mu-a)^2 + 2(X-\mu)(\mu-a)$$

$$E(X-a)^2 = E(X-\mu)^2 + E(\mu-a)^2 + 2E(\mu-a)(X-\mu)$$

$$E(X-a)^2 = var(X) + (\mu-a)^2$$

Scanned by CamScanner

Chebyshev's inequality:

Let x be a non-negative random variable with finite expectation.
$t>0$ any positive real number.
Define a new random

$$Y = 0 \quad \text{if } X < t$$
$$Y = t \quad \text{if } X \geq t$$

$$E(Y) = 0 \cdot P(Y=0) + P(X \geq t) \cdot t$$

$$E(Y) = t \cdot P(X \geq t)$$

Note $X \geq Y$

$$E(X) \geq E(Y) = t \cdot P(X \geq t)$$

$$\Rightarrow P(X \geq t) \leq \frac{E(X)}{t}$$

Chebyshev's inequality:
X be a random variable with mean μ & variance σ^2. For any real $t>0$,

$$P(|X-\mu| > t) \leq \frac{\sigma^2}{t^2}$$

Consider $(X-\mu)^2$

$$P((X-\mu)^2 \geq t^2) \leq \frac{E(X-\mu)^2}{t^2} \quad \text{.... from } (*)$$

Scanned by CamScanner

Moment generating function (MGF)

$M_X(t) = E(e^{tx})$ for a given random variable X.

In case of a discrete random variable X with p.m.f. $p(X=x_i) = f(x_i)$ for $x_i \in R_x$,

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

$$= \sum_{x_i} e^{tx_i} p(X=x_i)$$

Ex:

$X \sim$ Bernoulli(p)

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

$$M_X(t) = e^{t}p + 1 - p$$

Ex:

$X \sim$ Binomial(n, p)

$$M_X(t) = \sum_{x=0}^{n} e^{tx} \binom{n}{x} p^x (1-p)^{n-x}$$

$$= \sum_{x=0}^{n} \binom{n}{x} (e^t p)^x \underbrace{(1-p)}_{q}^{n-x}$$

$$= (q + pe^t)^n$$

Scanned by CamScanner

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

$$M_X(t) = \sum_{x=0}^{\infty} e^{tx} \cdot \frac{e^{-\lambda} \lambda^x}{\lfloor x}$$

$$= e^{-\lambda} \cdot \sum_{x=0}^{\infty} \frac{(e^t \lambda)^x}{\lfloor x}$$

$$= e^{-\lambda} \cdot e^{\lambda e^t}$$

Consider $X \sim \text{Bernoulli}(p)$

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

$$\frac{d}{dt} M_X(t) = pe^t$$

$X \sim \text{Binomial}(n, p)$

$$M_X(t) = \{(1-p) + pe^t\}^n$$

$$\frac{d}{dt} M_X(t) =$$

$$E(e^{tx}) = E\left\{1 + tx + \frac{t^2 x^2}{\lfloor 2} + \frac{t^3 x^3}{\lfloor 3} + \cdots\right)$$

$$= E(1) + E(tx) + E\left(\frac{t^2 x^2}{\lfloor 2}\right) + E\left(\frac{t^3 x^3}{\lfloor 3}\right) + \cdots$$

$$= 1 + t\,E(X) + \frac{t^2}{\lfloor 2} E(x^2) + \frac{t^3}{\lfloor 3} E(x^3) + \cdots$$

$$\frac{d}{dt} M_X(t) = E(x) + \frac{2t}{\lfloor 2} E(x^2) + \frac{3t^2}{\lfloor 3} E(x^3) + \cdots$$

Scanned by CamScanner

$$\frac{d}{dt} M_x(t) \Big|_{t=0} = E(X)$$

$$\frac{d^2}{dt^2} M_x(t) \Big|_{t=0} = E(X^2)$$

i) $M_x(t=0) = 1$; $M_x(t)$ is always defined for $t=0$.

ii) $M_x(t)$ as a function of t should be defined in a small interval from 0 to t.

Let X be a random variable with Moment Generating Function $M_x(t)$. Find the Moment Generating Function of the random variable $aX+b$ for some $a, b \in \mathbb{R}$, $a \neq 0$.

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

$$= E\left(e^{aXt} \cdot e^{bt}\right) = e^{bt} \cdot E\left(e^{atX}\right)$$

$$= e^{bt} \cdot M_x(at)$$

Expectation of a continuous r.v.

Scanned by CamScanner

$$= \frac{\lambda^{\alpha}}{\Gamma(\alpha)} \cdot \frac{\alpha \Gamma\alpha}{\lambda^{\alpha+1}}$$

$$E(X) = \frac{\alpha}{\lambda}$$

Ex:

$X \sim Gamma(1, \lambda)$; $\lambda > 0$

$$f_X(x) = \lambda e^{-\lambda x} \quad x > 0$$
$$= 0 \quad \text{otherwise}$$

$Gamma(1, \lambda) = exp(\lambda)$; $\lambda > 0$

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

Ex: Let $f_X(x)$ denote the density of Cauchy random variable

$$f_X(x) = \frac{1}{\pi(x^2+1)} \quad -\infty < x < \infty$$

$$E(|X|) = \int_{-\infty}^{\infty} \frac{|x|}{\pi(x^2+1)} dx = \int_{0}^{\infty} \frac{2x}{\pi(x^2+1)} = \frac{1}{\pi} \lim_{k \to \infty} \int_{0}^{k} \frac{2x}{1+x^2} dx$$

Scanned by CamScanner

classmate

Date ______
Page ______

Moments of a continuous random variable:

1. $E(X^m) = \int_{-\infty}^{\infty} x^m f_X(x)\,dx$: m^{th} raw moment

2. $E((X-\sigma)^m) = \int_{-\infty}^{\infty} (x-\sigma)^m f_X(x)\,dx$: m^{th} central moment

$m = 1, 2, \ldots$

Variance of $X = \sigma^2 = E(X-\mu)^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f_X(x)\,dx$ where $\mu = E(X)$

Ex: $X \sim gamma(\alpha, \lambda)$

$$E(X^m) = \int_0^{\infty} x^m \cdot \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}\,dx$$

$$= \frac{\lambda^\alpha}{\Gamma\alpha} \int_0^{\infty} x^{m+\alpha-1} \cdot e^{-\lambda x}\,dx$$

$$= \frac{\lambda^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha+m)}{\lambda^{\alpha+m}}$$

$$= \frac{\alpha(\alpha+1)\cdots(\alpha+m-1)}{\lambda^m}$$

Variance of $X = \sigma^2 = E(X^2) - (E(X))^2$

$$= \frac{\alpha(\alpha+1)}{\lambda^2} - \frac{\alpha^2}{\lambda^2}$$

$$\sigma^2 = \frac{\alpha}{\lambda^2}$$

Scanned by CamScanner

Ex: In particular, if $X \sim \exp(\lambda)$

$$E(X^m) = \frac{m!}{\lambda^m}$$

$$Var(X) = \frac{1}{\lambda^2}$$

Ex:

Easy exercise

$X \sim U[a,b]$

$$E(X^2) = \int_a^b x^2 \cdot \frac{1}{b-a}\, dx = \frac{b^3-a^3}{3(b-a)} = \frac{b^2+a^2+ab}{3}$$

$$Var(X) = \frac{a^2+ab+b^2}{3} - \left[\frac{a+b}{2}\right]^2$$

$$\parallel$$

$$E(X^2) - (a+b)^2/4$$

Advantage of symmetry of a r.v. X.

Let X be a symmetric r.v.

$\Rightarrow$ X and $-X$ have same density

For any integer $m \in \mathbb{N}$,
X^m and $(-X)^m$ have same density

For m odd
$\Rightarrow$ X^m and $-X^m$ have same density.

$$\Rightarrow E(X^m) = E(-X^m) = -E(X^m)$$

$$\Rightarrow E(X^m) = 0$$

Scanned by CamScanner

In particular, mean of all symmetric densities is zero. $(G - N(0, \sigma^2))$

Moment Generating Functions (MGF):

$$M_x(t) = E(e^{tx}) \quad \text{if it exists.}$$

Ex: $x \sim N(\mu, \sigma^2)$

$$f_x(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} \qquad -\infty < x < \infty$$

$$= \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\} \qquad -\infty < x < \infty$$

$$M_x(t) = \int_{-\infty}^{\infty} e^{tx} \cdot \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2\sigma^2}(x-\mu)^2} dx$$

$$y = x - \mu$$

$$M_x(t) = \int_{-\infty}^{\infty} e^{(y+\mu)t} \cdot \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{y^2}{2\sigma^2}} dy$$

$$= e^{\mu t} \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{\left(ty - \frac{y^2}{2\sigma^2}\right)} dy$$

Notice: $$ty - \frac{y^2}{2\sigma^2} = -\frac{(y - \sigma^2 t)^2}{2\sigma^2} + \frac{\sigma^2 t^2}{2}$$

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

Scanned by CamScanner

classmate
Date
Page

$$= e^{\mu t + \frac{\sigma^2 t^2}{2}} \cdot \underbrace{\int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y-\sigma^2 t)^2}{2\sigma^2}} dy}_{1 \quad N(\sigma^2 t, \sigma^2)}$$

$$M_x(t) = e^{\mu t + \frac{\sigma^2 t^2}{2}}$$

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

$$E(x) = \text{mean of } x = \frac{d}{dt} M_x(t)\Big|_{t=0} = \left[\frac{d}{dt} e^{\mu t + \frac{1}{2}\sigma^2 t^2}\right]_{t=0}$$

$$= (\mu + \sigma^2 t) e^{\mu t + \frac{1}{2}\sigma^2 t^2}\Big]_{t=0}$$

$$E(x) = \mu$$

$$E(x^2) = \left[\frac{d^2}{dt^2} M_x(t)\right]_{t=0}$$

$$= \left[\frac{d}{dt} \mu e^{\mu t + \frac{1}{2}\sigma^2 t^2} + \frac{d}{dt} \sigma^2 t e^{\mu t + \frac{1}{2}\sigma^2 t^2}\right]_{t=0}$$

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

$$\text{Var}(x) = E(x^2) - (E(x))^2 = \sigma^2 + \mu^2 - \mu^2 = \sigma^2$$

Scanned by CamScanner

Ex: Let $Y \sim N(0,1)$

$$M_Y(t) = e^{t^2/2} \qquad -\infty < t < \infty$$

Define $X = \mu + \sigma Y$

$$M_X(t) = M_{\mu+\sigma Y}(t) = e^{\mu t} \cdot M_Y(\sigma t) = e^{\mu t} \cdot e^{\sigma^2 t^2/2}$$

$$= e^{\mu t + \frac{1}{2}\sigma^2 t^2}$$

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

In general, if $X \sim N(\text{mean}, \text{variance})$

$$M_X(t) = e^{(\text{mean})t + (\text{variance})\frac{t^2}{2}}$$

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

Define: $Y = \frac{X-\mu}{\sigma} = \underbrace{-\frac{\mu}{\sigma}}_{b} + \underbrace{\frac{1}{\sigma}}_{a} X$

$$M_{aX+b}(t) = M_Y(t) = e^{t^2/2} \Longrightarrow \{\text{Proof}\}$$

Example: $X \sim \text{gamma}(\alpha, \lambda)$

$$M_X(t) = \int_0^\infty e^{tx} \cdot \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}\, dx$$

$$= \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty x^{\alpha-1} e^{-(\lambda-t)}\, dx$$

Scanned by CamScanner

$$= \frac{\lambda^{\alpha}}{\Gamma(\alpha)} \cdot \frac{\Gamma(\alpha)}{(\lambda-t)^{\alpha}}$$

$$M_x(t) = \left(\frac{\lambda}{\lambda-t}\right)^{\alpha} \qquad -\infty < t < \lambda$$

Ex: Exponential density with parameter λ $M_x(t) = \frac{\lambda}{\lambda - t}$

$$\exp(\lambda) = \text{gamma}(1, \lambda)$$

Quartiles :

A point $q_1 \in \mathbb{R}$ is called as 1^{st} quartile

$$P(x \leq q_1) = \frac{1}{4}$$

A point $m \in \mathbb{R}$ is called as 2^{nd} quartile or median if

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

A point $q_3 \in \mathbb{R}$, is called as 3^{rd} quartile if

$$P(X \leq q_3) = 3/4$$

Mode:

Scanned by CamScanner

Ex: Let $X \sim$ Binomial $(3, 1/2)$

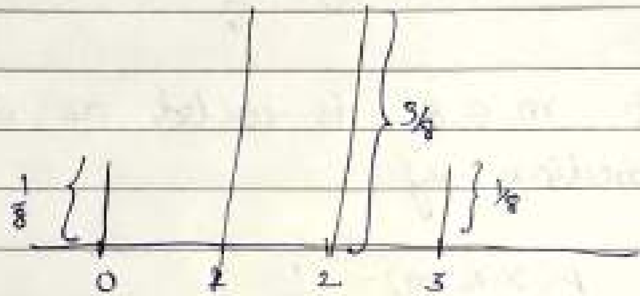
Compute median & mode.

$$p_X(x) = \binom{3}{x}\left(\frac{1}{2}\right)^{x}\left(\frac{1}{2}\right)^{3-x} \qquad x = 0, 1, 2, 3$$

$$= 0 \qquad \text{otherwise}$$

x	0	1	2	3
$p(x)$	1/8	3/8	3/8	1/8

mean = 3/2
mode = 1, 2
median = 1



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

mean $(X) = \mu$ median$(X) = \mu$

$$P(X \le m) = 0.5$$

$$\Rightarrow P\left(\frac{X-\mu}{\sigma} \le \frac{m-\mu}{\sigma}\right) = 0.5$$

$$\Rightarrow \Phi\left(\frac{m-\mu}{\sigma}\right) = 0.5$$

$$\Rightarrow \frac{m-\mu}{\sigma} = 0$$

$$\Rightarrow m = \mu$$

$mode(X) = \mu$

In this density, mean = mode = median = μ.

