classmate
Date
Page

- **Random Experiment:**

An experiment is known as a random exp. if (I) it has more than one outcome

(II) the outcome of a particular trial is not known in advance.

(III) it can be repeated contably many time in identical conditions.

Example:

(I) tossing a coin

(II) Rolling a die.

(III) random arrangement of 52 cards.

- **Sample Space:**

Collection of all possible outcomes of a random experiment.

(I) $\Omega = \{H, T\}$

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

(III) $\Omega = \{\pi \mid \pi$ is a permutation of 52 cards$\}$

$\hookrightarrow |\Omega| = 52!$

cardinality

Scanned by CamScanner

classmate
Date
Page

* Classical definition of Probability:

If the sample space (Ω) of a random experiment is finite and $A \in \Omega$ then probability of occurence of A is defined as:

$$P(A) = \frac{|A|}{|\Omega|}$$ → cardinality

Ex: $a_1 + a_2 + a_3 + \dots + a_r = n$, $r < n$
such that $a_i \in \mathbb{N} \cup \{0\}$.
What is the ~~prob~~ probability that all $a_i \in \mathbb{N}$.

$$\Rightarrow \frac{{}^{n+r-1}C_r - \sum_{p=1}^{n} {}^{r+p-1}C_p}{{}^{n+r-1}C_{r-1}} = \frac{{}^{n-1}C_{r-1}}{{}^{n+r-1}C_{r-1}}$$

Solution:

$\rightarrow A : \sum_{i=1}^{r} a_i = n$, $a_i \in \mathbb{N}$

$$\underbrace{1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1}_{n}$$

out of $(n-1)$ gaps collect $(r-1)$ plus sign

$$A : {}^{n-1}C_{r-1}$$

$\Omega : \sum_{i=1}^{r} a_i = n$, $a_i \in \mathbb{N} \cup \{0\}$

$|\Omega| = \left({}^{n+r-1}C_{r-1}\right)$ which is possible number of arrangement of '$r-1$' '+' symbols and n many '1' symbols. Such that

Scanned by CamScanner

(i) If '+' symbol comes in the beginning, then $a_1 = 0$ or in the end then $a_r = 0$ and if there is no '1' symbol in between two '+' symbols then consider $a_i = 0$.
Rest are same as counting for A

$$P(A) = \frac{{}^{n-1}C_{r-1}}{{}^{n+r-1}C_{r-1}}$$

Ex: what is the probability that any randomly chosen natural number is an even number?
⇒

Here comes the

Frequency definition of Probability.

If Ω is the sample space and A is a subset of Ω such that $|\Omega|$ is not finite then consider Ω_n and A_n satisfying:

$$\lim_{n\to\infty} \Omega_n = \Omega \quad \text{and} \quad \lim_{n\to\infty} A_n = A$$

then

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

NOW, $\Omega_{n_0} = \{1, 2, \ldots\ldots, n_0\}$ $\quad n_0 \in N$

$$A_{n_0} = \{2, 4, \ldots\ldots \left[\frac{n_0}{2}\right] \times 2\}$$

$$P(A) = \lim_{n_0\to\infty} \frac{|A_{n_0}|}{|\Omega_{n_0}|} = \frac{1}{2}$$

Scanned by CamScanner

classmate
Date
Page

- Algebra :

A collection $\mathcal{A}$ of subset of Ω is called as algebra if :

(I) $\Omega \in \mathcal{A}$

(II) any $A \subset \Omega$ and $A \in \mathcal{A}$ then
$A^c \in \mathcal{A}$

(III) If $A_1, A_2, \ldots, A_n \in \mathcal{A}$ and they are subsets of Ω then

$$\bigcup_{i=1}^{n} A_i \in \mathcal{A}$$

- σ- Algebra :

An Algebra $\mathcal{A}$ of Ω is said to be a σ-algebra or σ-field, if $A_1, A_2, \ldots A_n \in \mathcal{A}$ and are subsets of then

$$\lim_{n \to \infty} \bigcup_{i=1}^{n} A_i = \bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$$

★ Conditional Probability :-

→ "Probability of A given B occurs".

→ The probability changes on applying condition, it may increase or decrease.

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

then conditional probability of A given B:

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

Read as "Probability of A given B"

Example: R: Rolling a fair die
A: 2 appears on the top face of the die
B: even number appears on the top face.

then, $P(A/B) = \frac{\# A}{\# B} = \frac{1}{3}$ | $A = \{2\}$, $B = \{2, 4, 6\}$

Actually it is: $\frac{P(A \cap B)}{P(B)} \Rightarrow \frac{\# A \cap B}{\# B}$ here

Since $A \cap B = A$ here

Note:
"Conditional probab. is nothing" but alteration of sample space $\mathcal{P}(\Omega)$. That's it!!

Scanned by CamScanner

classmate
Date
Page

* Product Rule:

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

Note:
If $P(A) > 0$

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

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

(to be calculated)

So, $\boxed{\underbrace{P(B|A)}_{\text{likelihood}} \underbrace{P(A)}_{\text{Prior}} = \underbrace{P(A|B)}_{\text{posterior}} \cdot \underbrace{P(B)}_{\text{Normalization factor}}}$ ∴ Bayes theorem / Rule

→ Bayesian Inference is based on this!

or, $\boxed{P(A|B) = \frac{P(B|A)\, P(A)}{P(B)}}$

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

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

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

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

↑ Disjoint union

So, $P(A) = P(A \cap B_1) + P(A \cap B_2) + \ldots + P(A \cap B_n)$

Scanned by CamScanner

Using product Rule:

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

or, $$\boxed{P(A) = \sum_{i=1}^{n} P(A|B_i)\cdot P(B_i)}$$

Total law of probability

Note:

Modified Baye's Rule:

$$P(B_j|A) = \frac{P(A|B_j)\cdot P(B_j)}{\sum_{i=1}^{n} P(A|B_i)\cdot P(B_i)}$$

→ It's so worse!

→ But very useful!

Ex:

① grade vs Attendance statistics

☆ Independance of events:

Conditional probability of A given B is

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

let's understand with an example.

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

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

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

$\Rightarrow$ $P(A_2 / A_1) = \frac{1}{3}$, $P(A_2) = 1/4$

$P(B_1 / B_2) = 2/3$, $P(B_1) = 3/4$

$P(A) = 1/2$, $P(A/B) = 1/2$, $P(B) = 1/2$

$P(B/A) = 1/2$

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

Note:

Pairwise Independance of events :-

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

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

$P(C / A \cap B) = 1$. — leads to the definition of mutually Independance

$\downarrow$

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

$\downarrow$

(A, B), (B, C), (C, A) should be pairwise independent.

- **Pairwise Independent:**

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

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

- **Mutually Independent:**

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

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

Note:

We can genralise the definition of mutual independence to n-events $A_1, A_2, \ldots, A_n$ in an inductive way:

- Any subcollection of $A_1, \ldots, A_n$, containing atleast 2 and atmost $(n-1)$ event, is mutually independent
- $P(A_1 \cap A_2 \cap A_3 \cap \ldots \cap A_n) = P(A_1) \cdot P(A_2) \ldots P(A_n)$

- **Application** -

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

Scanned by CamScanner

Experiment: Tossing a coin : } R

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

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

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

NOW we set up a probability space when one repeats R 'n' number of times with the assumption that each event in independent from other.

