Random Variables

Suppose $(\Omega, \mathcal{B}, P)$ is a probability space. A random variable assigns numerical values to the points in the sample space.

Let us denote by X a r.v. on $(\Omega, \mathcal{B}, P)$. Then X is a function from Ω into $\mathbb{R}$.

$$X : \Omega \to \mathbb{R}$$

We put a condition of measurable on X.

<u>Examples</u>: 1. Suppose n matches of

chess are played between players A and B.

$X \to$ number of wins by A

$Y \to$ number of wins by B

$X \to 0, 1, 2, \ldots, n$

$Y \to 0, 1, 2, \ldots, n$

$X+Y+Z=n$

$Z \to$ number of draws

$Z \to 0, 1, 2, \ldots, n$

2. No of covid cases per day in WB / KGP during month of Sept 2021

$X \to 0, 1, 2 \dots\dots$

3. No. of trials needed for first direct hit on target by a player in archery competition

$X \to 1, 2, 3, \dots\dots$

4. Life of mobile charger (hrs)

$X \to (0, 100000)$

5. The amount of rainfall in an area during a week in monsoon (cm)

$X \to (0, 100)$

Types of Random Variables

Discrete R.V. : If a random variable takes finite or countably infinite number of values, it is called a discrete r.v.

Continuous R.V. : If a random variables takes values over an interval, then it is called a

Continuous r.v.

__Mixed R.V.__ : If a r.v. takes values over points as well as over intervals then it is called a mixed r.v.

Probability Distribution of a discrete __R.V.__

Let X be a discrete r.v. taking

values $x_1, x_2, \ldots \in \mathcal{X}$

The probability distribution of X is described by a probability <u>mass</u> <u>function</u> $p_X(k)$ satisfying

(i) $P(X = x_i) = p_X(x_i) \geq 0$ for all $x_i \in \mathcal{X}$

(ii) $$\sum_{x_i \in \mathcal{X}} p_X(x_i) = 1$$

<u>Example</u> : Suppose two fair dice are tossed and X is the sum observed.

$X \to 2, 3, 4, \cdots, 12$

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

$$p_X(3) = P(X=3) = P\left(\{(1,2),(2,1)\}\right) = \frac{2}{36}$$

$$p_X(4) = P(\{(1,3),(2,2),(3,1)\} = \frac{3}{36}$$

$$p_X(5) = \frac{4}{36}, \quad p_X(6) = \frac{5}{36}, \quad p_X(7) = \frac{6}{36}$$

$$p_X(8) = \frac{5}{36}, \quad p_X(9) = \frac{4}{36}, \quad p_X(10) = \frac{3}{36}$$

$p_X(11) = \frac{2}{36}$, $p_X(6) = \frac{1}{36}$

$Y =$ absolute difference of the numbers on upper faces

$Y \to 0, 1, 2, 3, 4, 5$

$$P(Y=0) = P(\{(1,1), (2,2), (3,3), (4,4), (5,5), (6,6)\}) = \frac{1}{6}$$

$$P(Y=1) = P\left(\{(1,2),(2,1),(2,3),(3,2),(3,4),(4,3),(4,5),(5,4),(5,6),(6,5)\}\right) = \frac{10}{36}$$

$$P(Y=2) = P\left(\{(1,3)(3,1),(2,4),(4,2),(3,5),(5,3),(4,6),(6,4)\}\right) = 8/36$$

$$P(Y=3) = P\left(\{(1,4),(4,1),(2,5),(5,1),(3,6),(6,3)\}\right) = 6/36$$

$$P(Y=4) = P\left(\{(1,5),(5,1),(2,6),(6,4)\}\right) = 4/36$$

$$P(Y=5) = P\left(\{(1,6),(6,1)\}\right) = 2/36$$

<u>Example</u>: A pack of 5 IC's

contains one defective. The chips are tested one by one until the defective is detected. Find the prob distⁿ of the number of testings reqd.