Ex: $X \sim Uniform(a, b)$

$$mean(X) = \frac{a+b}{2}$$

$F(m) = 0.5$ $\quad \frac{m-a}{b-a} = \frac{1}{2} \Rightarrow \boxed{m = \frac{a+b}{2}}$

Ex: $X \sim exp(\lambda)$

$$F(m) = \frac{1}{2}$$

$$1 - e^{-\lambda m} = \frac{1}{2}$$

$$\Rightarrow \frac{1}{2} = e^{-\lambda m} \qquad e^{\lambda m} = 2$$

$$m = \frac{\ln 2}{\lambda}$$

$$\mu = \frac{1}{\lambda}$$

Ex: A point is chosen randomly between the interval [-10, 10] (by uniform probability principle). Let X be a random variable defined in such a way that X denotes the co-ordinate of the chosen point if the point belongs to [-5, 5]. X takes value -5 if the point belongs to [-10, -5] and X takes value 5 if point belongs to [5, 10]. Compute CDF of X.

Ex: X is a cts. r.v. with p.d.f.

$$f(x) = \frac{1}{2} e^{-|x|} \qquad -\infty < x < \infty$$

Compute $P(1 \leq |x| \leq 2)$

Scanned by CamScanner

Random Vectors: $(\Omega, \mathcal{F}, P)$

Let $X_1, X_2, \ldots, X_r$ be r discrete random variables on $(\Omega, \mathcal{F}, P)$

Define $X = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ \vdots \\ X_r \end{bmatrix}$ is an r-dimensional vector

For an $\omega \in \Omega$

$$X(\omega) = \begin{bmatrix} X_1(\omega) \\ X_2(\omega) \\ \vdots \\ X_r(\omega) \end{bmatrix} \in \mathbb{R}^r$$

Suppose $X_1(\omega) = x_1$, $X_2(\omega) = x_2$, $\ldots$ $X_r(\omega) = x_r$

$$X(\omega) = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_r \end{bmatrix} \in \mathbb{R}^r$$

Definition: Let $X = \begin{pmatrix} X_1 \\ \vdots \\ X_r \end{pmatrix}$ be a vector where X_i is a random variable on $(\Omega, \mathcal{F}, P)$. For every $\underline{x} \in \mathbb{R}^r$ the set $\{\omega \in \Omega \mid \underline{X}(\omega) = \underline{x}\} \in \mathcal{F}$. Then $X: \Omega \to \mathbb{R}^r$ is called as an r-dimensional random vector.

We are interested in $Prob(\underline{X} = \underline{x})$.

Scanned by CamScanner

Let $\underline{X} = \begin{pmatrix} X_1(\omega) \\ X_2(\omega) \\ \vdots \\ X_r(\omega) \end{pmatrix}$ be a discrete random vector

If $\underline{x}$ denotes the value assumed by the random vector $\underline{X}$, then

$\{ x : P(\underline{X}(\omega) = \underline{X}) > 0 \}$ is finite or countably infinite

Definition: The discrete density / discrete joint p.m.f of the random vector $\underline{X}$ is defined as:

$$f(x_1, x_2, \dots, x_r) = \text{Prob}(X_1 = x_1, X_2 = x_2, \dots, X_r = x_r)$$

where $\underline{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_r \end{bmatrix}$ $\quad \underline{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_r \end{bmatrix}$

In the vector notation

$$f(\underline{x}) = P(\underline{X} = \underline{x}) \quad \forall \; \underline{x} \in \mathbb{R}^r$$

For a subset $A \subseteq \mathbb{R}^r$,

$$P(\underline{X} \in A) = \sum_{\underline{x} \in A} f(\underline{x})$$

Definition:

A function f is called as discrete joint p.m.f. if:

i) $f(\underline{x}) \geq 0 \quad \forall \; \underline{x} \in \mathbb{R}^r$

ii) $\{ x : f(x) \neq 0 \}$ is finite or countably infinite.

We denote the elements of this set as $\underline{x}_1, \underline{x}_2, \underline{x}_3, \dots$

iii) $\sum_i f(x_i) = 1$

Scanned by CamScanner

Ex:

$x_1 \backslash x_2$	1	2	3	4
1	1/4	1/8	1/16	1/16
2	1/16	1/16	1/4	1/8

$\underline{X} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ $\qquad P(\underline{X} = \underline{x})$

$\Rightarrow P(X_1 = x_1 \,;\, X_2 = x_2)$

$= f(x_1, x_2) = f(\underline{X})$

Very easy to verify that this table is a discrete joint pmf.

$Prob(x_1 \geq x_2) = \frac{1}{4} + \frac{1}{16} + \frac{1}{16}$

$Prob(X_1 = 1) = \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \frac{1}{16} = \frac{1}{2}$

$= Prob(X_1 = 1, X_2 = 1) + Prob(X_1 = 1; X_2 = 2) + Prob(X_1 = 1, X_2 = 3) + Prob(X_1 = 1, X_2 = 4)$

$Prob(X_1 = 1) = \sum_{x_2} f(1, x_2) = 1/2$

$Prob(X_1 = 2) = \sum_{x_2} f(2, x_2) = 1/2$

$$\boxed{f_{x_1}(x_1) = \sum_{x_2} f(x_1, x_2) \quad \text{for } x_1 \in R_{x_1}}$$

marginal p.m.f of x_2

Independent random variables:

Let $X_1, X_2, \ldots X_r$ be r discrete random variables with p.m.fs
$f_1, f_2, \ldots, f_r$ respectively. The random variables $X_1, \ldots, X_r$ are called as mutually independent if their joint p.m.f. f is given by

$$f(x_1, x_2, \ldots, x_r) = f_1(x_1) \cdot f_2(x_2) \cdots f_r(x_r)$$

Notation:

$$f(x_1, x_2, \ldots, x_r)$$

$$= Prob(X_1 = x_1, X_2 = x_2, \ldots, X_r = x_r)$$

Define $A_1 \subseteq \Omega$ such that $A_1 = [X_1 = x_1]$

$$A_1 = \{\omega : X_1(\omega) = x_1\}$$

$$= Prob(A_1 \cap A_2 \cap A_3 \cap \cdots A_r)$$
$$= Prob(A_1) \cdot Prob(A_2) \cdots Prob(A_r)$$
$$= Prob(X_1 = x_1) \cdot Prob(X_2 = x_2) \cdots Prob(X_r = x_r)$$
$$= f_1(x_1) \cdot f_2(x_2) \cdots f_r(x_r)$$

$x_1 \backslash x_2$	0	1
0	1/4	1/4
1	1/4	1/4

Ex: Let X_1 and X_2 be two independent random variables each with geometric distribution with parameter p

Find the ~~[illegible]~~ distribution $\min(X_1, X_2)$

$$\text{Prob}[\min(X_1, X_2) \geq z]$$

$$= \text{Prob}[X_1 \geq z \ ; \ X_2 \geq z]$$

$$= \text{Prob}(X_1 \geq z_1) \cdot \text{Prob}(X_2 \geq z_2)$$ independence of X_1 & X_2

$$= (1-p)^z \cdot (1-p)^z = (1-p)^{2z}$$

$$\min(X_1, X_2) \sim \text{geometric}\left(1-(1-p)^2\right)$$

Sum of independent random variables:

Let X & Y be two independent random variables

Let $x_1, x_2, \ldots$ be the distinct values taken by X.

Interested in the event

$$\{X+Y = z\} \quad \forall z$$

$$\| $$

$$\bigcup_i \{X = x_i, Y = z - x_i\}$$ ← Note this union is disjoint.

$$P(X+Y = z) = P\left(\bigcup_i \{X = x_i, Y = z - x_i\}\right)$$

$$= \sum_i P(X = x_i, Y = z - x_i)$$

$$= \sum_i P(X = x_i)\, P(Y = z - x_i)$$

Scanned by CamScanner

$$\boxed{f_{X+Y}(z) = \sum_{x} f_X(x) \cdot f_Y(z-x)}$$

Expectation:

Let $\underline{X} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_r \end{pmatrix}$ be a given discrete / random vector with joint p.m.f.

$$f_{X_1, \ldots, X_r}(x_1, \ldots, x_r)$$

$$E(h(X_1, \ldots, X_r)) = \sum_{x_1, x_2, \ldots x_r} h(x_1, \ldots, x_r) f_{X_1, X_2, \ldots X_r}(x_1, \ldots, x_r)$$

For $r = 2$, $h(x_1, x_2) = x_1 + x_2$

$$E(X_1 + X_2) = \sum_{x_1} \sum_{x_2} (x_1 + x_2) f_{X_1, X_2}(x_1, x_2)$$

$r = 2$,

$h(x_1, x_2) = x_1$

$$E(X_1) = \sum_{x_1} \sum_{x_2} x_1 f_{X_1, X_2}(x_1, x_2) = \sum_{x_1} x_1 \boxed{\sum_{x_2} f_{X_1, X_2}(x_1, x_2)}$$

$$\boxed{E(X_1) = \sum_{x_1} x_1 f_{X_1}(x_1)}$$

In general, take $h(x_1, \ldots, x_r) = x_i$ for $1 \le i \le r$

$$\boxed{E(x_i) = \sum x_i f_{x_i}(x_i)}$$

$$\sum_{x_1} \sum_{x_2} f_{x_1, x_2}(x_1, x_2) = 1$$

$$\parallel$$

$$\sum_{x_i} f_{x_1}(x_i) \Rightarrow \text{Marginal density}$$

Let X and Y be independent discrete random variable with joint pmf $f_{x,y}(x,y)$

$$E(XY) = \sum_x \sum_y xy\, f_{x,y}(x,y)$$

$$= \sum_x \sum_y xy\, f_x(x) \cdot f_y(y)$$

$$= \sum_x x f_x(x) \cdot \sum_y y f_y(y) = E(X) \cdot E(Y)$$

$$\boxed{E(XY) = E(X) \cdot E(Y)}$$

$$E(\psi_1(x)\ \psi_2(y))$$

$$E(\psi_1(x), \psi_2(y)) = \sum_x \sum_y \psi_1(x) \cdot \psi_2(y)\, f_{x,y}(x,y)$$

$$= \sum_x \psi_1(x) \cdot f_x(x) \ \sum_y \psi_2(y) \cdot f_y(y)$$

Scanned by CamScanner

X & Y are independent discrete r.v.s with M.G.F. of $X+Y$ joint pmf $f_{X,Y}(x,y)$

$$M_{X+Y}(t) = E(e^{t(x+y)}) = E(e^{tx} \cdot e^{ty})$$

$$= E(e^{tX}) \cdot E(e^{tY}) \quad \text{independent}$$

$$= M_X(t) \cdot M_Y(t)$$

$$M_{X+Y}(t) = M_X(t) \cdot M_Y(t)$$

Generalising this result,

$X_1, X_2, \ldots, X_n$ are mutually independent discrete r.v.s

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

Ex: Let $X_i \sim \text{Bernouli}(p)$ for $i = 1, 2, \ldots, n$ be iid (independent & identically distributed)

$$Y = \sum_{i=1}^{n} X_i$$

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

$$= \prod_{i=1}^{n} (1-p+pe^t) = (1-p+pe^t)^n$$

Binomial(n, p) = Sum of n independent Bernouli(p)

Scanned by CamScanner

Important Observation:

X & Y are independent

$\Rightarrow \quad E(XY) = E(X) \cdot E(Y)$ already proved.

Q: Is the converse true? NO

Ex: Let (X, Y) be a discrete r.v. with range $R_{(X,Y)} = \{(0,1), (0,-1), (1,0), (-1,0)\}$ with each outcome is equally likely

$$p(x,y) = \begin{cases} \frac{1}{4} & \text{for } (x,y) \in R_{X,Y} \\ 0 & \text{otherwise} \end{cases}$$

$E(X) = 0 \qquad E(Y) = 0$

Now, $XY = 0 \quad \Rightarrow \quad E(XY) = 0$

Observe that:

$p(X=0) = 1/2 \qquad p(Y=0) = 1/2$

$P(X=0; Y=0) = 0 \neq P(X=0) \cdot P(Y=0) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$

Remember:

$E(X+Y) = E(X) + E(Y)$

$E(aX) = aE(X)$

$E(X+C) = E(X) + C$

Scanned by CamScanner

Sum of variances:

Let X & Y be discrete random variables.

$$var(X+Y) = E(X+Y - E(X+Y))^2$$

$$= E(X+Y - E(X) - E(Y))^2$$

$$= E[(X - E(X)) + (Y - E(Y))]^2$$

$$= E(X - E(X))^2 + E(Y - E(Y))^2 + 2\underbrace{E\{(X - E(X)) \cdot (Y - E(Y))\}}_{cov(X,Y)}$$

$$var(X+Y) = var(X) + var(Y) + 2cov(X,Y)$$

<u>Note:</u>

$$cov(X,Y) = E\{(X - E(X)) \cdot (Y - E(Y))\}$$

$$= E(XY - XE(Y) - YE(X) + E(X) \cdot E(Y))$$

$$= E(XY) - E(X) \cdot E(Y) - E(X) \cdot E(Y) + E(X) \cdot E(Y)$$

$$\boxed{cov(X,Y) = E(XY) - E(X) \cdot E(Y)}$$

Corollary:

If X & Y are independent, then $cov(X,Y) = 0$

converse is NOT true.