For this compound experiment :

Job 1 $\Omega = \{ y_1 y_2 \cdots y_n \mid y_1, y_2, \ldots, y_n \in \{0,1\} \}$

Job 2 $\mathcal{F} \rightarrow$ Set of all the events to which probability can be assigned. $\rightarrow$ Powerset of Ω

Job 3 lets now pick a sequence from Ω and assign probability to it. (Probability Assignment)

- Probability of K successes in n repetitions at specified locations.

Consider an element $\omega \in \Omega$

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

here $y_1, y_2, \ldots y_K = 1$; $y_{K+1}, y_{K+2}, \ldots y_n = 0$

$\therefore \omega = (\underbrace{1111 \cdots 1}_{K \text{ times}} \underbrace{000 \cdots 0}_{(n-K) \text{ times}})$

Bernoulli trial

we define $A_i :=$ Success at i^{th} toss. i for $i = 1, 2, \ldots n$

NOW, we can say that all events are pairwise independent from the assumptions.

$$P(\omega) = P(y_1 y_2 \ldots y_n)$$

$$= P(A_1 \cap A_2 \cap A_3 \ldots \cap A_k \cap A^c_{k+1} \cap A^c_{k+2} \ldots \cap A^c_n)$$

$$= P(A_1) \cdot P(A_2) \ldots P(A_k) \cdot P(A^c_{k+1}) \cdot P(A^c_{k+2}) \ldots P($$

→ mutually independence from Assumption

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

Note:

In the above steps, A_i were mutually independent but we applied the rule ~~of~~ to $(A_1 \ldots A_k)$ and $(A^c_{k+1} \ldots A^c_n)$.

Q. Does that hold for compliment also ??
A. Yup! It does.

We can prove using A, B :
if $P(A \cap B) = P(A) \cdot P(B)$
does $P(A \cap B^c) = P(A) \cdot P(B^c)$ hold?

- Now we can find the probability of exactly k successes in n coin tosses using $P(\omega)$

$$^nC_k \cdot P(\omega) = \binom{n}{k} p^k (1-p)^{n-k}$$

find the k-locations

classmate
Date
Page

Ex: A box containing 10 balls

6 – red
4 – blue

what is the probability that first ball is blue and second is red?

$\therefore \quad P(A_1/A_2) = \dfrac{P(A_1 \cap A_2)}{P(A_2)}$

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

A_1 = red ball in first trial
A_2 = blue ball in second trial

$$P(A_2|A_1) = \frac{P(A_2 \cap A_1)}{P(A_1)} = \frac{P(A_2)\cdot P(A_1)}{P(A_1)} = P(A_2) = 0.4$$

→ Sampling with Replacement

Case(2) : Pick a ball, donot put it back in the box & pick a ball again

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

Thus, A_1 & A_2 are not independent.

Increase 'n' and keep the proportions of red balls and blue balls same.
→ A_1 & A_2 start behaving as independent events!!

Scanned by CamScanner

Uniform probability spaces :-

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

- Given set is finite and discrete :

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

- While $[0,1]$ is not finite.
  → Probability of every singleton element is zero. Because if we assign each of the some non-zero probability, the sum of all the probabilities shall reach ∞. So, all are zero.

→ Uniform propability Principle -

Probability of two subsets of Ω with equal size be equal.

Case(1):

Here we associate size or cardinality of the set.

As a consequence, we can associate probability of $\left(\frac{1}{n}\right)$ to each singleton set.

Case(2): The given set Ω is finite and an interval of R

$\Omega = [0,1]$

Scanned by CamScanner

Here, 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 $(b-a)$.

$\therefore P((a,b)) = \underline{(b-a)} \quad \forall (a,b) \in [0,1]$

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

This is the continuous uniform probability space.

Note:

1) P(Singleton set) $= 0$.

2) Consider the set of rational numbers in $[0,1]$ denoted as $\mathbb{Q} \cap \{0,1\}$

Then from (1) and the fact that $\mathbb{Q} \cap \{0,1\}$ is countable,

$$\boxed{P(\mathbb{Q} \cap \{0,1\}) = 0}$$

Note:

If we can come up with a bijective function from set of natural number to a set, then that set is a countable set.

Scanned by CamScanner

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

⇒ $1 - 0.1 \times 0.2 \times 0.3 \times 0.4$

$1 - 0.0024$

⇒ 0.9976

from independency of component failure

$A_1 \cap A_2 \cap A_3 \cap A_4$ = Machine fails

∴ required probab. $= 1 - P(A_1 \cap A_2 \cap A_3 \cap A_4)$

$= 1 - P(A_1) \cdot P(A_2) P(A_3) \cdot P(A_4)$

Ex: 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 min. no. of components are req. in the parallel assembly so that the engine has 99% of sccen probability?

⇒

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

$(0.05)^n = 0.01$, $\boxed{n = 2}$

Concept of independence

2 - | 12

Ex: In a course students are allowed to register after the teacher consents. It is observed that 20% times after obtaining the consent, student fails to register. There are only 100 seats available in the class. If a teacher consented 102 students, what is the probability that all the students will be accomodated in the class?

$\Rightarrow$ $1 - (0.8)^{102} - {}^{102}C_1 (0.2)(0.8)^{101}$

Ex: $P(P_3 / \text{fault}) = \dfrac{P(P_3 \text{ and fails})}{P(\text{fault})}$ ③

$=$

Ex: (I) $P(6 / \text{one has appeared}) = 1/6$.

1

2 (II) $P(>6) = \left(\frac{21}{36}\right) \Rightarrow 1/36 \Rightarrow \frac{1}{21}$

$2x + \frac{5}{36} = 1$

(3,3)
(5,1),
(4,2)

$x = \left(\frac{31}{72}\right)$

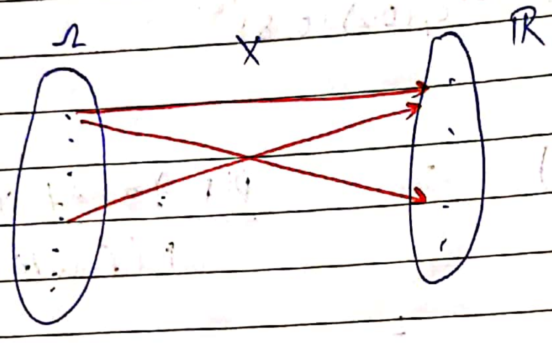
classmate

Date
Page

★ Random Experiment :-

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

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



$\forall x \in \mathbb{R}$, $\underbrace{\{\omega : X(\omega) = x\}}_{\subseteq \Omega} = X^{-1}(\omega)$

inverse image of X

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

$P(X=x) = P\{\omega \mid X(\omega) = x\}$

- 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.