$X \rightarrow$ no of testings

$\rightarrow 1, 2, 3, 4$

$$p_X(1) = P(X=1) = \frac{1}{5}, \quad p_X(2) = \frac{4}{5} \cdot \frac{1}{4} = \frac{1}{5}$$

$$p_X(3) = \frac{4}{5} \cdot \frac{3}{4} \cdot \frac{1}{3} = \frac{1}{5}, \quad p_X(4) = \frac{2}{5}$$

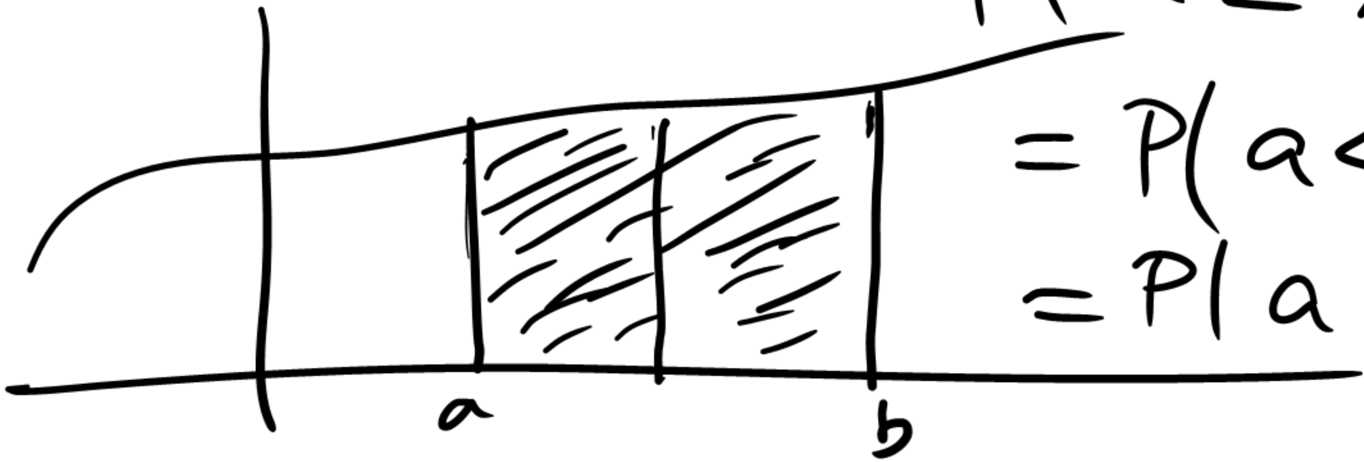
The <u>Probability Distⁿ</u> of <u>a</u> <u>Continuous</u> r.v : The prob. distⁿ of a continuous r.v. X is described by a prob. density function $f_X(x)$ satisfying

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

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

(iii) $$\int_a^b f_X(x)\,dx = P(a < X < b)$$

$$= P(a \leq X < b)$$



$$= P(a < X \leq b)$$

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

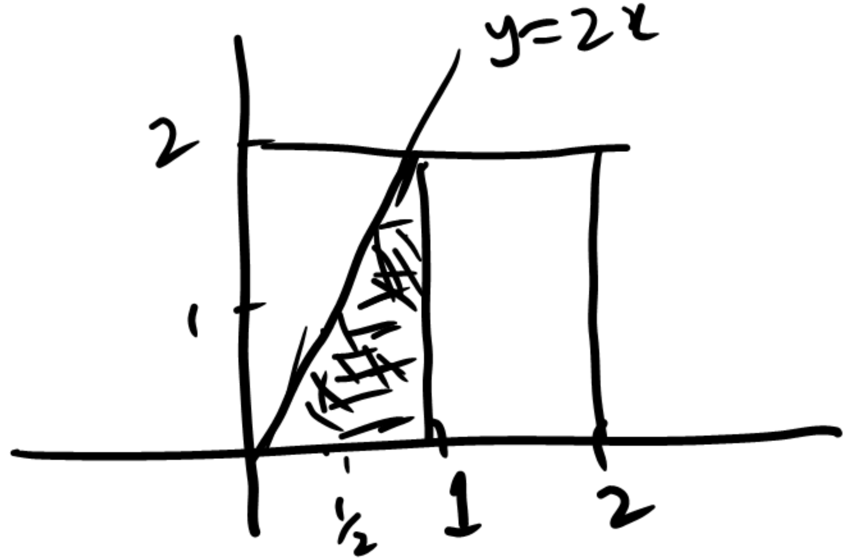
$$P(X=c) = 0 \quad \forall \quad c \in \mathbb{R}$$

for continuous r.v. X

<u>Example:</u> X is a continuous r.v.