$cov(X,Y) = 0 \not\Rightarrow$ X & Y are independent.

If X & Y are independent r.v.s

$$var(X+Y) = var(X) + var(Y)$$

Scanned by CamScanner

Date
Page

$$Var(aX) = E(aX - E(aX))^2 = E[aX - aE(X)]^2$$
$$= a^2 E(X - E(X))^2$$
$$= a^2 Var(X)$$

$$\boxed{Var(aX) = a^2 Var(X)}$$

- Var(c) where c is a constant

$$E(c - E(c))^2 = E(c-c)^2 = 0$$

Let $X_1, X_2, \ldots, X_n$ be independent random variables with variances $\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2$.

$$Var(X_1 + X_2 + \cdots + X_n)$$
$$= Var(X_1) + Var(X_2) + \cdots + Var(X_n)$$
$$= \sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2$$

Let $X_1, X_2, \ldots, X_n$ be iid with mean μ and variance σ^2.

$$E(\bar{X}) = E\left(\frac{X_1 + X_2 + \cdots + X_n}{n}\right) = E\left(\frac{X_1}{n}\right) + \cdots + E\left(\frac{X_n}{n}\right)$$
$$= \frac{1}{n}\left[E(X_1) + E(X_2) + \cdots + E(X_n)\right]$$
$$= \frac{1}{n}\mu + \frac{1}{n}\mu + \cdots = \mu$$

$$E(\bar{X}) = \mu$$

$$Var(\bar{X}) = Var\left(\frac{X_1 + X_2 + \cdots + X_n}{n}\right)$$
$$= Var\left(\frac{X_1}{n}\right) + Var\left(\frac{X_2}{n}\right) + \cdots + Var\left(\frac{X_n}{n}\right)$$

Scanned by CamScanner

$$= \frac{1}{n^2}\left[\sum_{i=1}^{n} Var(X_i)\right] = \frac{n \cdot \sigma^2}{n^2}$$

$$\boxed{Var(\bar{X}) = \frac{\sigma^2}{n}}$$

Correlation Coefficient:

Let X, Y be two discrete r.v.s. Then correlation coefficient $\rho(x, y)$ (rho) is defined as:

$$\rho_{XY} = \rho = \rho(X,Y) = \frac{cov(X,Y)}{\sqrt{Var(X) \cdot Var(Y)}} = \frac{cov(X,Y)}{SD(X) \cdot SD(Y)}$$

If X & Y are independent $\Rightarrow \rho_{XY} = 0$

Theorem: Schwartz Inequality

Let X & Y be random variables with finite second order moments.

$$[E(XY)]^2 \leq E(X^2) \cdot E(Y^2)$$

Futhermore, equality holds when $P(Y=0) = 1$ or $P(X = aY) = 1$ for some $a \in \mathbb{R}$.

Proof:
Easy to see that when
$P(Y=0)=1$ or $P(X = aY) = 1$, equality holds.

In order to prove inequality, for any $\lambda \in \mathbb{R}$.

$$0 \leq E(X - \lambda Y)^2 = \lambda^2 E(Y^2) - 2\lambda E(XY) + E(X^2)$$

Since the above expression is a quadratic in λ and $E(Y^2) > 0$, the minimum value of this quadratic expression is achieved at

$$2\lambda E(Y^2) - 2E(XY) = 0$$

$$\lambda = \frac{E(XY)}{E(Y^2)}$$

The minimum value at the point $\frac{E(XY)}{E(Y^2)}$ is given by:

$$\left[\frac{E(XY)}{E(Y^2)}\right]^2 \cdot E(Y^2) - 2\left[\frac{E(XY)}{E(Y^2)}\right] \cdot E(XY) + E(X^2) \geq 0$$

$$\Rightarrow \quad - \frac{[E(XY)]^2}{E(Y^2)} + E(X^2) \geq 0$$

$$\Rightarrow \quad \boxed{[E(XY)]^2 \leq E(X^2) \cdot E(Y^2)}$$

Importance of Schwartz inequality:

Apply this Schwartz inequality to two random variables $X - E(X)$ and $Y - E(Y)$

$$\left[E\{(X - E(X))(Y - E(Y))\}\right]^2 \leq E(X - E(X))^2 \cdot E(Y - E(Y))^2$$

$$\Rightarrow \quad [Cov(X, Y)]^2 \leq Var(X) \cdot Var(Y)$$

Scanned by CamScanner

$$\Rightarrow \quad \rho_{XY}^2 = \frac{[\mathrm{cov}(X,Y)]^2}{\mathrm{Var}(X)\cdot \mathrm{Var}(Y)} \le 1$$

$$\Rightarrow \quad |\rho_{xy}| \le 1$$

$$\Rightarrow \quad -1 \le \rho_{XY} \le 1$$

$$\boxed{\rho(X,Y) = 1 \quad \Leftrightarrow \quad P(X = aY) = 1}$$

Recall: If X is non-negative r.v. with finite expectation then for $t > 0$

$$P(X \ge t) \le \frac{E(X)}{t}$$

Chebyshev inequality:

$$P(|X - \mu| > t) \le \frac{\sigma^2}{t^2}.$$

Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables

Let $\mu = E(X_i)$

$$E\left(\frac{S_n}{n}\right) = E\left(\frac{X_1 + X_2 + \cdots + X_n}{n}\right) = \mu$$

$$\mathrm{Var}\left(\frac{S_n}{n}\right) = \mathrm{Var}\left(\frac{X_1 + X_2 + \cdots + X_n}{n}\right) = \frac{\sigma^2}{n}$$

Chebyshev's inequality $\Rightarrow$

$$P\left(\left|\frac{S_n}{n} - \mu\right| \ge \delta\right) \le \frac{\mathrm{Var}(S_n/n)}{\delta^2} = \frac{\sigma^2}{n\delta^2}$$

Scanned by CamScanner

In particular,

$$\lim_{n \to \infty} P\left(\left| \frac{S_n}{n} - \mu \right| \geq \delta \right) = 0$$

WLLN

Weak Law of Large Numbers

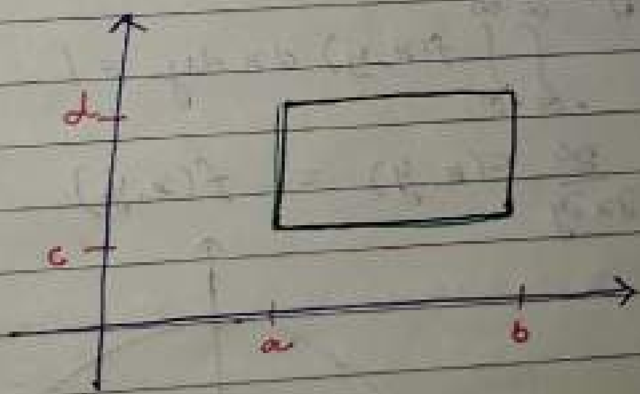
Joint Continuous random variables

X and Y are continuous r.v.s on the same probability space.

$$F(x, y) = \text{Prob}(X \leq x, Y \leq y)$$

$$-\infty < x, y < \infty$$

Rectangle $R = \{ (x, y) \mid a < x \leq b \ , \ c < y \leq d \}$



$$P((X, Y) \in R) = P(a < X \leq b, \ c < Y \leq d)$$

$$= F(b, d) - F(a, d) - F(b, c) + F(a, c)$$

Marginal Distributions:

Scanned by CamScanner

$F_X(x) = P(X \le x) = F(x, \infty) = \lim_{y \to \infty} F(x, y)$ Marginal CDF of X

$F_Y(y) = P(Y \le y) = F(\infty, y) = \lim_{x \to \infty} F(x, y)$ Marginal CDF of Y

If there exist a non-negative function: $f(x,y)$ over $\mathbb{R}^2$ such that

$$F(x, y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f(u, v)\, dv\, du,$$

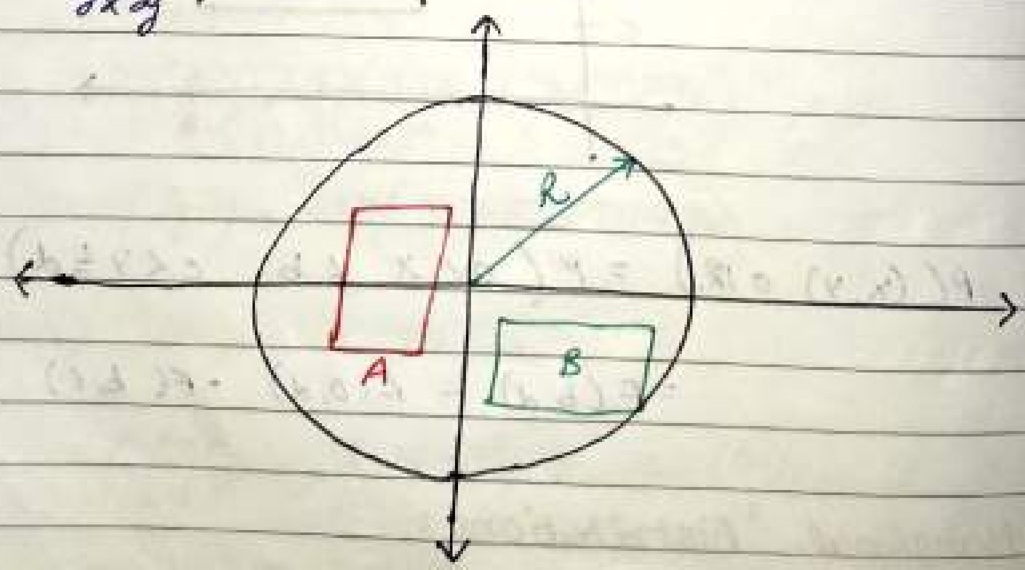
then $f(x,y)$ is called joint pdf of (X, Y)

$$P\big((X,Y) \in A\big) = \iint_A f(x,y)\, dx\, dy$$

Obviously,

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\, dx\, dy = 1$$

and $\dfrac{\partial^2}{\partial x \partial y} F(x,y) = f(x,y)$



classmate
Date
Page

$$f(x,y) = \begin{cases} \dfrac{1}{\pi R^2} & (x,y) \in \text{circle } x^2+y^2 \le R^2 \\ = 0 & \end{cases}$$

$$P((x,y) \in A) = \iint_A f(x,y)\,dx\,dy$$

$$= \frac{\text{Area of } A}{\pi R^2}$$

$$F_X(x) = F(x,\infty) = \lim_{y\to\infty} F(x,y)$$ Marginal pdf of x

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy = \int_{-\infty}^{\infty} f_{XY}(x,y)\,dy$$

$$f_X(x) = \int_{-\sqrt{R^2-x^2}}^{\sqrt{R^2-x^2}} \frac{1}{\pi R^2}\cdot dy = \frac{2\sqrt{R^2-x^2}}{\pi R^2}$$

$$f_X(x) = \begin{cases} \dfrac{2\sqrt{R^2-x^2}}{\pi R^2} & -R < x < R \\ 0 & \text{otherwise} \end{cases}$$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x,y)\,dx =$$ Marginal pdf of Y.

$$f_Y(y) = \begin{cases} \dfrac{2\sqrt{R^2-y^2}}{\pi R^2} & -R < y < R \\ 0 & \text{otherwise} \end{cases}$$

Scanned by CamScanner

Independence: X and Y are called independent r.v.s if and only if

$$f_{XY}(x,y) = f_X(x) \cdot f_Y(y)$$

(joint pdf is the product of marginal pdfs)

In this case X & Y are <u>NOT</u> independent.

Easy Way to generate examples:

Let X & Y be two independent continuous r.v.s with pdfs $f_1(x)$ & $f_2(y)$ respectively.

$$f_{XY}(x,y) = f_1(x) \cdot f_2(y)$$

$$f_{XY}(x,y) \geq 0 \qquad \forall\, x, y$$

$$\iint_{x\,y} f_{XY}(x,y)\,dx\,dy = \int_x f_X(x)\,dx \cdot \int_y f_Y(y)\,dy$$

Ex:

$$X \sim N(0,1) \qquad Y \sim N(0,1)$$

$$\varphi_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \qquad -\infty < x < \infty$$

$$\varphi_Y(y) = \frac{1}{\sqrt{2\pi}} e^{-y^2/2} \qquad -\infty < y < \infty$$

$$f_{XY}(x,y) = \varphi_X(x) \cdot \varphi_Y(y)$$

$$= \frac{1}{2\pi} e^{-(x^2+y^2)/2} \qquad -\infty < x, y < \infty$$

Scanned by CamScanner

Example: Let X and Y have joint density

$$f(x,y) = ce^{-(x^2-xy+y^2)/2} \qquad -\infty < x,y < \infty$$

Find c??