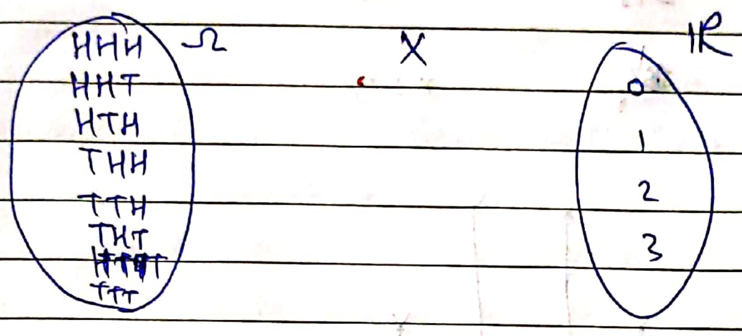
Expt: 3 fair coins tossed independently

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

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

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

X: no. of heads in 3 tosses



$x = 2 \qquad X^{-1}(2) = \{\omega \in \Omega \mid X(\omega) = 2\}$
$\qquad\qquad\qquad\qquad = \{HHT, HTH, THH\}$

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

$x = 8.7$

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

Here, $R_X = \{0,1,2,3\}$ and hence X is a discrete random variable

classmate
Date
Page

★ Probability mass function:

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

$$f(x) = 1/8 \qquad x = 0 \qquad\qquad f(x) = 0 \;\; \forall \; x = \text{else}$$

$$= 3/8 \qquad x = 2$$

$$= 3/8 \qquad x = 1$$

$$= 1/8 \qquad x = 3$$

Pictorial:

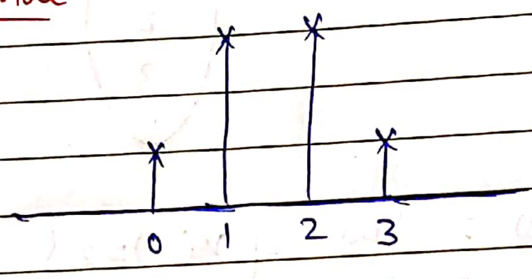


Table:

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

Ex: **Bernoulli Random Variable:**

experiment: Tossing a coin.

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

let's define

$$X(\text{Heads}) = 1, \quad X(\text{tails}) = 0$$

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

classmate
Date
Page

$$f(x) = \begin{cases} 1-p & ; \quad x=0 \\ p & ; \quad x=1 \\ 0 & \text{otherwise} \end{cases} \Bigg\}\ \text{pmf}$$

Representation :-

$X \sim$ Bernoulli(p) $\quad R_X = \{0, 1\}$

$\downarrow$ probability of success.

$\longrightarrow$ X follows Bernoulli p $\downarrow$ parameter of the distribution

Ex: Binomiadl

experiment : n independent Bernoulli trials are performed.

X : number of successes in a trials

$R_X = \{0, 1, 2, \ldots, n\}$

$f(x) = p(X = x)$

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

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

Representation:-

$x \sim$ Binomial (n, p)

Scanned by CamScanner

Ex: Geometric

experiment : Perform independent Bernoulli trials (p) until the first success

$\Omega = \{ S, FS, FSS, FFFS, \ldots \}$

$p \quad (1-p)p \quad (1-p)^2 p \quad (1-p)^3 p$

X: The number of failures before the first successes.

$R_X = \{0, 1, 2, 3, \ldots\}$ Discrete random Variable

$f(x) = p(1-p)^x \qquad x = 0, 1, 2, \ldots$

$= 0 \qquad$ otherwise

Note:

For any random experiment : $P(\Omega) = 1$

let X be a random variable defined in Ω.
let R_X be the range of this random variables.

$$R_X = \{x_1, x_2, x_3, \ldots x_n, \ldots\}$$

(Discrete random variable case)

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

$= 1$

All are ~~disjoin~~ disjoints since f is a function and there should be only one

- Random expt. $\longrightarrow \Omega, \mathcal{F}, P \longrightarrow X \longrightarrow$ pmf

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

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

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

Ex: Uniform pmf:

$$f(x) = \frac{1}{s} \qquad x \in \{x_1, x_2, \ldots x_s\}$$
$$= 0 \qquad \text{otherwise}$$

Ex: Geometric pmf:

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

Ex: Negative Binomial pmf

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

Scanned by CamScanner

experiment : performing independent Bernoulli(p) trials until α number of successes are observed.

For $\alpha = 2$

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

Ex: Poisson density (λ) $(\lambda > 0)$

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

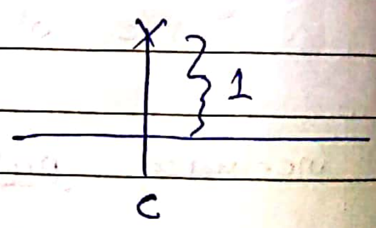
$$= 0 \qquad \text{otherwise}$$

* Special examples of random variable :-

Ex: Constant random variable

$X(\omega) = c \quad \forall\ \omega \in \Omega$

$P(X = c) = 1, \quad P(X \neq c) = 0$



Ex: Indicator random variable

let $A \in \mathcal{F}$

$$X_A(\omega) = 1 \qquad \forall\ \omega \in A$$

$$= 0 \qquad \forall\ \omega \in A^c$$

Scanned by CamScanner

classmate
Date
Page

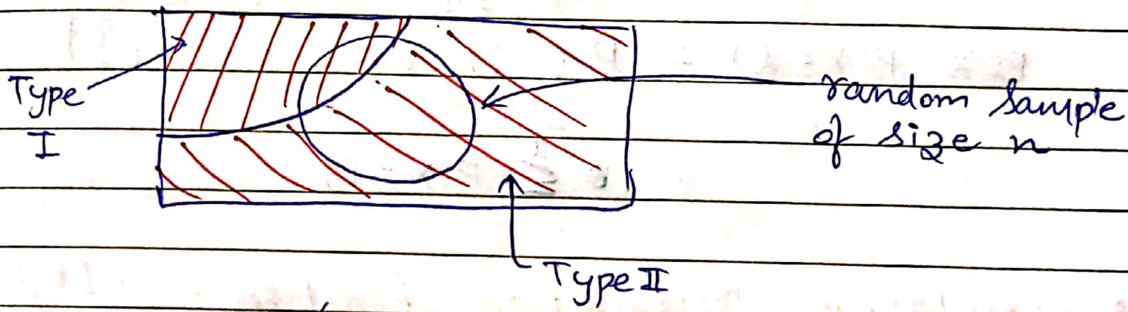
Ex: Hyper geometric distribution:

Population of r objects

Type I objects $\rightarrow r_1$
Type II objects $\rightarrow r_2 = r - r_1$

let a sample of size n is chosen from this population $(n \le r)$

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



$$P(X=x) = \frac{\binom{r_1}{x}\binom{r-r_1}{n-x}}{\binom{r}{n}} \qquad x = 0, 1, 2, 3, \ldots n$$

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

Scanned by CamScanner

* Computations with pmf :-

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

X : Discrete random variable with pmf $f(x)$

Event : $X \le t$ for some real no. t

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

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

↳ Disjoint union

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

$$= \sum_{i=-\infty}^{[t]} P(X = i)$$

• Cumulative Distribution function :- (CDF)