with pdf $f_X(x) = \begin{cases} 2x, & 0 < x < 1 \\ 0, & \text{otherwise} \end{cases}$

$$\int_{-\infty}^{\infty} f_X(x)\, dx = \int_0^1 2x\, dx = x^2 \Big|_0^1 = 1$$



$$P\left(X<\frac{1}{2}\right) = \int_0^{1/2} 2x\,dx = x^2\Big|_0^{1/2} = \frac{1}{4}$$

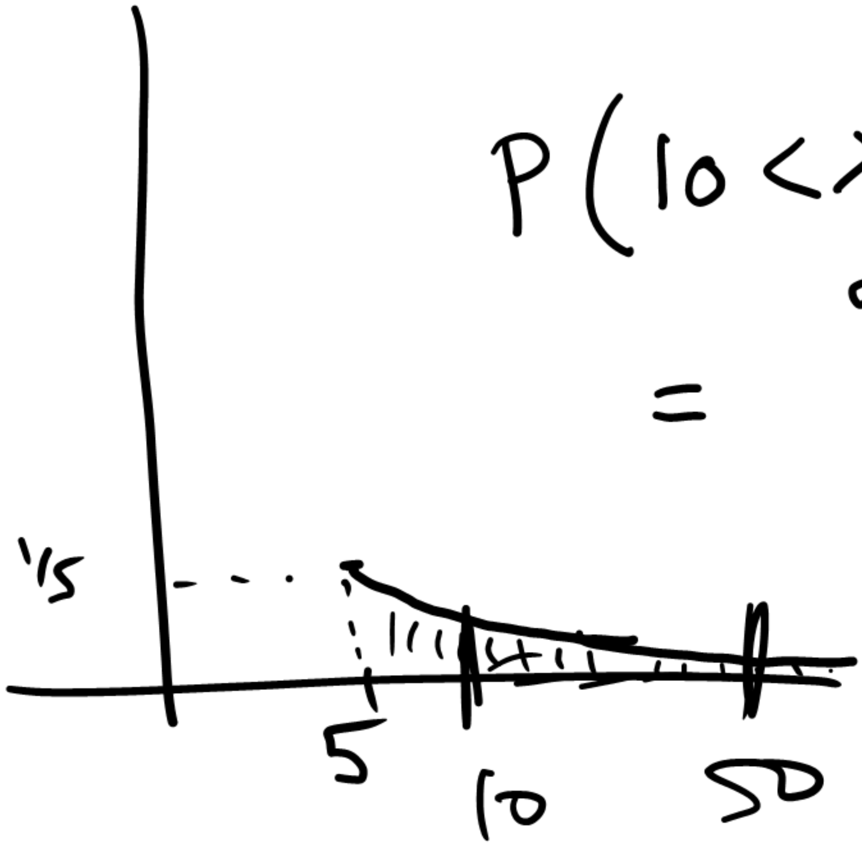
2. Let X be a continuous r.v. with pdf $f_X(x) = \begin{cases} \dfrac{5}{x^2}, & x > 5 \end{cases}$

$$\begin{cases} 0, & x \le 5 \end{cases}$$

$$\int_5^\infty \frac{5}{x^2}\,dx = -\frac{5}{x}\Big|_5^\infty = 1$$

$$P(10 < X < 50)$$

$$= \int_{10}^{50} \frac{5}{x^2}\,dx$$



$$= -\frac{5}{x}\Bigg|_{10}^{50} = \frac{1}{2} - \frac{1}{10} = \frac{2}{5}$$

<u>Cumulative Distribution Function</u>

<u>of a R.V. X</u> : (cdf)

$$F_X(x) = P(X \le x), \; x \in \mathbb{R}$$

$$= P(X \in (-\infty, x])$$

Dice example:

$X \to$ sum on two dice $\to 2, 3, \ldots, 12$

$$
F_X(x) = P(X \le x) = \begin{cases} 0, & x < 2 \\ \frac{1}{36}, & 2 \le x < 3 \\ \frac{3}{36}, & 3 \le x < 4 \\ 6/36, & 4 \le x < 5 \\ 10/36, & 5 \le x < 6 \end{cases}
$$