Marginal of X:

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy$$

$$= c\int_{-\infty}^{\infty} e^{-(x^2-xy+y^2)/2}\,dy$$

$$= c\int_{-\infty}^{\infty} e^{-\left[(y-x/2)^2 + \frac{3x^2}{4}\right]/2}\,dy$$

$$= ce^{-\frac{3x^2}{8}}\int_{-\infty}^{\infty} e^{-(y-x/2)^2/2}\,dy$$

$$= ce^{-\frac{3x^2}{8}}\int_{-\infty}^{\infty} e^{-u^2/2}\,du$$

$$f_X(x) = \sqrt{2\pi}\,c\,e^{-3x^2/8}$$

$$\sqrt{2\pi}\cdot c = \frac{1}{\sqrt{2\pi\cdot\frac{4}{3}}}$$

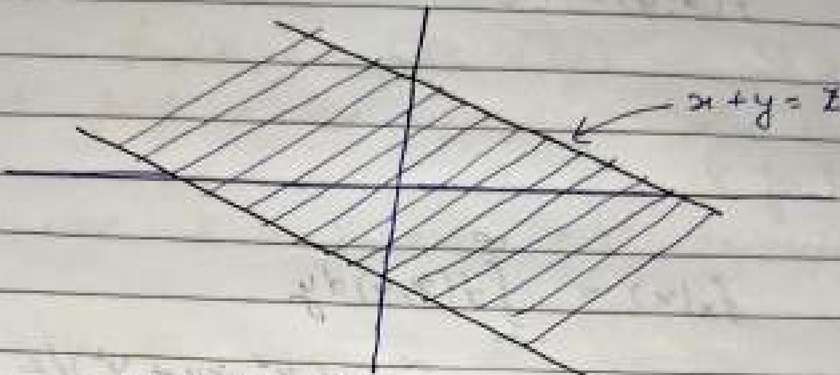
$$c = \frac{\sqrt{3}}{4\pi}$$

$$f_Y(y) = \sqrt{2\pi}\,c\,e^{-3y^2/8}$$

Scanned by CamScanner

Distribution of Sums $\{ \phi(x,y) = X+Y \}$

$$A_z = \{ (x,y) \mid x+y \le z \}$$



$$F_Z(z) = \iint_{A_z} f(x,y)\,dx\,dy$$

$$= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{z-x} f(x,y)\,dy \right] dx$$

Substitute $y = v - x$ in the inner integral.

$$F_Z(z) = \int_{-\infty}^{\infty} \left[\int_{-\infty}^{z} f(x, v-x)\,dv \right] dx$$

↳ convergent, change of variable can be made

$$F_Z(z) = \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} f(x, v-x)\,dv \right] dx$$

Thus the pdf of $X+Y$ is given by

$$f_{X+Y}(x) = \int_{-\infty}^{\infty} f(x, z-x)\,dz \qquad -\infty < z < \infty$$

Scanned by CamScanner

$$F_Z(x) = \iint_{A_z} f(x,y)\, dx\, dy$$

$$= \int_{-\infty}^{\infty}\left[\int_{-\infty}^{z-x} f(x,y)\, dy\right] dx$$

Substitute $y = v - x$ in the inner integral

$$F_Z(x) = \int_{-\infty}^{\infty}\left[\int_{-\infty}^{x} f(x, v-x)\, dv\right] dx$$

$$F_Z(x) = \int_{-\infty}^{z}\boxed{\left[\int_{-\infty}^{\infty} f(x, v-x)\, dx\right]} dv$$

Thus, the pdf of $X+Y$ is given by:

$$f_{X+Y}(x) = \int_{-\infty}^{\infty} f(x, z-x)\, dx \qquad -\infty < z < \infty$$

If X & Y are independent

$$f_{X+Y}(z) = \int_{-\infty}^{\infty} f_X(x)\cdot f_Y(z-x)\, dx \qquad -\infty < z < \infty$$

If X & Y are non-negative independent random variables.

$$f_{X+Y}(z) \begin{cases} = \int_{0}^{z} f_X(x)\cdot f_Y(z-x)\, dx & 0 < z < \infty \\ = 0 & \text{elsewhere} \end{cases}$$

Scanned by CamScanner

classmate
Date
Page

Ex:

$X, Y \sim \exp(\lambda)$ are independent

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

$$f_Y(y) = \begin{cases} \lambda e^{-\lambda y} & y \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

$$f_{X+Y}(z) = \int_0^z f_X(x) \cdot f_Y(z-x)\,dx \qquad z \geq 0$$

$$= 0 \qquad \text{otherwise}$$

For $z \geq 0$

$$f_{X+Y}(z) = \int_0^z \lambda e^{-\lambda x} \cdot \lambda e^{-\lambda(z-x)}\,dx$$

$$= \lambda^2 e^{-\lambda z} \int_0^z dx = \lambda^2 z e^{-\lambda z}$$

$$f_{X+Y}(z) = \lambda^2 z e^{-\lambda z} \qquad z \geq 0$$

$$= 0 \qquad \text{otherwise}$$

Thus $X + Y \sim \text{gamma}(2, \lambda)$

$$\text{pdf of gamma}(\alpha, \lambda) = \begin{cases} \dfrac{\lambda^\alpha}{\Gamma \alpha} \displaystyle\int_0^\infty x^{(\alpha-1)} e^{-\lambda x}\,dx \\ 0 \end{cases}$$

Scanned by CamScanner

Ex:

X, Y are iid $U(0,1)$

$$f_{X+Y} = ??$$

$$f_{X+Y} = \int_0^z f_X(x) \cdot f_Y(z-x)\, dx$$

$f_X(x) \cdot f_Y(z-x)$ takes only values zero or 1

$f_X(x) f_Y(z-x)$ takes value 1 when $0 \le x \le 1$ and $0 \le z-x \le 1$

If $0 \le z \le 1$, then the integrand has value 1 on the set $0 \le x \le z$.

$$f_{X+Y}(z) = z \qquad 0 \le z \le 1$$

If $1 < z \le 2$, then the integrand has value 1 on the set

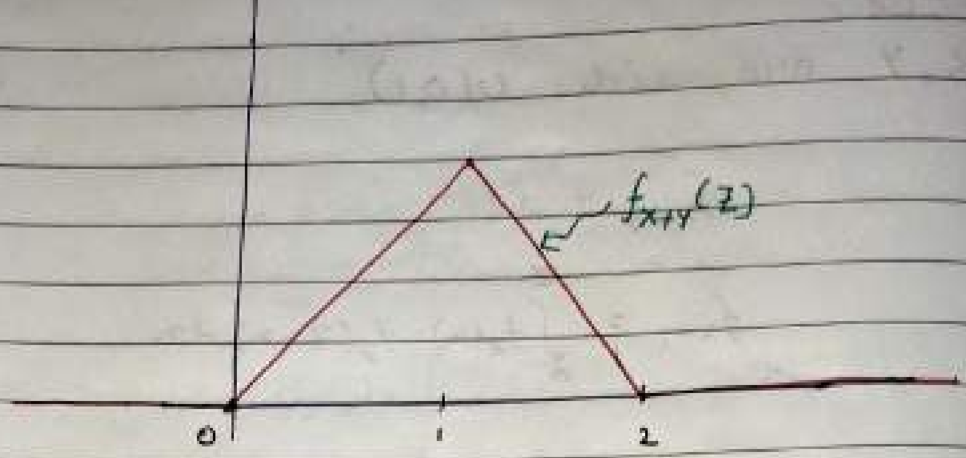
$$z-1 \le x \le 1$$

$$f_{X+Y}(z) = 2 - z \qquad 1 < z \le 2$$

Thus,

$$f_{X+Y}(z) = \begin{cases} 0 & z < 0 \\ z & 0 \le z \le 1 \\ 2-z & 1 < z \le 2 \\ 0 & z > 2 \end{cases}$$

Scanned by CamScanner



Ex:

Let $X \sim \Gamma(\alpha_1, \lambda)$ and $Y \sim \Gamma(\alpha_2, \lambda)$ be independent r.v.s. Then

$$X+Y \sim \Gamma(\alpha_1 + \alpha_2, \lambda)$$

$$f_X(x) = \frac{\lambda^{\alpha_1} \cdot x^{\alpha_1 - 1} e^{-\lambda x}}{\Gamma(\alpha_1)} \qquad x > 0$$

$$f_Y(y) = \frac{\lambda^{\alpha_2} \cdot y^{\alpha_2 - 1} \cdot e^{-\lambda y}}{\Gamma(\alpha_2)} \qquad y > 0$$

For $z > 0$,

$$f_{X+Y}(z) = \frac{\lambda^{\alpha_1 + \alpha_2}}{\Gamma\alpha_1 \cdot \Gamma\alpha_2} \int_0^z e^{-\lambda z} \cdot x^{\alpha_1 - 1} (z-x)^{\alpha_2 - 1} \, dx$$

Ex: Let X, Y be independent random variable with

$$X \sim N(\mu_1, \sigma_1^2)$$
$$Y \sim N(\mu_2, \sigma_2^2)$$

Then what is the pdf of $X+Y$??

Scanned by CamScanner

(Another approach to MGF)

$$M_{X+Y}(t) = E(e^{t(X+Y)})$$
$$= E(e^{tX} \cdot e^{tY})$$
$$= E(e^{tX}) \cdot E(e^{tY})$$ since X & Y are independent

$$M_{X+Y}(t) = M_X(t) \cdot M_Y(t)$$
$$= e^{\mu_1 t + \frac{\sigma_1^2 t^2}{2}} \cdot e^{\mu_2 t + \frac{\sigma_2^2 t^2}{2}}$$

$$M_{X+Y}(t) = e^{(\mu_1+\mu_2)t + \frac{(\sigma_1^2+\sigma_2^2)t^2}{2}}$$

$$X+Y \sim N(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2)$$

If X & Y are continuous random variables with joint pdf $f_{X,Y}(x,y)$

$$E(\phi(X,Y)) = \iint \phi(x,y) f_{X,Y}(x,y)\,dx\,dy$$

If X & Y are independent

$$E(\phi(X,Y)) = \iint \phi(x,y) \cdot f_X(x) \cdot f_Y(y)\,dx\,dy$$

If X & Y are independent and $\phi(X,Y) = \phi_1(X) \cdot \phi_2(Y)$

$$E(\phi(X,Y)) = \iint \phi_1(X) \cdot \phi_2(Y) \cdot f_X(x) \cdot f_Y(y)\,dx\,dy$$
$$= \int \phi_1(x) f_X(x)\,dx \cdot \int \phi_2(y) f_Y(y)\,dy$$
$$= E\{\phi_1(x)\} \cdot E\{\phi_2(y)\}$$

Scanned by CamScanner

Generalize the result:

1. $X_i \sim exp(\lambda)$ are iid for $i = 1, 2, 3, \ldots, n$

$$\sum_{i=1}^{n} X_i \sim gamma(n, \lambda)$$

2. $X_i \sim gamma(\alpha_i, \lambda)$ are independent for $i = 1, 2, 3, \ldots, n$

$$\sum_{i=1}^{n} X_i \sim gamma\left(\sum_{i=1}^{n} \alpha_i, \lambda\right)$$

3. $X_i \sim N(\mu_i, \sigma$

Distribution on quotients:

Let X & Y be two continuous random variables with joint pdf $f_{XY}(x,y)$ or $f(x,y)$.

What is density for $Z = Y/X$.

$$A_Z = \{(x,y) \mid y/x \leq z\}$$

If $x < 0$, then $y/x \leq z \iff y \geq xz$

$A_Z = \{(x,y) \mid x < 0 \text{ and } y \geq xz\} \cup \{(x,y) \mid x > 0, y \leq$

$$F_{Y/X}(z) = \iint_{A_Z} f(x,y)\, dx\, dy$$

Scanned by CamScanner

$$F_{Y/X}(z) = \int_{-\infty}^{0}\left[\int_{xz}^{\infty} f(x,y)\,dy\right]dx + \int_{0}^{\infty}\left[\int_{-\infty}^{xz} f(x,y)\,dy\right]dx$$

Substitute $y = xv$ ($dy = x\,dv$) in the inner integral

$$F_{Y/X}(z) = \int_{-\infty}^{0}\left(\int_{z}^{-\infty} x f(x, xv)\,dv\right)dx + \int_{0}^{\infty}\left[\int_{-\infty}^{z} x f(x, xv)\,dv\right]dx$$

$$F_{Y/X}(z) = \int_{-\infty}^{z}\boxed{\left[\int_{-\infty}^{\infty} |x| f(x, xv)\,dx\right]}dv$$

$$F_{Y/X}(z) = \int_{-\infty}^{\infty} |x| f(x, xv)\,dx$$

For X & Y, non-negative and independent.

$$F_{Y/X}(z) = \int_{0}^{\infty} x f_X(x) f_Y(xv)\,dx \qquad 0 < z < \infty$$

Ex:

Let X & Y be independent random variables with densities $\Gamma(\alpha, \lambda)$ and $\Gamma(\alpha_2, \lambda)$ respectively.

Prove:
$$f_{Y/X}(z) = \frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma\alpha_1 \cdot \Gamma\alpha_2} \cdot \frac{z^{\alpha_2 - 1}}{(z+1)^{\alpha_1+\alpha_2}} \qquad 0 < z < \infty$$

$$= 0 \qquad \text{elsewhere}$$

Scanned by CamScanner

Recall:

$$f_X(x) = \frac{\lambda^{\alpha_1} \cdot x^{\alpha_1 - 1} e^{-\lambda x}}{\Gamma(\alpha_1)} \qquad x > 0$$

$$f_Y(y) = \frac{\lambda^{\alpha_2} \cdot y^{\alpha_2 - 1} \cdot e^{-\lambda y}}{\Gamma(\alpha_2)} \qquad y > 0$$

$$f_{Y/X}(z) = \frac{\lambda^{\alpha_1+\alpha_2}}{\Gamma\alpha_1 \cdot \Gamma\alpha_2} \cdot z^{\alpha_2 - 1} \int_0^\infty x^{\alpha_1+\alpha_2-1} e^{-x\lambda(z+1)} \, dx$$

$$= \frac{\lambda^{\alpha_1+\alpha_2} \cdot z^{\alpha_2-1}}{\Gamma\alpha_1 \cdot \Gamma\alpha_2} \cdot \frac{\Gamma\alpha_1+\alpha_2}{[\lambda(z+1)]^{\alpha_1+\alpha_2}}$$

Ex: Let X & Y be independent $N(0,\sigma^2)$ random variables. Find the density of Y^2/X^2.

$$X^2, Y^2 \sim \Gamma\left(\frac{1}{2}, \frac{1}{2\sigma^2}\right)$$

$$f_{Y^2/X^2}(z) = \frac{1}{\pi(z+1)\sqrt{z}} \qquad 0 < z < \infty$$

$$= 0 \qquad \text{otherwise}$$

Homework:

X & Y are independent $N(0,\sigma^2)$ random variables. Find the density of Y/X^2. ?