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

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

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

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

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

$= 0$ $\quad$ Otherwise

Scanned by CamScanner

clearly for $t < 0$; $F(t) = 0$

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

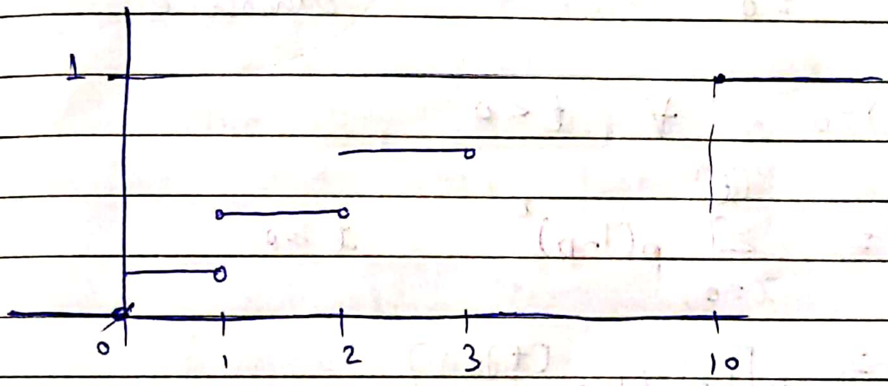
for any ~~t~~ $0 \le t < 1$:

$$F(t) = 1/10$$

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

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

Step function



CDF : let X be a 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 $t_1 < t_2 \Rightarrow F(t_1) \le F(t_2)$
- $F(-\infty) = 0$ and $F(\infty) = 1$

Scanned by CamScanner

classmate
Date
Page

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

- Right continuity of F

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

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

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

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

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

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

$$= 1-(1-p)^{[t]+1}$$

$$\boxed{\therefore F(t) = 1-(1-p)^{[t]+1}} \quad \forall\ t \geq 0$$

- For this geometric random variable:

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

$x-1$

Events

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

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

Disjoint.

$$\boxed{\therefore P(X \geq x) = (1-p)^{x}}$$

Prove: $P(X > n+m \mid X > n) = P(X > m)$

memoryless property

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

Note: True for geometric random variable

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

Note:

Example of above property:

Suppose there are 10 failures, now what is the probability that 15 failures will occur.

~ This situation is same as probability of 5 failures from a fresh start.

- Let's extend our study to continuous r.v(s).

Consider, X as 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 r.v. X $(F_X(x))$.

Scanned by CamScanner

$\delta \quad \delta \quad \delta \quad \delta$

- **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 $F(x)$ is a non-decreasing function

Very important → 5. Right continuity:

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

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:**

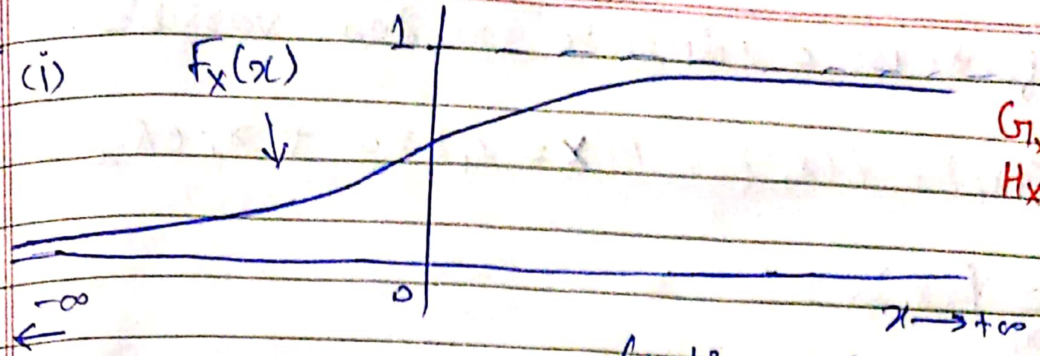
If a function $F_X(x)$ satisfies the properties stated above, then $F_X(x)$ can be represented as sum of two functions say $G_X(x)$ and $H_X(x)$ as

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

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:

(i)

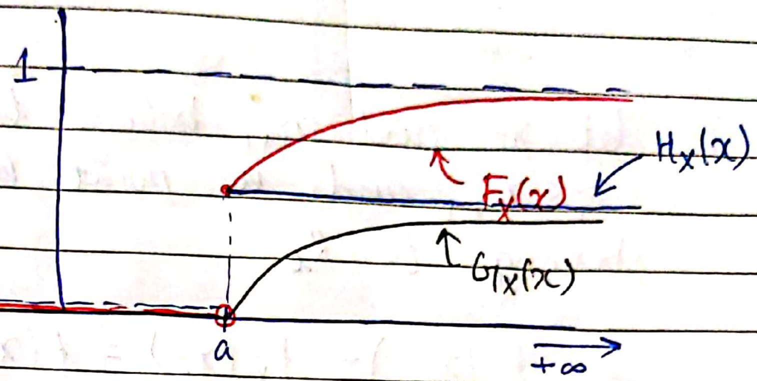


$G_X(x) = F_X(x)$
$H_X(x) = 0$

Continuous Variable

(ii)

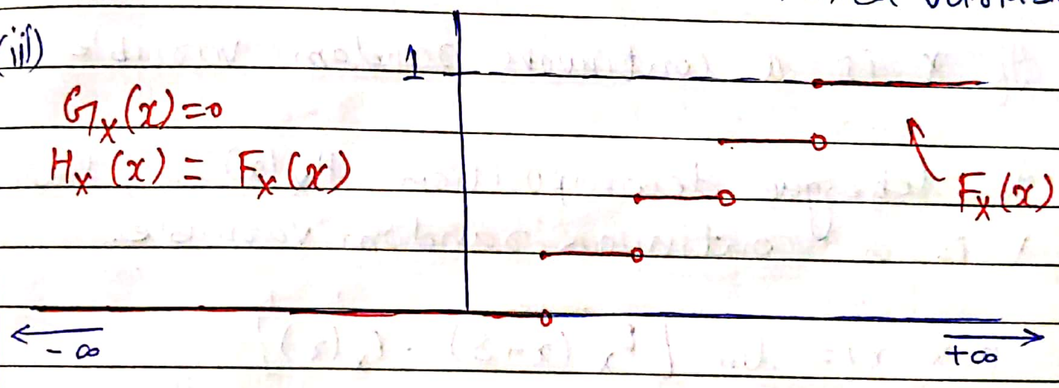
$F_X(x) = \underbrace{G_X(x)}_{\text{Some}} + \underbrace{H_X(x)}_{\text{Some}}$



Mixed Variable

(iii)

$G_X(x) = 0$
$H_X(x) = F_X(x)$



Discrete variable

Note:

From above discussion, we conclude :

→ In Lebesgue decomposition if $G_X(x) \equiv 0$ (identically) for every $x \in \mathbb{R}$, then the random variable X is called as discrete random variable.

If $H_X(x) \equiv 0$ (identically) then X is called as continuous random variable.

→ Otherwise, X is called mixed r. variable.

** If X is a discrete random variable:

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

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

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

** If X is a continuous random variable

In lebesgue decomposition $H_X(x) \equiv 0$, then X is a continuous random variable.