$$= 15/36, \quad 6 \le x < 7$$

$$= 21/36, \quad 7 \le x < 8$$

$$= 26/36, \quad 8 \le x < 9$$

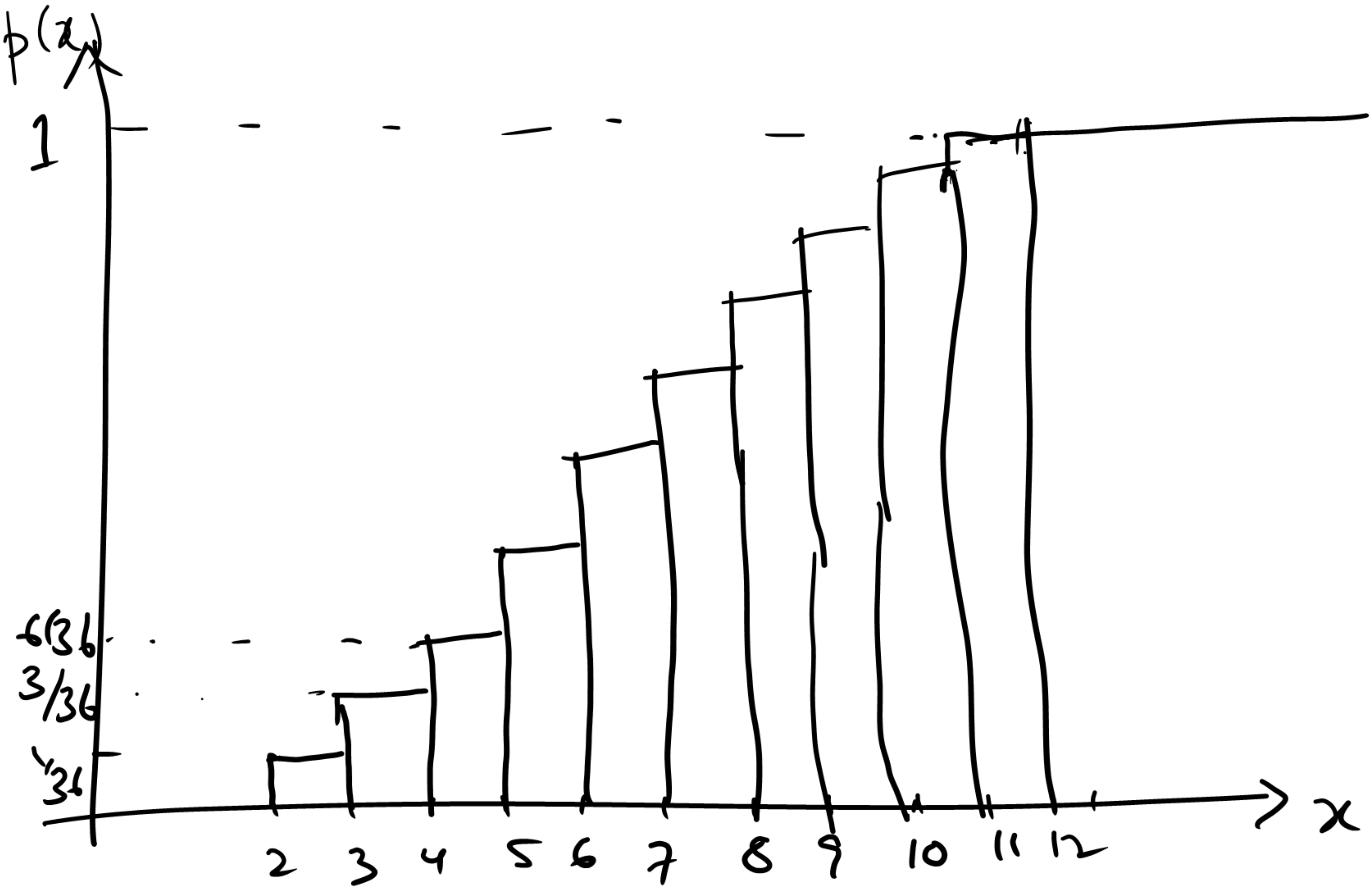
$$= 30/36, \quad 9 \le x < 10$$

$$= 33/36, \quad 10 \le x < 11$$

$$= 35/36, \quad 11 \le x < 12$$

$$= 1, \quad x \geq 12$$

The cdf of a discrete r.v. has a finite or countably infinite number of discontinuities. So we get it as a step function


p(x)
1
6/36
3/36
1/36
2
3
4
5
6
7
8
9
10
11
12
x

For a discrete r.v. X with values

$x_1, x_2, \ldots \in \mathcal{X}$

$$F_X(x) = \sum_{x_i \leq x} p_X(x_i)$$

$$F_X(x_i) - F_X(x_{i-1}) = p_X(x_i)$$

Ex. Write cdf of r.v. Y in this problem (*)