Conditional Densities:

Let (X, Y) be a discrete random vector.

Conditional probability

$$P(Y=y \mid X=x) = \frac{P(X=x, Y=y)}{P(X=x)} = \frac{f(x,y)}{f_X(x)}$$

where $f(x,y)$ is joint pmf of (X, Y) and $f_X(x)$ is the marginal pmf of X.

Definition: Let X and Y be continuous random variable with joint pdf $f(x,y)$. Then the conditional density of Y given X, denoted as $f_{Y/X}(Y|x)$ is defined as

$$f_{Y/X}(y|x) = \begin{cases} \frac{f(x,y)}{f_X(x)} & 0 < f_X(x) < \infty \\ 0 & \text{otherwise} \end{cases}$$

$$P(a \le Y \le b / X=x) = \int_a^b f_{Y/X}(y/x)\, dy$$

Also, observe

$$f(x,y) = f_X(x) \cdot f_{Y/X}(y/x)$$

If X & Y are independent,

$$f_{Y/X}(y/x) = f_Y(y)$$

Scanned by CamScanner

Example:

$$f(x,y) = \frac{\sqrt{3}}{4\pi} e^{-(x^2 - xy + y^2)/2} \qquad -\infty < x, y < \infty$$

$$X \sim N(0, 4/3)$$

$$f_{Y/X}(y/x) = \frac{f(x,y)}{f_X(x)}$$

$$= \frac{\frac{\sqrt{3}}{4\pi} e^{-(x^2 - xy + y^2)/2}}{\frac{\sqrt{3}}{2\sqrt{2}\pi} e^{-3x^2/8}}$$

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

Thus $f_{Y/X}(y|x)$ is $N\left(\frac{x}{2}, 1\right)$

$\text{Prob}(0 \le Y \le 2 \mid X = 0) = \Phi(2) - \Phi(0)$
where $\Phi(x)$ is CDF of $N(0,1)$

$\text{Prob}(0 \le Y \le 2 \mid X = 2) = 2\Phi(1) - 1$

Example:
Let $X \sim U[0,1]$ and the random variable $Y \sim U[0, X]$ Find joint density of X, Y and marginal density of Y.

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

Scanned by CamScanner

$$f_{Y|X}(y|x) = \begin{cases} 1/x & 0<y<x<1 \\ 0 & \text{otherwise} \end{cases}$$

$$f(x,y) = f_X(x) \cdot f_{Y|X}(y|x)$$

$$= \begin{cases} 1/x & 0<y<x<1 \\ 0 & \text{elsewhere} \end{cases}$$

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx = \int_{y}^{1} \frac{1}{x}\,dx$$

$$f_Y(y) = \begin{cases} -\log y & 0<y<1 \\ 0 & \text{elsewhere} \end{cases}$$

Baye's rule:

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)}$$

$$= \frac{f_X(x) \cdot f_{Y|X}(y|x)}{\int_{-\infty}^{\infty} f_X(x) \cdot f_{Y|X}(y|x)\,dx}$$

Scanned by CamScanner

Bivariate Normal Density:

Random vector (X_1, X_2) is said to follow bivariate normal density if its joint pdf is given by

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{\frac{-1}{2(1-\rho^2)}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right) + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right]\right\}$$

$$-\infty < x_1 < \infty$$
$$-\infty < x_2 < \infty$$

Probability Computation:

$$P(a_1 \le X_1 \le b_1, a_2 \le X_2 \le b_2) = \int_{a_1}^{b_1}\int_{a_2}^{b_2} f(x_1, x_2)\, dx_1\, dx_2$$

Range of parameter

$$-\infty < \mu_1 < \infty$$
$$-\infty < \mu_2 < \infty$$
$$\sigma_1 > 0, \quad \sigma_2 > 0$$
$$-1 < \rho < 1$$

The Marginal Densities:

$$f_{X_1}(x_1) = \int_{-\infty}^{\infty} f(x_1, x_2)\, dx_2$$

$$= \frac{1}{\sigma_1\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2} \qquad -\infty < x_1 < \infty$$

Marginal density of X_1 is $N(\mu_1, \sigma_1^2)$

$$f_{X_2}(x_2) = \int_{-\infty}^{\infty} f(x_1, x_2)\, dx_1$$

$$= \frac{1}{\sigma_2\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x_2-\mu_2}{\sigma_2}\right)^2}$$

$-\infty < x_2 < \infty$

Marginal Density of X_2 is $N(\mu_2, \sigma_2^2)$:

$E(X_1) = \mu_1$, $var(X_1) = \sigma_1^2$
$E(X_2) = \mu_2$, $var(X_2) = \sigma_2^2$

$$cov(X_1, X_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x_1-\mu_1)(x_2-\mu_2)\, f(x_1,x_2)\, dx_1\, dx_2$$

$$= E(X_1-\mu_1)(X_2-\mu_2)$$

Then

$$\rho = \frac{cov(X_1, X_2)}{\sigma_1 \cdot \sigma_2}$$

In general, independence $\Rightarrow$ uncorrelatedness and converse is <u>not</u> true.

In case of bivariate normal density, if X_1 & X_2 are uncorrelated $(\rho = 0) \Rightarrow X_1$ & X_2 are independent
(joint = product of marginals)

The conditional densities:

$$f_{X_1|X_2} = \frac{f(x_1, x_2)}{f_{X_2}(x_2)}$$

$$= \frac{1}{\sigma_1\sqrt{2\pi}\sqrt{1-\rho^2}} \cdot \exp\left[\frac{-1}{2\sigma_1^2(1-\rho^2)} \left\{x_1 - \left(\mu_1 + \rho\left(\frac{\sigma_1}{\sigma_2}\right)(x_2-\mu_2)\right)\right\}^2\right]$$

$$f_{X_2|X_1}(x_2|x_1) = \frac{f(x_1, x_2)}{f_{X_1}(x_1)}$$

$$= \frac{1}{\sigma_2\sqrt{2\pi}\sqrt{1-\rho^2}} \exp\left\{ \frac{-1}{2\sigma_2^2(1-\rho^2)} \left\{ x_2 - \left[\mu_2 + \rho\left(\frac{\sigma_2}{\sigma_1}\right) \times (x_1-\mu_1)\right]\right\}\right\}$$

Thus, conditional densities $X_2 | X_1$ and $X_1 | X_2$ are both normal densities.

$$X_2 | X_1 = x_1 \sim N\left(\mu_2 + \rho\left(\frac{\sigma_2}{\sigma_1}\right)(x_1 - \mu_1),\ \sigma_2^2(1-\rho^2)\right)$$

$$X_1 | X_2 = x_2 \sim N\left(\mu_1 + \rho\left(\frac{\sigma_1}{\sigma_2}\right)(x_2 - \mu_2),\ \sigma_1^2(1-\rho^2)\right)$$

Example:

(X_1, X_2) is a bivariate normal random variable with parameter:

$\mu_1 = 0.2$, $\mu_2 = 1100$, $\sigma_1^2 = 0.02$, $\sigma_2^2 = 525$, $\rho = 0.9$

compute $E(X_2 | X_1)$

$$E(X_2 | X_1) = \mu_2 + \rho\left(\frac{\sigma_2}{\sigma_1}\right)(x_1 - \mu_1)$$

$$= 1100 + 0.9\left(\frac{\sqrt{525}}{\sqrt{0.02}}\right)(x_1 - 0.2)$$

$$= 145.8x_1 + 1070.84$$

$$E(X_2 | X_1 = 1) = 145.8(1) + 1070.84$$

Scanned by CamScanner

$P(X_2 \geq 1080 / X_1 = 1)$

$$Var(X_2 | X_1 = 1) = \sigma_2^2 (1 - \rho^2) = 525(1 - 0.81) = 99.75$$

Note that $Y = X_2 | X_1 = 1$ is a Normal density with mean μ_Y and variance σ_Y^2 given by

$$\mu_Y = 1216.64 \quad \sigma_Y^2 = 99.75$$

$$P(X_2 \geq 1080 | X_1 = 1) = P(Y \geq 1080)$$

$$= P\left(\frac{Y - \mu_Y}{\sigma_Y} \geq \frac{1080 - \mu_Y}{\sigma_Y}\right)$$

$$= P\left(Z \geq \frac{1080 - 1216.64}{\sqrt{99.75}}\right) = P(Z \geq -13.6)$$

$$= 1 - \Phi(-13.6) = \Phi(13.6) \approx 1$$

Change of Variable Formula:

Let $X_1, X_2, \ldots, X_n$ be random variables with joint PDF $f(x_1, x_2, \ldots, x_n)$

Define $\varphi_i = \sum_{j=1}^{n} a_{ij} X_j \qquad i = 1, 2, \ldots, n$

$$A = [a_{ij}]$$

$$= \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ a_{21} & & a_{2n} \\ \vdots & & \vdots \\ a_{n1} & & a_{nn} \end{bmatrix}$$ then clearly,

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = A \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix}$$

If $|A| \neq 0$, then $B = A^{-1}$

$$\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix} = B \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \qquad X_i = \sum_{j=1}^{n} b_{ij} Y_j$$

$i = 1, 2, 3, \dots, n$

$$\boxed{f(y_1, y_2, \dots, y_n) = \frac{1}{|\det(A)|} f(x_1, x_2, \dots, x_n)}$$

Non-linear Transformation case:-

$$Y_i = g_i(X_1, X_2, \dots, X_n) \; ; \; i = 1, 2, 3, \dots, n$$

$$J(x_1, x_2, \dots, x_n) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \dots & \frac{\partial f_1}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial f_n}{\partial x_1} & & \frac{\partial f_n}{\partial x_n} \end{bmatrix}$$

$$f(y_1, y_2, \dots, y_n) = \frac{1}{|J(x_1, x_2, \dots, x_n)|} f(x_1, x_2, x_3, \dots, x_n)$$

Example:

Let $X_1, X_2, X_3, \dots, X_n$ be iid rvs $\exp(\lambda)$.

$Y_i = X_1 + X_2 + \dots + X_i \qquad i = 1, 2, 3, \dots, n$

$$Y_1 = X_1$$
$$Y_2 = X_1 + X_2$$
$$\vdots$$
$$Y_n = X_1 + X_2 + \dots + X_n$$

$$A = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 0 & \cdots & 0 \\ 1 & 1 & 1 & \cdots & 0 \\ 1 & \vdots & \vdots & & \vdots \\ 1 & 1 & 1 & \cdots & 0 \end{bmatrix}$$

$$|\det(A)| = 1$$

Joint density of Y_i's is same as the joint density of X_i's in this case

Summary:

Joint rvs

discrete
joint pdfs
↓
marginals
conditionals
independence

continuous
joint distributions
↓
joint pdfs
marginals
conditionals
independence

(Valid for both)
[Mean, Median, Mode] – (Bivarite Normal)
Mgfs
→ Conditional Expectations and variances
Covariance and correlation

Transformations
- distribution of sums
convolution / mgf
- Distribution of quotients

$\Rightarrow \mathbb{R}^2 \to \mathbb{R}$

$X_1, \dots, X_n \longrightarrow X_1 + X_2 + \cdots + X_4$

$X_1, X_2 \longrightarrow X_1 / X_2$

$\left. \varphi_i = g_i(x_1, x_2, \dots, x_n) \; ; i = 1, 2, 3, \dots, n \right] \Rightarrow R^n \rightarrow R^n$

Sampling Distributions:

1. Standard Normal : $N(0,1)$:

x^2 random variable with n degrees of freedom

Let $X_1, X_2, \dots, X_n$ be iid $\sim N(0, \sigma^2)$ { Normal random variable has additive property

$$\Rightarrow \frac{X_1 + X_2 + \cdots + X_n}{n} \sim N\left(0, \frac{\sigma^2}{n}\right)$$

$$\frac{X_1 + X_2 + \cdots + X_n}{\sigma\sqrt{n}} = \left(\frac{X_1 + \cdots + X_n}{n}\right) \Big/ (\sigma/\sqrt{n})$$

$$\downarrow$$

$$N(0,1) \qquad = \frac{N(\mu, \sigma^2) - \mu}{\sigma}$$

In particular

$$\frac{X_i}{\sigma} \sim N(0,1)$$

$$\Rightarrow \frac{X_i^2}{\sigma^2} \sim \Gamma\left(\frac{1}{2}, \frac{1}{2}\right) \qquad i = 1, 2, \dots, n$$

From the additive property of $\Gamma(\alpha_i, \lambda)$

We know. thus,

$$\frac{X_1^2}{\sigma^2} + \frac{X_2^2}{\sigma^2} + \cdots + \frac{X_n^2}{\sigma^2} \sim \underbrace{\Gamma\left(\frac{n}{2}, \frac{1}{2}\right)}$$

x^2 density with n degree of freedom

Scanned by CamScanner

$n \rightarrow$ how many squared r.v.s have been added to obtain the random variable.

deg. of freedom = no. of random variables (normal) participating in the sum.