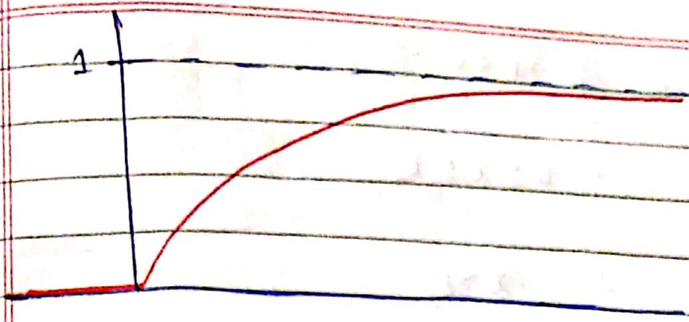
$$P(X = x) = \lim_{\delta \to 0} \left[F_X(x+\delta) - F_X(x) \right]$$

$$= 0$$

Here, the probability for each X is zero but still probability space exists.

→ Same as probability of singleton sets in a large set.

Scanned by CamScanner

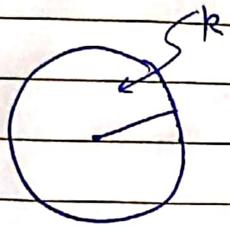


define $f_X(x) = \frac{d}{dx} F_X(x)$

(always exists except for a "few" points)
in the above examples at $x=0$, $f_X(0) = DNE$.

$f_X(x)$: p.d.f.

Ex:



Experiment: Throw dart on this circular ~~pass~~ board.

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

$\rightarrow$ Uniform probability space

let's define

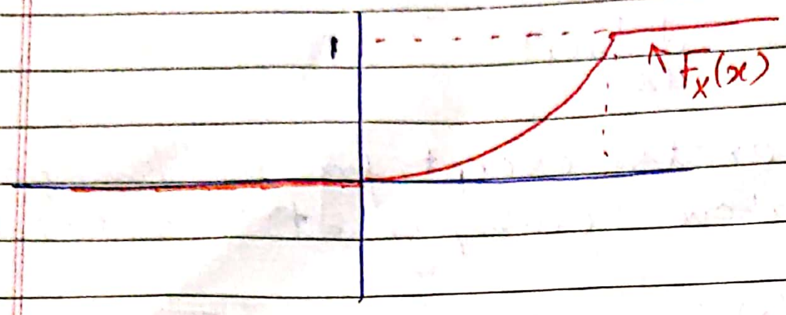
X: Distance of the dart from the Center of the board (bull's eye)

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

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



then pdf

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

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

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

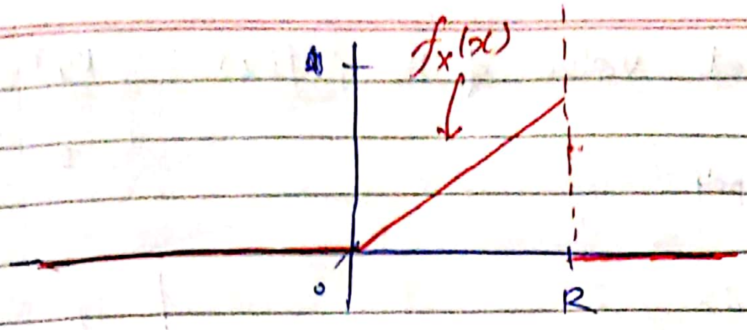
$$P(a \le x \le b) = F_X(b) - F_X(a)$$

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

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

Scanned by CamScanner

classmate
Date
Page



Note:

- To get probability of the variable at any point, we use probability mass function (pmf)

- 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 $f_X(x)$ 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

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

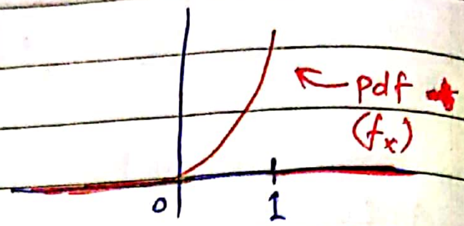
Scanned by CamScanner

Ex: (1) For what value of K, $f(x) = \begin{cases} Kx^3 & 0 \le x \le 1 \\ 0 & \text{Otherwise} \end{cases}$

is a pdf?

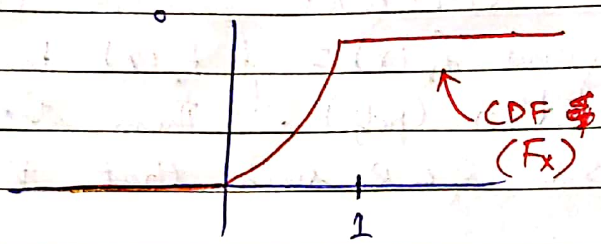
$\Rightarrow \int_0^1 Kx^3 dx = 1$

$\frac{Kx^4}{4}\Big|_0^1 \Rightarrow \boxed{K = 4}$



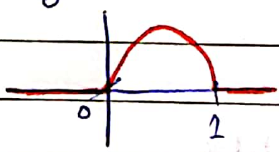
For $y \in (0,1)$

$$F_x(y) = \int_0^y 4x^3 dx = y^4.$$



Ex: $f(x) = \begin{cases} 6x(1-x) & ; \; x \in [0,1) \\ 0 & \text{else} \end{cases}$

$\Rightarrow$ (1) $6\int_0^1 x(1-x)\, dx = 6\left[\frac{x^2}{2} - \frac{x^3}{3}\right]\Big|_0^1$



$\Rightarrow 6\left[\frac{1}{2} - \frac{1}{3}\right] = \underline{\underline{1}}$

Note

PDF value can cross one.

classmate
Date
Page

★ Change of variable

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

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

We know,

$$G(y) = P(Y \le y) \qquad \forall y \in R$$

$$= P(x^2 \le y)$$

$$= P(-\sqrt{y} \le x \le \sqrt{y}) = F_X(\sqrt{y}) - F_X(-\sqrt{y})$$

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

take derivative to get density

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

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

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

Scanned by CamScanner

classmate
Date
Page

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

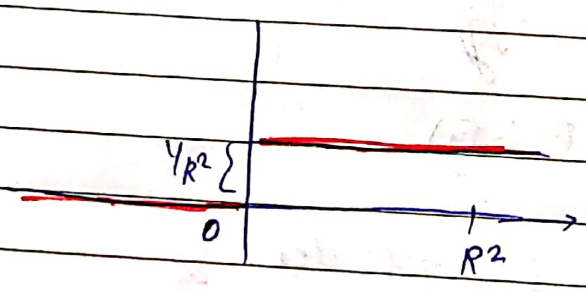
Find the density of $Y = X^2$.

→ let g(y) be the density of Y.