If X is a continuous r.v. with pdf $f_X(x)$, then

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

$$f_X(x) = \frac{d}{dx} F_X(x) \quad \text{for } x \in \mathbb{R}$$

$$\underline{\underline{(a.e.)}}$$

$$f_X(x) = \begin{cases} 2x, & 0 < x < 1 \\ 0, & \text{otherwise} \end{cases}$$

$$\begin{aligned} F_X(x) &= 0, & x \le 0 \\ &= x^2, & 0 < x < 1 \end{aligned}$$

$$= 1, \qquad x \geq 1$$

$$f_X(x) = \begin{cases} \dfrac{5}{x^2}, & x > 5 \\ 0, & x \leq 5 \end{cases}$$

$$\int_5^x \frac{5}{t^2}\,dt$$

$$F_X(x) = \begin{cases} 0, & x \leq 5 \\ 1 - \dfrac{5}{x}, & x > 5 \end{cases}$$

The cdf $F_X(x)$ satisfies the following properties:

(i) $\lim\limits_{x \to -\infty} F_X(x) = 0$

(ii) $\lim\limits_{x \to \infty} F_X(x) = 1$

(iii) F_X is a nondecreasing fn.

i.e. if $x_1 < x_2$, $F_X(x_1) \le F_X(x_2)$

(iv) F_X is continuous from right at every point, that is,

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

If $F(\cdot)$ is function satisfying the above 4 properties then there exists a r.v. X which has F as its cdf.

<u>Mixed R.V.</u>

$X \to$ waiting time at a traffic signals

$P(X=0) = \frac{1}{4}$, $f_X(x) = \frac{3}{4}$, $0<x<1$

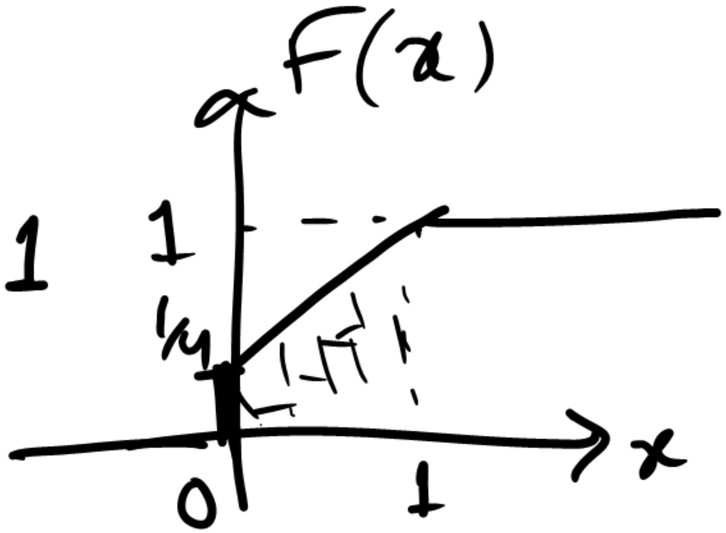
$$F_X(x) = 0, \qquad x < 0$$

$$= \frac{1}{4} \qquad x = 0$$

$$= \frac{1}{4} + \frac{3}{4}x, \qquad 0 < x < 1$$

$$= 1, \qquad x \geq 1$$

$$F_X(x) = \begin{cases} 0, & x < 0 \\ \dfrac{3x+1}{4}, & 0 \leq x < 1 \\ 1, & x \geq 1 \end{cases}$$



Expectation : Let X be a discrete r.v. with pmf $p_X(x_i)$, $x_i \in \mathcal{X}$.

We define the expected value of X as

$$E(X) = \sum_{x_i \in \mathcal{X}} x_i \, p_X(x_i) \rightarrow \text{mean of } X$$

average of X

provided the series on the right

converges absolutely.