Notice: χ_n^2 is the sum of squares of 'n' independent standard normal variables.

Theorem:
Let $Y_1, Y_2, \ldots, Y_n$ be the independent χ^2 random variable with deg. of freedom $K_1, K_2, \ldots, K_m$ respectively

Then $Y_1 + Y_2 + Y_3 + \cdots + Y_m \sim \chi^2_{K_1} + \cdots + \chi^2_{K_m}$

Ratio of χ^2 densities:

$$Y_1 \sim \chi^2_{K_1} \qquad Y_2 \sim \chi^2_{K_2}$$

and Y_1 & Y_2 are independent.

Then the random variable defined by the ratio $(Y_1/K_1)/(Y_2/K_2)$ is called an 'F' distributed random variable with (K_1, K_2) degrees of freedom.

$$Y_1 \sim \Gamma\left(\frac{K_1}{2}, \frac{1}{2}\right)$$

The density of Y_1/Y_2 is denoted as $F(K_1, K_2)$

$$Y_2 \sim \Gamma\left(\frac{K_2}{2}, \frac{1}{2}\right)$$

Scanned by CamScanner

Particular Case: $k_1 = 1 \quad = F(1, k_2)$

$f(1, k_2) = $ square of standard normal
(sum of squares of std. normal
divided by dof)

t- distribution:

Let X be the standard normal random variable and Y be χ^2 random variable with n dof and X and Y are independent of each other.

$$X \sim N(0,1) \qquad Y \sim \chi_n^2$$

Then $X / \sqrt{Y/n}$ is said to follow t- distribution with 'n' degrees of freedom.

Clearly: $t_n^2 = F(1, n)$

Central Limit Theorem:

Let $X_1, X_2, \ldots, X_n$ be iid with μ and variance (σ^2)
To study the distribution $\sum_{i=1}^{n} X_i = S_n$

Notice: $E(S_n) = n\mu \qquad var(S_n) = n\sigma^2$

Suppose X_1 has density f, then

$f_{S_n}(x) = \sum_y f_{S_{n-1}}(y) f(x-y)$ if X_i's are discrete

$f_{S_n}(x) = \int_{-\infty}^{\infty} f_{S_{n-1}}(y) f(x-y)\, dy$ if X_i's are continuous

Note:

$$S_n^* = \frac{S_n - ES_n}{\sqrt{var(S_n)}} = \frac{S_n - n\mu}{\sigma\sqrt{n}}$$

$$\sim N(0,1)$$

Scanned by CamScanner

Then, $\lim_{n\to\infty} P\left[\frac{S_n - n\mu}{\sigma\sqrt{n}} \le x\right] = \Phi(x) \qquad -\infty < x < \infty$

where $\Phi(x)$ is the cdf of $N(0,1)$

Asymptotically, the cdfs are matching.

Note that $P\left[\frac{S_n - n\mu}{\sigma\sqrt{n}} \le x\right] = F_{S_n^*}(x) = \text{CDF of } S_n^*$

$\Rightarrow$ CLT states

$$\lim_{n\to\infty} F_{S_n^*}(x) = \Phi(x) \qquad -\infty < x < \infty$$

Observe: $P(S_n \le x) \approx \Phi\left(\frac{x - n\mu}{\sigma\sqrt{n}}\right) = \Phi\left(\frac{x - ES_n}{\sqrt{var(S_n)}}\right)$

$$P\left[\frac{S_n - n\mu}{\sigma\sqrt{n}} \le \frac{x - n\mu}{\sigma\sqrt{n}}\right]$$

generally $n \ge 25$ these approximations are really very good.

Example:

Let $X_i \sim exp(1)$ for $i = 1, 2, \ldots, n$

$S_n = X_1 + X_2 + \cdots + X_n$

$Prob(S_n \le x) \simeq \Phi\left(\frac{x-n}{\sqrt{n}}\right)$

Ex: Suppose life of a bulb after it is installed follows exponential distribution with mean = 10 days. As soon as the bulb burns out another bulb with same characteristics is installed. What is the probability that 50 bulbs are required in one year?

Sol: X_i = Life of i^{th} bulb after it is installed

$X_i \sim exp(1/10) \qquad S_{50} = X_1 + X_2 + \cdots + X_{50}$

$$P(S_{50} \le 365) \approx \Phi\left(\frac{\overset{E(S_n)}{365 - 500}}{\underset{\sqrt{var(S_n)}}{\sqrt{5000}}}\right)$$

The probability of interest

$= \Phi(-1.91)$

$P(S_{50} \le 365) = 0.028$

Scanned by CamScanner

$$P(S_n \le x) \approx \Phi\left(\frac{x-\mu}{\sigma\sqrt{n}}\right) \quad \forall\ x$$

Let f_{S_n} be pdf of S_n, then

$$f_{S_n} \approx \frac{1}{\sigma\sqrt{n}}\,\phi\left(\frac{x-\mu n}{\sigma\sqrt{n}}\right) \qquad -\infty < x < \infty$$

Φ: CDF of $n(0,1)$

φ: Φ' pdf of $n(0,1)$

Ex:

$$X_i \sim \exp(1) \quad \text{for} \quad i = 1, 2, 3, \ldots, n$$
iid

$$S_n = X_1 + X_2 + \cdots + X_n$$

$$f_{S_n}(x) \approx \frac{1}{\sqrt{n}}\,\phi\left(\frac{x-n}{\sqrt{n}}\right) \qquad -\infty < x < \infty$$

Approximation for discrete random variable:

$$f_{S_n}(x) = \text{Prob}(S_n = x) > 0$$
then x is a possible value of S_n

- x is a possible value of X, then x is an integer
- If '$a \in \mathbb{R}$' is a possible value of X, then the set
  $\{ x - a \mid x \text{ is a possible value of } X\}$ has gcd 1.

Scanned by CamScanner

Ruling out r.v.s. $P(X_1 = 1) = P(X_1 = 3) = \frac{1}{2}$

Example:

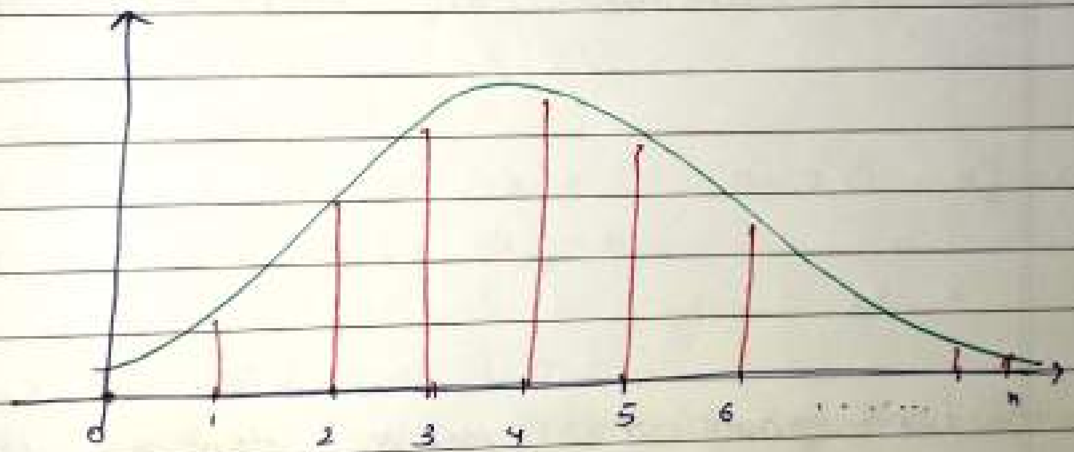
Let $X_1 \sim \text{Bernouli}(p)$; $\mu = p$, $\sigma^2 = p(1-p)$

$X_1, X_2, \ldots, X_n$ are iid Bernouli(p).

$S_n = X_1 + X_2 + \cdots + X_n \sim \text{Binomial}(n, p)$

$$f_{S_n}(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

$$\approx \frac{1}{\sqrt{np(1-p)}} \, \phi\left(\frac{x - np}{\sqrt{np(1-p)}}\right)$$



Another approximation

$$\boxed{f_{S_n}(x) = \Phi\left(\frac{x + \frac{1}{2} - n\mu}{\sigma\sqrt{n}}\right) - \Phi\left(\frac{x - \frac{1}{2} - n\mu}{\sigma\sqrt{n}}\right)}$$

$$\text{Prob}(S_n \le x) = \Phi\left(\frac{x + \frac{1}{2} - n\mu}{\sigma\sqrt{n}}\right)$$

Scanned by CamScanner

Ex: $S_n \sim Bin(n, p)$ $n = 25$, $p = 0.6$

$P(S_n \geq 13) = 1 - P(S_n < 13) = 1 - P(S_n \leq 12)$

$$= 1 - \Phi\left(\frac{12 + \frac{1}{2} - 25 \times 0.6}{\sqrt{25(0.6)(0.4)}}\right)$$

Scanned by CamScanner

Statistics

Statistical Inference (Estimation and testing of hypothesis)

Let $X_1, X_2, \ldots, X_n$ be a random sample from $f_\theta(\cdot)$ and $F_\theta(\cdot)$ where $f_\theta(\cdot)$ is the pdf and F_θ is the C.D.F. of X_i's.

($X_1, X_2, \ldots, X_n$ a random sample means $X_1, X_2, X_3, \ldots, X_n$ are i.i.d random variables with density $f_\theta(\cdot)$)

Ex:

Let $X_1, X_2, \ldots, X_{10}$ be a random sample from Bernoulli(p) (Useful for statistical properties)

What do we observe?

$X_1 = 1$, $X_2 = 1$, $X_3 = 0$, $X_4 = 1$, $X_5 = 0$, $X_6 = 0$, $X_7 = 0$, $X_8 = 0$, $X_9 = 1$, $X_{10} = 0$

→ 1, 1, 0, 1, 0, 0, 0, 0, 1, 0

Data available from repetitions of the random experiment.

Question: Can you guess 'p' from the evidence/data

$$\hat{p} = \frac{\sum X_i}{n}$$

Note:

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

Scanned by CamScanner

$\overline{X}$ is a random variable and $\sum_{i=1}^{n} x_i/n$ is a realization of $\overline{X}$.

For different set of repetitions of the random expt. one would get different values of $\hat{p} = \sum_{i=1}^{n} x_i/n$. The variation in the values of $\hat{p}$ can be understood from the density of $\overline{X}$.

In this particular case, $X_1 + X_2 + \cdots + X_n \sim$ Binomial(n, p) and then notice,

$$E(\overline{X}) = p = E\left(\frac{X_1 + X_2 + \cdots + X_n}{n}\right) = \frac{np}{n}$$

$$Var(\overline{X}) = Var\left(\frac{X_1 + X_2 + \cdots + X_n}{n}\right) = \frac{1}{n^2} \cdot np(1-p)$$

$$Var(\overline{X}) = \frac{p(1-p)}{n}$$

Few definitions:

$X_1, X_2, \ldots, X_n$ a random sample with pdf $f_\theta()$

Def 1: Family of Distributions

$$\mathcal{F} = \{ f_\theta(\cdot) \mid f_\theta \text{ is pdf of } X \text{ and } \theta \in \Theta$$

e.g.
Let $X_1, \ldots, X_n$ be a random sample from Binomial (n, p).

$$B = \left\{ \binom{n}{x} p^x (1-p)^{n-x} \;\middle|\; n \in \mathbb{N}, p \in [0,1] \right\}$$

Scanned by CamScanner

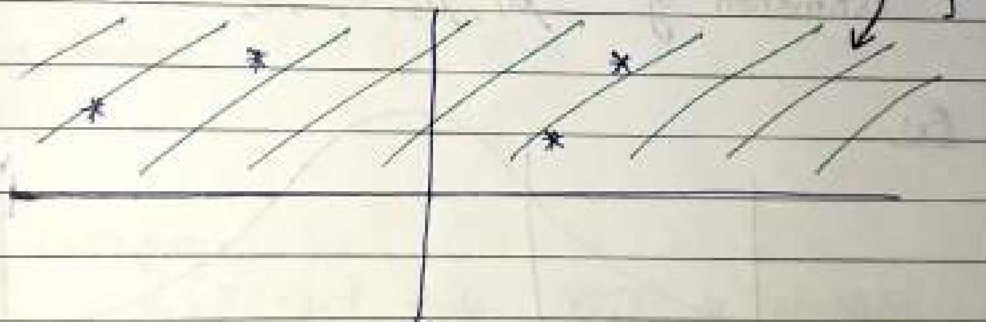
Here $\theta = [n, p]$: parameter vector

$\Theta = \mathbb{N} \times [0, 1]$: parameter space

Example 2:

Let $X_1, X_2, \ldots, X_n$ be a random sample from $N(\mu, \sigma^2)$.

$$N = \left\{ \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \;\middle|\; (\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}_+ \right\}$$



Note: Any vector denoted as * •, in the parameter space is a valid guess

Parametric estimation:

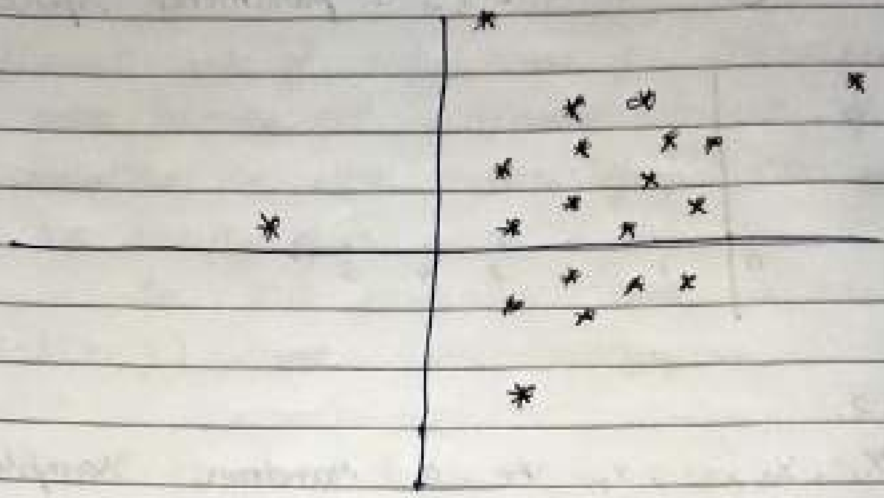
In this problem we want to identify the parameter values of the distribution from the given data set.