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

$$= \frac{1}{2\sqrt{y}}\left[\frac{2\sqrt{y}}{R^2} + 0\right] \quad \forall \quad y \in [0, R^2]$$

$$g(y) = \begin{cases} 1/R^2 & ; \quad y \in [0, R^2] \\ 0 & ; \quad \text{else} \end{cases}$$



Scanned by CamScanner

☆ Continuous uniform random variable:

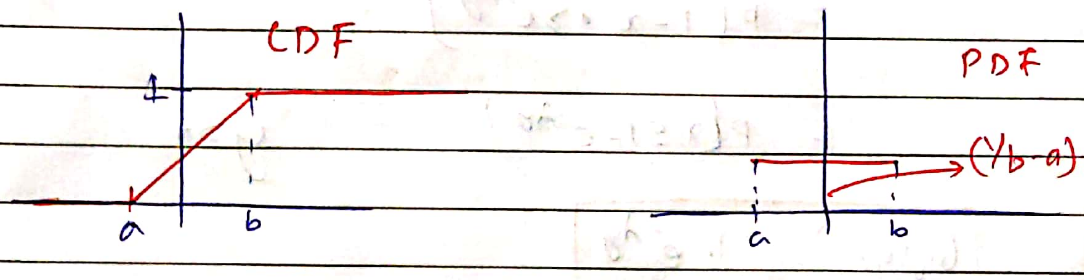
let x be a continuous random variable with the pdf defined as

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

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

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

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



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

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

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

Consider the transformations:

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

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

$$G_Y(y) = P(Y \le y)$$

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

$$= P[1-x \ge e^{-\lambda y}]$$

$$= P[x \le 1 - e^{-\lambda y}] \qquad y > 0$$

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

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

$$\boxed{g(y) = \begin{cases} \lambda e^{-\lambda y} & y > 0 \\ 0 & y \le 0 \end{cases}} \leftarrow \text{exponential density with parameter } \lambda (>0)$$

Scanned by CamScanner

classmate

Date

Page

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

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

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

For any real numbers $x, y > 0$

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

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

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

So,

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

or,

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

Memoryless property of exponential density.

Scanned by CamScanner

Theorem: Let φ be differentiable function which is strictly increasing or strictly decreasing on an interval I. Let $\varphi(I)$ 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 $Y = \varphi(X)$ whose density is given by

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

→ Trick to remember — looks like chain rule

Wherever you see x replace by $\varphi^{-1}(y)$.

Proof:

Ex: let X be a continuous r.v. 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?

→ Theorem is valid.

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

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

$$\text{or, } x = \left(\frac{y-a}{b}\right) \left(= \varphi^{-1}(y)\right)$$

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

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

Symmetric ~~[illegible]~~ 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 it's pdf $f_X(x)$ is a symmetric density.

Note

If X is a symm. r.v. with CDF $F_X(x)$. Then $F_X(0) =$

Scanned by CamScanner

classmate
Date
Page

Ex: $g(x) = \dfrac{1}{1+x^2}$ $\quad -\infty < x < \infty$

Is $g(x)$ a pdf ??

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

we modify the $g(x)$

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

Cauchy density $\rightarrow$ Symmetric

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

now, $c = \displaystyle\int_{-\infty}^{\infty} e^{-x^2/2}\,dx$ $\hookrightarrow$ can be integrated using gamma integ

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

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

now, $x = r\cos\theta,\ y = r\sin\theta$

$$\int_{0}^{\infty}\int_{0}^{2\pi} r e^{-r^2/2}\,d\theta\,dr$$

* Jacobian $\rightarrow$ $\dfrac{\partial x\,\partial y}{\partial r\,\partial \theta} = \begin{vmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{vmatrix}$

$\Rightarrow r(\cos^2\theta + \sin^2\theta) = r$

$\therefore c^2 = 2\pi$

$c = \sqrt{2\pi}$

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

Very Important

let define :

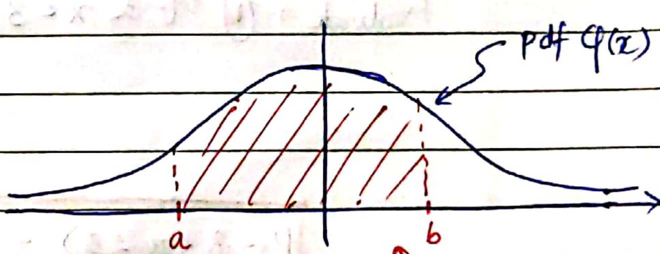
$$\boxed{\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \qquad -\infty < x < \infty}$$

Standard normal density

Symmetric

Random Variable corresponding to this density is called as "Standard normal random Variable".

Range R_X is full $\mathbb{R}$



now we would like to evaluate:

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

let Φ_X denote the CDF of std. normal r.v.

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

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

classmate
Date
Page

Scanned by CamScanner

— It seems that it is enough to know

$$\Phi_X(a) \quad \forall x \in \mathbb{R}$$

So, $$\Phi_X(a) = \int_{-\infty}^{a} \varphi(x)\,dx$$

or, $$\Phi_X(x) = \int_{-\infty}^{x} \varphi(x)\,dx$$

Since, $\varphi(x)$ is symmetric :

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

So, we can evaluate at $x=-a$ if we know at $x=a$.

Evaluate : (i) $P(-3 < x < 3) = 1 - 2\,\Phi_X(-3)$
$= 2\,\Phi_X(3) - 1$
$= 0.9974$

$$\boxed{P(-3 < x < 3) = 0.9973}$$

Very important (6-σ relation or something)

(ii) $P(-0.5 < x < -0.25) = \Phi_X(-0.25) - \Phi_X(-0.25)$
$= \Phi_X(0.5) - \Phi_X(0.25)$
$= 0.0928$ ☺

Note:

The probability is evaluated using Standard Normal table.

Scanned by CamScanner

classmate

Date

Page

Let X be a standard normal random variable
Define

$$Y = \mu + \sigma X \quad , \quad \sigma > 0$$

let $g(y)$ be the density of y:

$$g(y) = \frac{1}{\sigma}\varphi\left(\frac{y-\mu}{\sigma}\right)$$

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

$$\boxed{\therefore g(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(y-\mu)^2}{2\sigma^2}}}$$

Normal density with two parameters μ and σ^2.

$\mu \downarrow$ Mean, $\sigma^2 \downarrow$ Variance

$\downarrow$

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

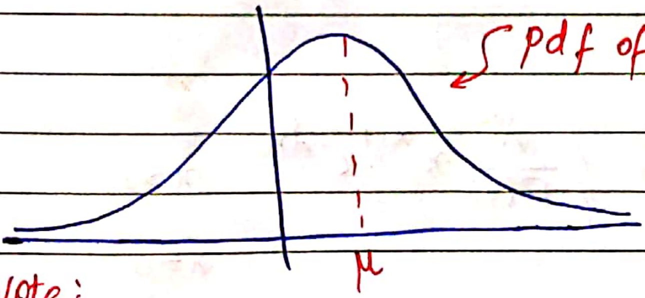
now let's calculate $P(Y)$:

$$P(a < Y < b) = P(a < \mu + \sigma X < b)$$

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

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

There is diff: Impossible Improbable



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

Note:

now, $P(\mu - 3\sigma < Y < \mu + 3\sigma) = P(-3 < X < 3) = 0.9973$

[6σ distance] $\longrightarrow$ Extremely important!!

classmate
Date
Page

★ Gamma densities

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

$Y = X^2$

Find the density of Y

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

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

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

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

↓

Gamma density

- Gamma functions:

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

- Gamma pdf:

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

parameter of this density are λ, α

if $\alpha = 1/2$, $\lambda = \frac{1}{2\sigma^2}$ we get $g(y)$

Scanned by CamScanner

- ## Expectation of a random variable :-

random variable(s)

discrete
- R_X
countably finite / infinite
- p.m.f.

continuous
- p.d.f.