Dice $\rightarrow$ $X \rightarrow$ sum

$$E(X) = 2 \cdot \frac{1}{36} + 3 \cdot \frac{2}{36} + \cdots + 11 \cdot \frac{2}{36} + 12 \cdot \frac{1}{36}$$

$$= \frac{252}{36} = 7$$

$X \to$ the no of tests reqd to detect defective IC

$$E(X) = 1 \cdot \frac{1}{5} + 2 \cdot \frac{1}{5} + 3 \cdot \frac{1}{5} + 4 \cdot \frac{2}{5} = \frac{14}{5}$$

$$= 2.8$$

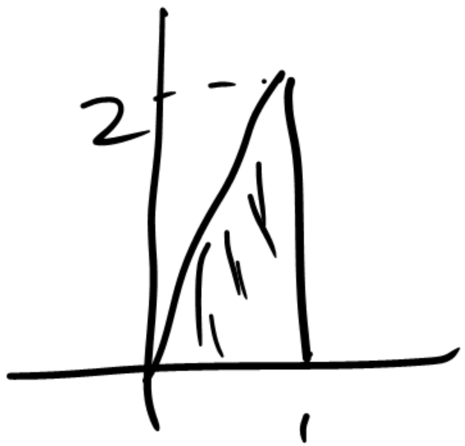
Let X be a continuous r.v. with

pdf $f_X(x)$. Then

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

provided the integral on the right converges absolutely.

$$f_X(x) = 2x, \quad 0 < x < 1$$



$$E(X) = \int_0^1 2x^2\,dx = \frac{2}{3}$$

$$f(x) = \frac{5}{x^2}, \quad x > 5, \quad E(X) = \int_5^\infty \frac{5}{x}\,dx$$

$$= 5 \log_e x \Big|_5^\infty \text{ diverges}$$

Here $E(X)$ does not exist

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

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

$$= 1$$

$$E(X) = \frac{1}{\pi} \int_{-\infty}^{\infty} \cdot \frac{x}{1+x^2}\,dx$$

diverges

E(X) does not exist

Moments:

Let $g : \mathbb{R} \to \mathbb{R}$

$Y = g(X)$ is a r.v.

$$E(Y) = E\,g(X)$$

$$= \sum_{x_i \in \mathcal{X}} g(x_i)\, p_X(x_i) \quad \text{discrete}$$

$$= \int_{-\infty}^{\infty} g(x) f_X(x)\, dx \qquad \text{continuous}$$

provided the series or integral on the right converges absolutely.

$$Y = aX + b, \qquad a, b \in \mathbb{R}$$

$$E(Y) = E(aX + b)$$

$$= \sum_{x_i \in \mathcal{X}} (a x_i + b)\, p_X(x_i)$$

$$= a \underbrace{\sum_{x_i \in \mathcal{X}} x_i\, p_X(x_i)}_{E(X)} + b \underbrace{\sum_{x_i \in \mathcal{X}} p_X(x_i)}_{1}$$

$$= a\, E(X) + b$$

ie Expectation is a linear function.

k^{th} noncentral moment of X

$$\mu'_k = E(X^k), \quad k = 1, 2, \ldots$$

$$\mu'_1 = E(X) \rightarrow \text{mean}$$

k^{th} central moment of X

$$\mu_k = E(X-\mu_1')^k, \quad k=1,2,\ldots$$

$$\mu_1 = 0 \quad \text{(always)}$$

$$\mu_2 = E(X-\mu_1')^2 = \text{Variance of } X$$

$\downarrow$

measure of dispersion / variation of r.v. X

Conventional notations $\mu_1' = \mu \to$ mean

$\sigma^2 = \mu_2 \to$ variance

$\sigma =$ standard deviation of r.v. X.

$$\mu_k = E(X-\mu_1')^k$$

$$= E\left[X^k - \binom{k}{1} X^{k-1}\mu_1' + \binom{k}{2} X^{k-2}\mu_1'^2 - \cdots + (-1)^{k+1}\mu_1'^k\right]$$

$$= \mu_k' - \binom{k}{1} \mu_{k-1}' \mu_1' + \cdots$$

$$+ (-1)^{k+1} \mu_1'^{\,k}$$

$$\mu_k' = E(X^k) = E\big((X - \mu_1') + \mu_1'\big)^k$$

$$= \mu_k + \binom{k}{1} \mu_k \mu_1' + \cdots + \mu_1'^{\,k}$$

$$\mu_2 = \mathrm{Var}(X) = \mu_2' - \mu_1'^{\,2}$$

$$\mathrm{Var}(X) = E(X^2) - \{E(X)\}^2 \geq 0$$

$$\Rightarrow \{E(X)\}^2 \leq E(X^2).$$