Non- parametric estimation:

In this problem we want to identify pdf or cdf directly from the data.

Examples:

Data in a 2-D plane represented by *.

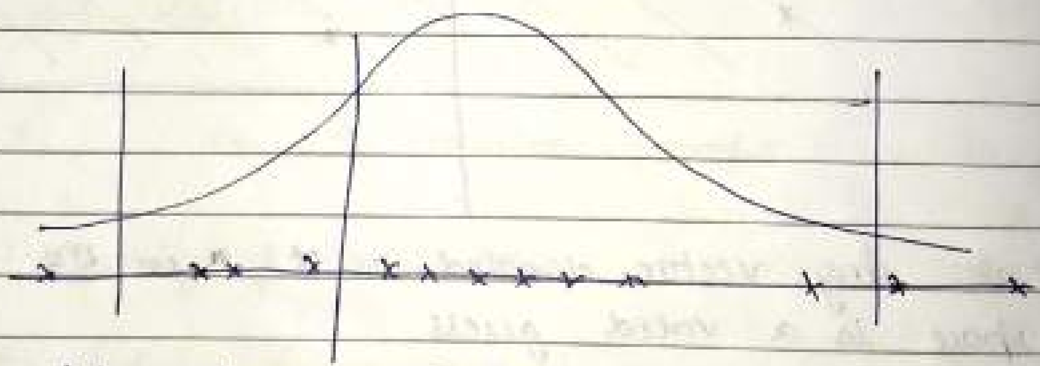


Data comes from bivariate normal density.

Parametric estimation:

Estimation of $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2$ and ρ

Ex:



Data 1-D, represented by *

parametric estimation (normal density) parameters to be estimated : μ, σ^2

Ex: Non-parametric estimation, Data 2-D.

Scanned by CamScanner

Point estimation in parametric set up:

1) parameter estimation (already understood)
2) Definition of estimator and estimate
3) Good properties of an estimator.
4) Methods of estimation

(MOM / MLE)

Definition: Statistic

A statistic is a function of data but it is free from any unknown parameter.

Exs: Let $X_1, X_2, \dots, X_n$ be a random sample from $N(\mu, \sigma^2)$

① Let μ be known

② μ is unknown

$$\sum_{i=1}^{n} (X_i - \mu)^2 \text{ is } \underline{\text{NOT}} \text{ a statistic}$$

However, $\sum_{i=1}^{n} (X_i - \bar{X})^2$ is a statistic.

③ Assume, μ, σ^2 unknown.

$$\frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{\sigma^2} \text{ is } \underline{\text{NOT}} \text{ a statistic.}$$

Let $X_1, X_2, \dots, X_n \sim$ (iid) or (independent identically distributed random variable) $f_\theta(\cdot)$

Scanned by CamScanner

where $f_\theta \sim \mathcal{F} = \{f_\theta / \theta \in \Theta\}$

Objective: We may be interested to

Ex: Let $X_1, X_2, \ldots, X_n$ be a random sample from $N(\mu, \sigma^2)$. Here $\theta = [\mu, \sigma^2]$: parameter vector

$g(\mu, \sigma^2) = \mu$ (mean)
$g(\mu, \sigma^2) = \sigma^2$ (variance)
$g(\mu, \sigma^2) = \sigma^2/\mu$ if $\mu \neq 0$ (coefficient of variance)

Ex: Let $X_1, X_2, \ldots, X_n$ be a random sample from $G(\alpha, \lambda)$.

$g(\alpha, \lambda) = \alpha/\lambda$ $g(\alpha, \lambda) = \alpha/\lambda^2$ $g(\alpha, \lambda) = \lambda$

Definition

Estimator: Let $X_1, X_2, \ldots, X_n$ be a random sample from $f_\theta(\cdot)$
A statistic $T(X_1, \ldots, X_n)$ when used to estimate a parameter function $g(\theta)$ is called an estimator of $g(\theta)$.

Note: $T(X_1, X_2, \ldots, X_n)$ being a function of random variable is itself a random variable.

Ex: Recall the exampl. of collecting data by tossing a coin 'n' times and then estimate the probability of success (p).

Let $X_1, X_2, \dots, X_n$ be a random sample from Bernouli(p). Here parameter $\theta = p$.

$g(\theta) = p$, $T(X_1, X_2, \dots, X_n) = \overline{X}$ is an 'estimator of p'

Estimate:

The value of $T(X_1, \dots, X_n)$ evaluated at the given data points $X_1 = x_1, X_2 = x_2, \dots, X_n = x_n$ is called as an estimate.

Examples of estimators:

Let $X_1, X_2, \dots, X_n$ be a random sample from $f_\theta(\cdot)$ with $E(X_i) = \mu$ $Var(X_i) = \sigma^2$ for $i = 1, 2, \dots, n$

Then μ can be estimated by

i) $T(X_1, X_2, \dots, X_n) = X_1$

ii) $T(X_1, X_2, \dots, X_n) = [5X_3 + 7X_9]/12$

iii) $T(X_1, X_2, \dots, X_n) = \sum_{i=1}^{n} a_i X_i$

where $\sum_{i=1}^{n} a_i = 1$, $a_i \in [0,1]$

iv) $T(X_1, X_2, \dots, X_n) = \overline{X}$

Ex: σ^2 can be estimated by

$$S_1^2(X_1, X_2, \dots, X_n) = \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X})^2$$

$$S_2^2(X_1, X_2, \dots, X_n) = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2$$

Scanned by CamScanner

Properties of estimators:

Unbaised estimator: Let $T(X_1, X_2, \ldots, X_n)$ be an estimator of $g(\theta)$ which satisfies

$$E(T(X_1, X_2, \ldots, X_n)) = g(\theta) \quad \forall\ \theta \in \Theta$$

then $T(X_1, \ldots, X_n)$ is said to be an unbaised estimator of $g(\theta)$.

Ex: Let $X_1, X_2, \ldots, X_n$ be a random sample from $f_\theta(\cdot)$.

Let $\mu = E(X_i)$ for $i = 1, 2, \ldots, n$

$\sigma^2 = Var(X_i)$

We need to estimate μ.

Let $f_\theta = N(\mu, \sigma^2)$

1. $T(X_1, X_2, \ldots, X_n) = X_1$: unbaised estimator

$$E(T(X_1, \ldots, X_n)) = E(X_1) = \mu$$

2. $T(X_1, \ldots, X_n) = \bar{X}$

$$E[T(X_1, \ldots, X_n)] = E(\bar{X}) = E\left(\frac{X_1 + X_2 + \cdots + X_n}{n}\right)$$

$$= \frac{\mu + \cdots + \mu}{n} = \frac{n\mu}{n} = \mu$$

3. $T(X_1, X_2, \ldots, X_n) = 3X_1$

$$E(T(X_1, X_2, \ldots, X_n)) = 3\mu \neq \mu \quad \text{(baised estimator)}$$

unless $\mu = 0$

Definition: Mean squared Error (MSE). Let $g(\theta)$ be a function of parameters which is estimated by an estimator $T(X_1, X_2, \ldots X_n)$. Then the MSE of the estimator for $g(\theta)$ is defined as $MSE(T(X_1, X_2, \ldots X_n))$ $= E(T(X_1, X_2, \ldots, X_n) - g(\theta))^2$

Scanned by CamScanner

$MSE(T(X_1, X_2, \ldots, X_n))$
$$= var(T(X_1, X_2, \ldots, X_n)) + [Bias(T(X_1, \ldots, X_n))]^2$$

Estimation:

- Estimator / statistic
- unbiasedness

Methods of Estimation:

1) Method of Moments

observations / data / random sample

$\{ X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n \}$ from the distribution $f_\theta(\cdot)$ Given

We want to estimate the parameters $\theta_1, \theta_2, \ldots, \theta_k$ where $\theta = [\theta_1, \theta_2, \ldots, \theta_k]$ is the parameter vector ?

Algo: Compute

1. theoretical moments from the pdf $f_\theta(\cdot)$
2. Empirical moments from the data

Example:

Question: Let $X_1, X_2, \ldots, X_n$ be a random sample from Gamma (α, λ).
Using method of moments, estimate α & λ.

Solution:

theoretical moments

Scanned by CamScanner

mean of Gamma(α, λ) = α/λ
Variance of Gamma(α, λ) = α/λ^2

2) Empirical Moments from the data sample of mean:
$$\bar{X} = \sum_{i=1}^{n} x_i / n$$

$$\text{sample variance} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

MOM estimators $\Rightarrow$
$$\frac{\alpha}{\lambda} = \bar{X} \quad ; \quad \frac{\alpha}{\lambda^2} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

<u>Question:</u>
Let 0.3, 0.01, 2.89, 4.49, 0.64 be a random sample from Gamma(α, λ) Using method of moments estimate α & λ

<u>Solution:</u>

$$\bar{x} = 1.65 \quad = \alpha/\lambda$$

$$\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 = 1.713 = \alpha/\lambda^2$$

$$\frac{\alpha}{\lambda^2} = \frac{\alpha}{\lambda}\cdot\frac{1}{\lambda} = 1.713$$

$$\Rightarrow 1.65 \times \frac{1}{\lambda} = 1.78$$

$$\Rightarrow \lambda = \frac{1.713}{1.65} = 1.038$$

$$\Rightarrow \lambda = \frac{1}{1.038} = 0.96$$

Scanned by CamScanner

Example

Q. Let 0.3, 0.1, 0.7, −2.1, 0.8 be a random sample from $N(\mu, \sigma^2)$. Estimate μ & σ^2 using method of moments.

Mean of $N(\mu, \sigma^2) = \mu$

Variance of $N(\mu, \sigma^2) = \sigma^2$

$$\mu = \bar{X} \quad , \quad \sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

$$\hat{\mu} = 0.12 \qquad \hat{\sigma}^2 = 1.58$$

Example:

Q. Let 1, 0, 0, 1, 1, 1, 0, 1, 1, 1 be a random sample from Bernoulli(p). Use MOM to estimate p.

$$\hat{p} = 0.7$$

Theoretical moments should exist to use MOM.

Maximum Likelihood Estimation:

Let $X_1, X_2, X_3, \ldots, X_n$ be a random sample from $f_\theta(\cdot)$ then the joint density of $X_1, X_2, \ldots, X_n$ is given by $f_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} f_\theta(x_i)$

Consider a case when $f_\theta(\cdot)$ is discrete

$(1-p)^x p \qquad x = 0, 1, 2, 3, 4, \ldots$

The likelihood function:

$$l(\theta) = \prod_{i=1}^{n} f_\theta(x_i) = \prod_{i=1}^{n} \Sigma\, f_\theta(x_i)$$

Scanned by CamScanner

Maximum likelihood estimate of θ,

$$\hat{\theta}_{MLE} = \arg\max_{\theta \in \Theta} l(\theta)$$

$$= \cancel{arg} \arg\max_{\theta \in \Theta} \log(l(\theta))$$

Examples:

Let $X_1, X_2, \ldots, X_n$ be a random sample from $N(\mu, 1)$. Find the MLE of μ.

$$l(\mu) = \prod_{i=1}^{n} f_\theta(x_i)$$

where $f_\theta(x_i) = N(\mu, 1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x_i-\mu)^2}$

$$l(\mu) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x_i-\mu)^2} \qquad -\infty < x_i < \infty$$

$$= \frac{1}{(2\pi)^{n/2}} \cdot e^{-\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2}$$

$$\log(l(\mu)) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2$$

Here maximizing $\log(l(\mu))$ is equivalent to minimizing sum of squares as shown below

$$\arg\max_{\mu} \log(l(\mu))$$

$$= \arg\min_{\mu} \sum_{i=1}^{n}(x_i-\mu)^2$$

Scanned by CamScanner

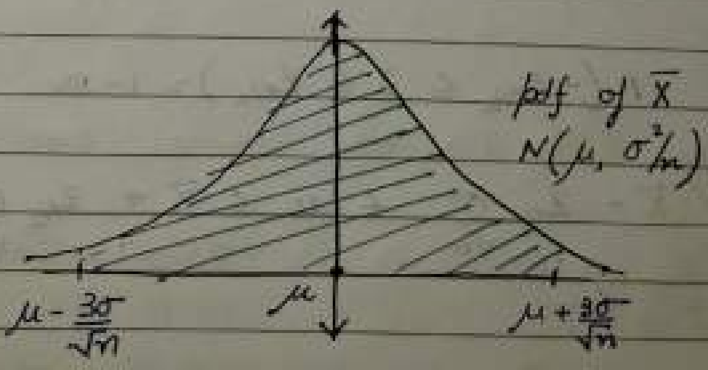
classmate
Date
Page

cost function = $\sum_{i=1}^{n} (x_i - \mu)^2$

$$\frac{d}{d\mu} \sum_{i=1}^{n} (x_i - \mu)^2 = -2 \sum_{i=1}^{n} (x_i - \mu) = 0$$

$$\Rightarrow \sum_{i=1}^{n} x_i = n\mu \Rightarrow \boxed{\hat{\mu} = \frac{\sum_{i=1}^{n} x_i}{n}}$$