Definition:- Let X be a discrete random variable with R_X as its range.

$$\boxed{E(X) = \sum_{x \in R_X} x \cdot P(X=x)}$$

→ finite sum. The sum should be convergent

Ex Let X be a discrete uniform r.v. ~~as~~ on $\{x_1, x_2, \ldots, x_n\} = R_X$

$$p(X=x_i) = \frac{1}{n} \quad \text{for } x_i \in R_X$$
$$= 0 \quad \text{otherwise}$$

$$E(X) = \sum_{x_i \in R_X} x_i \cdot P(X=x_i) = \sum_{x_i \in R_X} x_i \cdot P(x_i) = \frac{\sum_{x_i \in R_X} x_i}{n}$$

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

In general, $E(x)$ is weighted average.

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

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

← pmf of x in tabular form

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

$\boxed{\therefore\ E(X) = p}$

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

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

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

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

$$\left[\because\ j\binom{n}{j} = n\binom{n-1}{j-1}\right]$$

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

$$= np \sum_{j=1}^{n-1} p^{j-1} (1-p)^{n-j}$$

$$= np\,(p + 1 - p)^{n-1}$$

$$= \underline{np}$$

$\boxed{\therefore\ E(X) = np}$ → Counting the number of successes in n trials.

classmate
Date
Page

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

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

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

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

$$= e^{-\lambda} \sum_{i=0}^{\infty} \frac{i\lambda^{i}}{i!} \quad \underline{\lambda e^{\lambda}}$$

$$\boxed{\therefore E(X) = \lambda}$$

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

$$p(X=x) = (1-p)^{x} \cdot p \quad ; \; x = 0,1,2,\ldots$$

$$= 0 \quad ; \; \text{Otherwise}$$

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

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

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

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

Since the sum is convergant we interchange $\frac{d}{dp}$ and Σ.

$$= -p(1-p) \frac{d}{dp}\left(\frac{1}{p}\right) = -p(1-p)\frac{-1}{p^2}$$

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

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

Scanned by CamScanner

classmate
Date
Page

Ex: let X be a discrete r.v. with the pmf

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

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

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

$$= \left(1 - \frac{1}{2}\right) + \left(\frac{1}{2} - \frac{1}{3}\right) + \left(\frac{1}{3} - \frac{1}{4}\right) + \cdots$$

$$= \underline{\underline{1}}$$

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

Expectation doesnot exist

"we can't Expect anything"

- Properties of expectation :

1) let $Z = \varphi(x)$ be a function of the discrete random variable X.

$$E(Z) = E(\varphi(x)) = \sum_{x} \varphi(x)\, P(X = x)$$

provided that the expectation exists.

Ex:

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

$E(X) = 0, \quad E(X^2) = \underline{2}$

Scanned by CamScanner

classmate
Date
Page

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

1) For an $\alpha \in \mathbb{R}$ such that $P(X=\alpha)=1$
$E(X)=\alpha$

2) For some constant $c \in \mathbb{R}$, $E(cX)=cE(X)$

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

4) If $P(X \geq Y)=1$, then $E(X) \geq E(Y)$

moreover, $E(X)=E(Y)$ iff $P(X=Y)=1$

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

Ex: Hypergeometric

Population (r)

Type I: r_1

Type II: $r-r_1$

A sample of size n is drawn from the population.

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

Scanned by CamScanner

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

$$S_n = X_1 + X_2 + \dots + X_n$$

$$E(X_i) = \left(\frac{r_1}{r}\right)$$

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

Moments: $\longrightarrow$ To understand the shape/distribution of data.

$$E(X^r) = \sum_{x} x^r P(X=x) \qquad \text{for } r = 1, 2, \dots$$

These are called 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 $r=2$

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

↳ "How does x varies around the mean"

Scanned by CamScanner

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

Note:

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

$$\underbrace{E((X-\mu)^2)}_{\text{mean}} = \underbrace{\sigma^2}_{\text{Variance}}$$

- Interpretations of σ^2:

- Measure ~~o~~ variability of x around any real number

let's measure squared variability of x around a

$= E(X-a)^2$

↳ is always minimum ~~ano~~ ~~among~~ among various variabilities. if $a = \mu$.

let's ~~minini~~ minimise wrt a:

- Variance is the minimum error (Approximating a r.v. by a constant a and the error in the approximation is $E((X-a)^2)$).

$$E(X-a)^2 = E(X^2 - 2a\cancel{X} + a^2)$$
$$= E(X^2) - 2a E(X) + a^2$$

differentiate wrt a:

~~a = E(a)~~ $E(X) = a$

↳ mean

So the best choice is to approximate with mean.

Scanned by CamScanner

— Another interpretation of variance:

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

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

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

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

$\downarrow$

$E(X) - E(\mu)$

$\downarrow$

$E(X) - \mu$

$\downarrow$

$\mu - \mu = 0$

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

→ estimator biased trade offs.

now if we take $a = \mu$

$E((X-\mu)^2)$ will be minimum.

Kya kaata hai!!

Scanned by CamScanner

classmate
Date
Page

Theorem

Chebyshev's inequality :-

let X be a non-negative random variable with finite expectation.

$t > 0$ any positive real number

Define a new random variable Y from X

$Y = 0$ if $X < t$
$Y = t$ if $X \geq t$

$E(Y) = \cancel{0 \cdot P(X=} \; 0 \cdot P(Y=0) + P(X \geq t) \cdot t$

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

Note $X \geq Y$

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

$$\boxed{\therefore \quad P(X \geq t) \leq \frac{E(X)}{t}}$$

Formally :

let X be a random variable with mean μ & variance σ^2. For any real $t > 0$,

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

Scanned by CamScanner

Kind of Laplace Transform

Consider $(X-\mu)^2$

$$P((X-\mu)^2 \geq t^2) \leq \frac{E(X-\mu)^2}{t^2}$$

★ Moment generating functions:- (MGF)

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

If expectation exist MGF exist.

Properties

→ At $t=0$, $M_X(t) = 1$.

→ It should exist in a small interval around $t=0$. Important

→ In case of a discrete random variable X with pmf $p(X=x_i) = f(x_i) \ \forall x_i \in R_X$

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

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

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

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

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

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

Scanned by CamScanner

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

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

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

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

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

Note:

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

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

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

↳ Expectation at $t=0$

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

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

$$\frac{d}{dt} M_X(t) = n((1-p)+pe^t)^{n-1} \cancel{\frac{d}{dt}((1-p)+pe^t)} \, pe^t$$

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

↳ Expectation at $t=0$

For differentiation to exist, it should exist in a small interval around $t=0$.

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

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

or, $$M_x(t) = 1 + tE(x) + \frac{t^2}{2!}E(x^2) + \frac{t^3}{3!}E(x^3) + \cdots$$

$$\frac{d}{dt}M_x(t) = E(x) + tE(x^2) + \frac{t^2}{2!}E(x^3) + \cdots$$

So, $$\boxed{\left.\frac{d}{dt}M_x(t)\right|_{t=0} = E(X) \qquad \left.\frac{d^2}{dt^2}M_x(t)\right|_{t=0} = E(x^2)}$$

Variance σ^2:

$$\Rightarrow \left.\frac{d^2}{dt^2}M_x(t)\right|_{t=0} - \left(\left.\frac{d}{dt}M_x(t)\right|_{t=0}\right)$$

→ So, moment generating functions are very importantand convenient.