Let $x_1, x_2, \ldots, x_n$

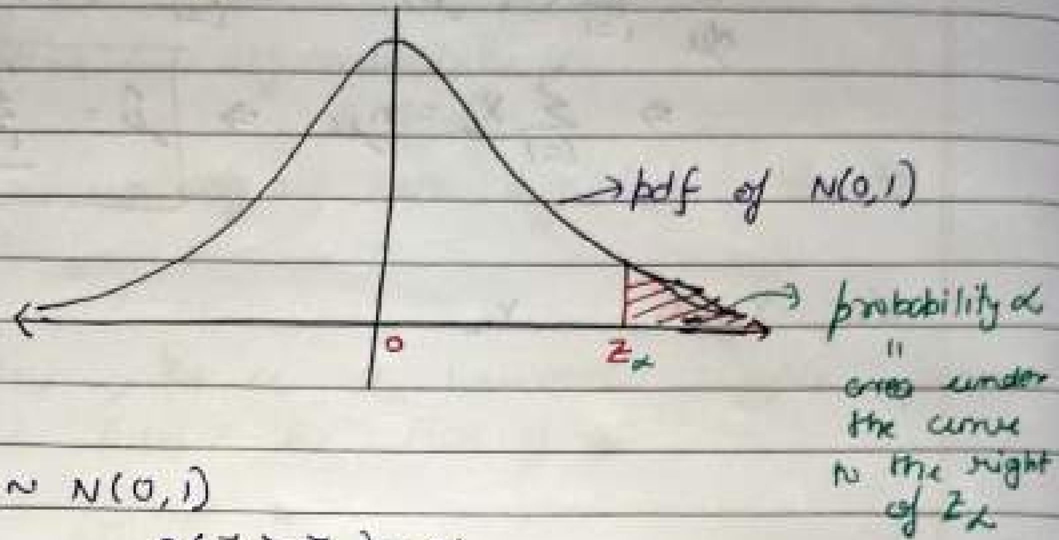


Scanned by CamScanner

classmate
Date
Page

3 s.d. rule of normal density

$$P\left(\mu - \frac{3\sigma}{\sqrt{n}} \le \bar{X} \le \mu + \frac{3\sigma}{\sqrt{n}}\right) \simeq 0.997$$

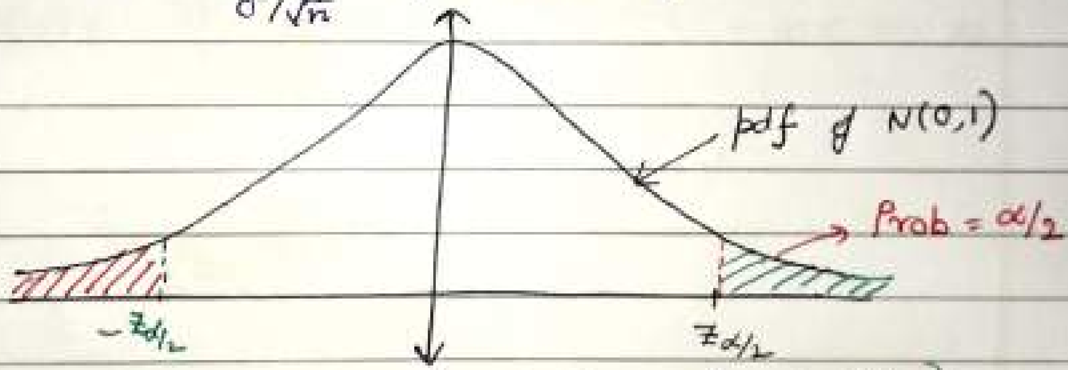


Let $Z \sim N(0,1)$

$$P(Z \ge z_\alpha) = \alpha$$

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$



It is clear that: $\left.\begin{array}{l} P(Z \ge z_{1-\alpha/2}) = 1 - \alpha/2 \\ \Rightarrow P(Z \le z_{1-\alpha/2}) = \alpha/2 \\ \Rightarrow P(Z \le -z_{\alpha/2}) = \alpha/2 \end{array}\right\}$ due to symmetry of N(0,1)

Thus, $P(-z_{\alpha/2} \le Z \le z_{\alpha/2}) = 1 - \alpha$

$$\Rightarrow P\left(\bar{X} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \le \mu \le \bar{X} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

Scanned by CamScanner

Take $\alpha = 0.1$ $\quad \alpha/2 = 0.05$ $\quad Z_{\alpha/2} = 1.65$

$$P\left(\bar{x} - 1.65\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + 1.65\frac{\sigma}{\sqrt{n}}\right) = 0.9$$

$100(1-\alpha)\%$ of confidence interval on μ when σ^2 is known for $N(\mu, \sigma^2)$

$$L = \bar{x} - Z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}} \qquad U = \bar{x} + Z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}$$

Example:

Let $0.3, -1.8, 1.9, 0.01, -0.3, 0.1$ be a random sample from $N(\mu, 1)$. Find 90% confidence interval on μ.

Soln: $1-\alpha = 0.9 \Rightarrow \alpha = 0.1 \quad \alpha/2 = 0.05$

$$Z_{\alpha/2} = 1.65$$

Here $\sigma = 1$, known $\quad n = 6 \quad \bar{x} = 0.035$

$$L = \bar{x} - Z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}$$

$$= 0.035 - 1.65 \times \frac{1}{\sqrt{6}} = -0.64$$

$$U = \bar{x} + Z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}$$

$$= 0.035 + 1.65 \times \frac{1}{\sqrt{6}} = 0.71$$

$$C.I. \rightarrow [-0.64, 0.71]$$

Exercise:

For the same sample compute 99% confidence interval on μ.

$1-\alpha = 0.99$

$\alpha = 0.01$ $\qquad P(Z \le Z_{\alpha/2}) = 0.995$

$\alpha/2 = 0.005$ $\qquad Z_{\alpha/2} = 2.58$

Scanned by CamScanner

$$L = \bar{X} - Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} = 0.035 - 2.58 \times \frac{1}{\sqrt{6}} = -1.04$$

$$U = 0.035 + 2.58 \times \frac{1}{\sqrt{6}} = 1.09$$

$$C.I = [-1.04, 1.09]$$

length of the CI

$$L = \bar{X} - Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \qquad U = \bar{X} + Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

$$U - L = 2 \cdot Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

Let $X_1, X_2, \ldots, X_n$ be a random sample from $N(\mu, \sigma^2)$

<u>Assumption:</u> μ & σ^2 both are unknown

$$\frac{X_i - \mu}{\sigma} \sim N(0,1) \quad \text{for } i = 1, 2, \ldots, n$$

$$\left(\frac{X_i - \mu}{\sigma}\right)^2 \sim \Gamma\left(\frac{1}{2}, \frac{1}{2}\right) = \chi^2_1 \quad \text{for } i = 1, 2, \ldots, n$$

$$\sum_{i=1}^{n} \frac{(X_i - \mu)^2}{\sigma^2} \sim \Gamma\left(\frac{n}{2}, \frac{1}{2}\right) = \chi^2_n$$

Consider

$$\sum_{i=1}^{n} (X_i - \mu)^2 = \sum_{i=1}^{n} \left[(X_i - \bar{X}) + (\bar{X} - \mu)\right]^2$$

$$= \sum_{i=1}^{n} (X_i - \bar{X})^2 + \sum_{i=1}^{n} (\bar{X} - \mu)^2 + \cancel{2 \sum_{i=1}^{n} (X_i - \bar{X})(\bar{X} - \mu)}^{0}$$

$$\sum_{i=1}^{n} (X_i - \mu)^2 = \sum_{i=1}^{n} (X_i - \bar{X})^2 + \sum_{i=1}^{n} (\bar{X} - \mu)^2$$

$$\sum_{i=1}^{n} (X_i - \mu)^2 = \sum_{i=1}^{n} (X_i - \bar{X})^2 + n(\bar{X} - \mu)^2$$

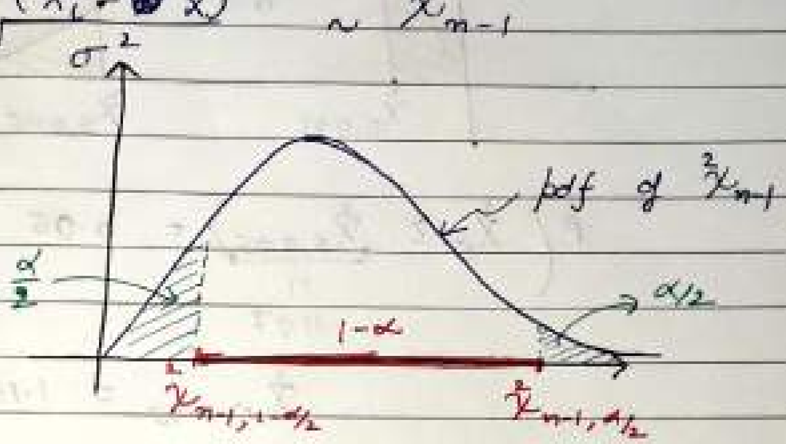
Scanned by CamScanner

$$\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\sigma^2} = \frac{\sum_{i=1}^{n}(X_i-\overline{X})^2}{\sigma^2} + \frac{(\overline{X}-\mu)^2}{\sigma^2/n}$$

$$\sim \qquad\qquad \sim \qquad\qquad \sim$$

$$\chi^2_n \qquad\qquad \chi^2_{n-1} \qquad\qquad \chi^2_n$$

Thus, $\dfrac{\sum_{i=1}^{n}(X_i-\overline{X})^2}{\sigma^2} \sim \chi^2_{n-1}$



Percent points

$$P(\chi^2_{n-1} \geq \chi^2_{n-1,\alpha/2}) = \alpha/2$$

$$P(\chi^2_{n-1} \geq \chi^2_{n-1,1-\alpha/2}) = 1-\alpha/2$$

$$P(\chi^2_{n-1} \leq \chi^2_{n-1,1-\alpha/2}) = \alpha/2$$

Thus,

$$P(\chi^2_{n-1,1-\alpha/2} \leq \chi^2_{n-1} \leq \chi^2_{n-1,\alpha/2}) = 1-\alpha$$

$$\Rightarrow P\left(\chi^2_{n-1,1-\alpha/2} \leq \frac{\sum_{i=1}^{n}(X_i-\overline{X})^2}{\sigma^2} \leq \chi^2_{n-1,\alpha/2}\right) = 1-\alpha$$

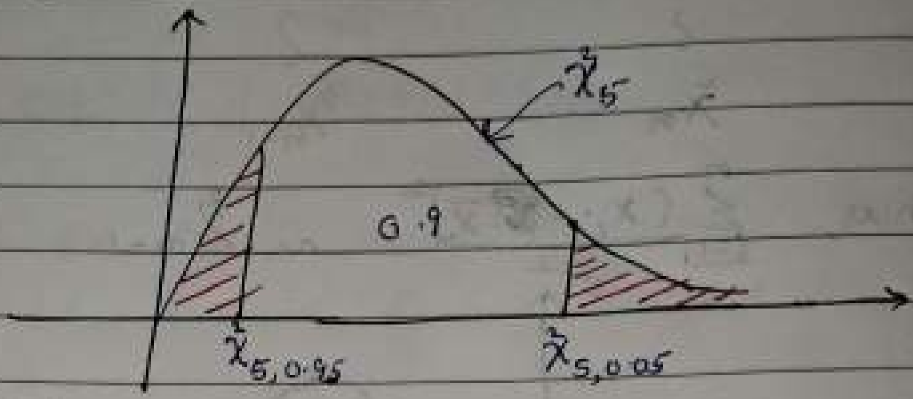
$$\Rightarrow P\left(\frac{\sum_{i=1}^{n}(X_i-\overline{X})^2}{\chi^2_{n-1,\alpha/2}} \leq \sigma^2 \leq \frac{\sum_{i=1}^{n}(X_i-\overline{X})^2}{\chi^2_{n-1,1-\alpha/2}}\right) = 1-\alpha$$

Example: Let 0.3, −1.8, 1.9, 0.01, −0.1, 0.1 be a random sample from $N(\mu, \sigma^2)$. Construct 90% confidence interval for σ^2.

Scanned by CamScanner

Soln: $1-\alpha = 0.9$ $\alpha = 0.1$ $\alpha/2 = 0.05$

Here $n = 6$



$$P\left(\chi^2_5 \geq \underbrace{\chi^2_{5,0.05}}_{= 11.07}\right) = 0.05$$

$$\chi^2_{5,0.95} = 1.15$$

$$L = \frac{\sum_{i=1}^{6} (x_i - \bar{x})^2}{11.07} = 0.635$$

$$U = \frac{\sum_{i=1}^{6} (x_i - \bar{x})^2}{1.15} = 6.12$$

Confidence interval on μ when σ^2 is unknown for $N(\mu, \sigma^2)$ population.

Scanned by CamScanner