Ex: Let X be a random variable with MGF $M_x(t)$. Find the MGF of the r.v. $aX+b$ for some $a, b \in \mathbb{R}$, $a \neq 0$

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

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

$$= e^{bt}E\left(e^{atx}\right)$$

$$\boxed{M_{aX+b}(t) = e^{bt}M_x(at)}$$

$$\left.\frac{d}{dt}M_{aX+b}(t)\right|_{t=0} = be^{bt}M_x(at) + e^{bt} \cdot a\frac{d}{dt}M_x(at)$$

$$= b + aE(X)$$

Scanned by CamScanner

- Expectation of a continuous random variable:-

Let X be a continuous random variable with density (pdf) $f_X(x)$. Then expectation of x is defined as $E(X)$.

$$E(X) = \int_{-\infty}^{\infty} x f_X(x)\,dx \quad \text{if it exists.}$$

$$\left[E(x) \text{ exists if } \int_{-\infty}^{\infty} |x| \cdot f_X(x)\,dx < \infty \right]$$

Ex:

Scanned by CamScanner

Gamma Distribution: $\left(\frac{1}{\text{mean}}\right)$ is the parameter of a distribution.

classmate Date Page

Ex: $X \sim \text{Gamma}(1, \lambda)$; $\lambda > 0$

$$f_X(x) = \lambda e^{-\lambda x} \qquad x > 0$$

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

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

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

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

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

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

$$= \frac{1}{\pi} \int_0^k \frac{2x}{\pi(x^2+1)} dx$$

$$= \text{Limit doesn't exist}$$

- Moments of a continuous random variable:-

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

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

$m = 1, 2, \ldots$

Scanned by CamScanner

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

where, $\mu = E(X)$

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

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

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

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

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

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

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

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

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

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

$$\sigma^2 = \frac{1}{\lambda^2}$$

Ex: Easy excercise:

$$X \sim U[a,b]$$

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

$$= \frac{a^2+ab+b^2}{3}$$

Variance of $X = \sigma^2 = \frac{a^2+ab+b^2}{3} - \left(\frac{(a+b)}{2}\right)^2$

$$= \frac{4(a^2+ab+b^2) - 3(a^2+b^2+2ab)}{12}$$

$$= \frac{a^2+b^2-2ab}{12}$$

$$\boxed{\sigma^2 = \frac{(a-b)^2}{12}}$$

Note:

Advantage of symmetry of a random variable X.

Let X be a symmetric random variable

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

For any integer $m \in \mathbb{N}$,

$\Rightarrow$ X^m and $(-X)^m$ have same density

For m odd

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

Scanned by CamScanner

classmate
Date
Page

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

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

In particular, mean of all symmetric densities $= 0$.
(Ex:- $N(0, \sigma^2)$)

let's define moment generating function to calculate mean and σ^2 for Normal distribution

$M_X(t) = E(e^{tx})$ if it exists.

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

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

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

lets put $y = x - \mu$

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

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

★ ★ ★
Note: We use a trick of completing the square, often used in statistics

So, we put

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

Scanned by CamScanner

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

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

(1)

we completed the square to generate this integration

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

So, mean $= \left.\frac{d}{dt} M_X(t)\right|_{t=0} = (\mu + \sigma^2 t)\, e^{\mu t + \frac{\sigma^2 t^2}{2}}$

$= \mu$

$E(x^2) = \left.\frac{d^2}{dt^2} M_X(t)\right|_{t=0} \Rightarrow \mu^2 + \sigma^2(1 + \cancel{\ })$

$\Rightarrow \mu^2 + \sigma^2$

$\therefore$ Variance $\Rightarrow E(x^2) - E(x))^2$

$\Rightarrow \mu^2 + \sigma^2 - \mu^2$

$\Rightarrow \underline{\sigma^2}$

So, for normal distributions, the two parameters are μ and σ^2

μ ↓ Mean, σ^2 ↓ Variance

Classmate Date ____ Page ____

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

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

$$M_Y(t) = e^{t^2/2}$$

define $X = \mu + \sigma Y \qquad \sigma > 0$

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

$$= e^{\mu t} M_Y(\sigma t)$$

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

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

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

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

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

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

Define : $Y = \frac{X-\mu}{\sigma}$

$$= \underbrace{\frac{-\mu}{\sigma}}_{b} + \underbrace{\frac{X}{\sigma}}_{a}$$

$$M_Y(t) = M_{aX+b}(t) = e^{-\mu t/\sigma} \cdot M_X\left(\frac{t}{\sigma}\right)$$

$$= e^{-\mu t/\sigma} \cdot e^{\mu/\sigma\, t + t^2/2} = e^{t^2/2}$$

$\therefore Y \sim N(0,1)$

Scanned by CamScanner

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

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

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

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

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

Ex: Exponential density with parameter λ.

$$M_X(t) = \frac{\lambda}{\lambda - t} \qquad \exp(\lambda) = \text{gamma}(1, \lambda)$$

★ Quartiles :-

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

$$P(X \le q_1) = \frac{1}{4}$$

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

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

Scanned by CamScanner

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

$$P(X \le q_3) = \frac{3}{4}$$

Mode:

X ~~xxx~~ is a random variable with density $f_X(x)$

$$M = \underset{x}{\text{avg max}}\ f_X(x)$$

Ex: let $X \sim \text{Binomial}(3, \frac{1}{2})$

Compute median & mode.

⇒ $\text{mean} = 3 \times \frac{1}{2} = \frac{3}{2}$

$$\text{pmf } f_X(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

$$f_X(x) \begin{cases} = \binom{n}{x} \left(\frac{1}{2}\right)^x \left(1-\frac{1}{2}\right)^{3-x} & x = 0,1,2,3 \\ = 0 & \text{otherwise} \end{cases}$$

$$f_X(x) = \binom{n}{x} \frac{1}{8}$$

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

median ⇒ 1, mode = 1, 2 → highest probable event

upto 1

50% probability has occured

Scanned by CamScanner

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

Calculate

median = ?

mode = ?

mean $(X) = \mu$

Variance $(X) = \sigma^2$

$\Rightarrow$ median $= \mu$

mode $= \mu$

$\rightarrow \quad P(X \leq m) = 0.5$

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

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

From table

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

$\Rightarrow$ median $= \mu$

Note:

For normal distribution,

Mean = Median = mode $= \mu$

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

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

median $(X) = a + \left(\frac{b-a}{2}\right) \Rightarrow \left(\frac{a+b}{2}\right)$ = mean

mode $(X) = (a, b)$

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

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

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

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

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

$$\underbrace{m}_{\text{median}} = \frac{1}{\lambda} \ln 2$$

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 coordinate 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.

$$\Rightarrow f_X(x) = \begin{cases} -5 & x \in [-10, -5] \\ x & x \in [-5, 5] \\ 5 & x \in (5, 10] \end{cases}$$

CDF. $F_X(x) = \begin{cases} 0 & x < -10 \\ & \\ & \end{cases}$

Scanned by CamScanner