Random Variables

A random variable is a real valued function defined on the sample space. Let (Ω, B, P) be a probability space. A random variable

$$X : \Omega \to R$$
$$B \to C$$
$$P \to Q$$

?? $P(A) \to Q(E)$

$$(\Omega, B, P) \xrightarrow{X} (R, C, Q)$$

Eg :- Suppose a fair coin is tossed once.

$\Omega = \{H, T\}$

$B = \{\phi, \{H\}, \{T\}, \Omega\}$

$P(\phi) = 0 \qquad P(\{H\}) = \frac{1}{2} \qquad P(\{T\}) = \frac{1}{2}$

$X \to$ no. of heads

$X(H) = 1 \quad , \quad X(T) = 0$

$C = \{\phi, \{1\}, \{0\}, R\}$ $\quad R \to \{0, 1\}$

$Q(\{1\}) = \frac{1}{2} \qquad Q(\{0\}) = \frac{1}{2}$

$P(X = 1) = \frac{1}{2} \qquad P(X = 0) = \frac{1}{2}$

Types of Random Variables

A random variable is said to be discrete if it takes finite or countable number of values.

For eg, the number of wins in a series of games, number of children in a family, no. of deaths in a hospital.

If the random variable takes a value over an internal it is said to be cont. r.v. For example, age of problem, length of a string

To describe prob. distr. of a R.V we have following cases.

① Probability Mass Function
If a R.V. X is discrete and it takes values $x_1, x_2, \dots \in X$ it's probabity distribution is described by a fn. called probability mass fn. $P_X(x)$ if it satisfies

① $P(X = x_i) = P_X(x_i), \quad x_i \in X$

② $0 \leq P_X(x_i) \leq 1 \quad \forall \quad x_i \in X$

③ $\sum_{x_i \in X} P_X(x_i) = 1$

Eg :- Consider tossing of two fair dice. Let X = sum on the two dice. $X \to 2, 3, \dots, 12$

Then,

$P_X(2) = P(X=2) = \frac{1}{36}, \quad P_X(3) = \frac{2}{36}$

Scanned with CamScanner

Eg :- Suppose there are 5 ATM's in an office and two are working. A person randomly selects three machines. Let x be the no. of working machines in his selection. Find probability distribution of x.

$x \to 0, 1, 2$

$$P_x(0) = \frac{{}^3C_3}{{}^5C_3} = \frac{1}{10}$$

$$P_x(1) = \frac{{}^2C_1 \cdot {}^3C_2}{{}^5C_3} = \frac{6}{10}$$

$$P_x(2) = \frac{{}^3C_1 \; {}^2C_2}{{}^5C_3} = \frac{3}{10}$$

Continuous R.V : The probability distribution of a cont. R.V. x is described by a prob. density fn: $f_x(x)$ satisfying :

(i) $f_x(x) \geq 0 \quad \forall \; x \in R$

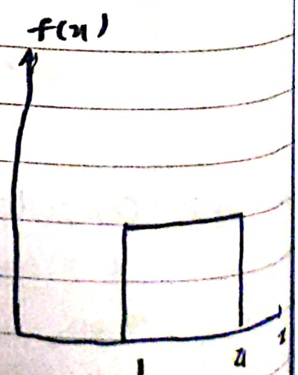
(ii) $\int_{-\infty}^{\infty} f_x(x)\,dx = 1$

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

Eg :- $f_x(x) = \begin{cases} 1/3 & 1 < x < 4 \\ 0 & \text{otherwise} \end{cases}$

$$\int_{-\infty}^{\infty} f_x(x)\,dx = \int_1^4 \frac{1}{3}\,dx = 1$$

$$P(2 < x < 3) = \int_2^3 \frac{1}{3}\,dx = \frac{1}{3}$$



Date
Page

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

$$\int_0^1 2x \, dx = 1$$

$$P\left(\frac{1}{4} < x < \frac{1}{3}\right) = \int_{1/4}^{1/3} 2x \, dx$$

$$= \frac{1}{9} - \frac{1}{16} = \frac{7}{16}$$

For cont. R.V the probability of a point is zero i.e

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

Cummulative Distribution function of a R.V:

For a R.V X, cdf $F_X(x)$ is defined as

$$F_X(x) = P(\{\omega : X(\omega) \leq x\})$$
$$= P(X \leq x)$$

Properities of CDF:

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

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

(iii) If $x_1 < x_2$, $F(x_1) \leq F(x_2)$

F - cont. from right every point. i.e

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

Conversely if a fn. F satisfies the above prop, then it is cdf of some RV, X

Scanned with CamScanner

If X is a cont. R.V, then the relationship between pdf and cdf is

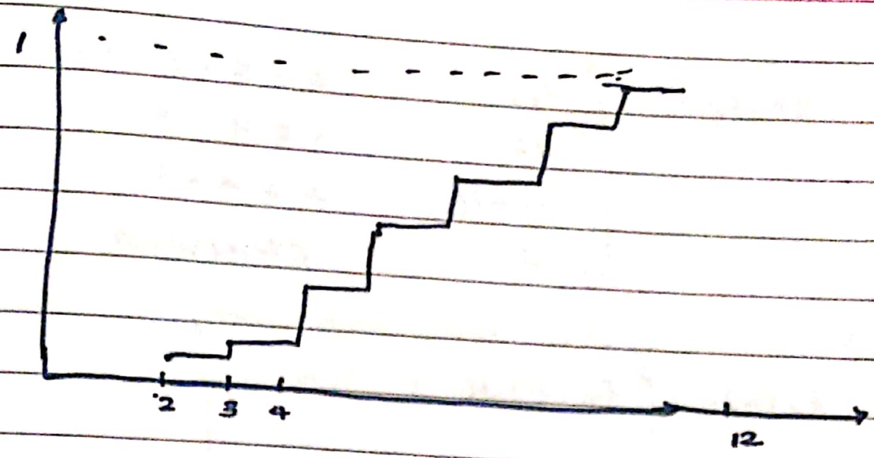
$$F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt \quad \text{and}$$

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

For dice problem $X \rightarrow$ sum

$$
\begin{aligned}
F_X(x) &= 0, & x < 2 \\
&= \frac{1}{36}, & 2 \le x < 3 \\
&= \frac{3}{36} & 3 \le x < 4 \\
&= \frac{6}{36} & 4 \le x < 5 \\
&= \frac{10}{36} & 5 \le x < 6 \\
&= \frac{15}{36} & 6 \le x < 7 \\
&= \frac{21}{36} & 7 \le x < 8 \\
&= \frac{26}{36} & 8 \le x < 9 \\
&= \frac{30}{36} & 9 \le x < 10 \\
&= \frac{33}{36} & 10 \le x < 11 \\
&= \frac{35}{36} & 11 \le x < 12 \\
&= 1 & x \geq 12
\end{aligned}
$$

Scanned with CamScanner



So cdf of a discrete R.v is a step - fn: and the size of discont. at finite or countably infinite no. of points

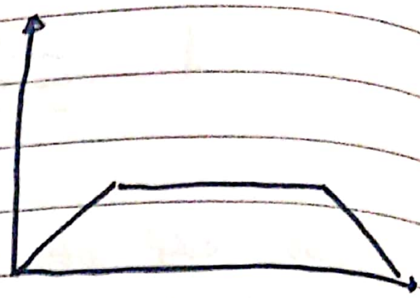
$$f_X(x) = \begin{cases} 2x & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\, dt = \begin{cases} 0 & x \le 0 \\ \int_0^x 2t\, dt & 0 < x \le 1 \\ \int_0^1 2t\, dt & x \geq 1 \end{cases}$$

$$= \begin{cases} 0 & x \le 0 \\ x^2 & 0 \le x < 1 \\ 1 & x \geq 1 \end{cases}$$

$$\frac{d}{dx} F_X(x) = \begin{cases} 2x & 0 < x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

Eq :- $f_X(x) = \begin{cases} x/2 & 0 < x < 1 \\ 1/2 & 1 \le x < 2 \\ (3-x)/2 & 2 \le x < 3 \\ 0 & \text{otherwise} \end{cases}$



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

$= 0 \qquad x \le 0$

$= \int_0^x \frac{t}{2}\,dt \qquad 0 \le x \le 1$

$= \int_0^1 \frac{t}{2}\,dt + \int_1^x \frac{1}{2}\,dt \qquad 1 \le x < 2$

$= \int_0^1 \frac{t}{2}\,dt + \int_1^2 \frac{1}{2}\,dt + \int_2^x \frac{3-t}{2}\,dt \qquad 2 \le x < 3$

$= \int_0^1 \frac{t}{2}\,dt + \int_1^2 \frac{1}{2}\,dt + \int_2^3 \frac{3-t}{2}\,dt \qquad x \ge 3$

so,

$F_X(x) = 0 \qquad x \le 0$

$= x^2/4 \qquad 0 < x \le 1$

$= \frac{1}{4} + \frac{x-1}{2}, \qquad 1 \le x < 2$

$= \frac{3}{4} + \left\{ -\frac{(3-t)^2}{4} \right\} \Big|_2^x \qquad 2 \le x < 3$

$= 1 \qquad x \ge 3$

Scanned with CamScanner

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

1) $\frac{d}{dx} F_X(x)$ This isn't cont. from right $x = 1/2$.
So not a cdf

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

$$\left.\begin{array}{l} f_X(x) = 1 \quad , \quad 0 < x < 1/2 \\ P(X = 1/2) = 1/2 \end{array}\right\} \text{Mixed R.V}$$

If a R.V, is partly discrete and partly cont., it is said to be a mixed R.V.

Let X be discrete with pmf $P_X(x_i)$,

$$E(X) = \sum_{x_i \in X} x_i \, P_X(x_i)$$

provided the series on the right is absolutely convergent.

Incase of cont. R.V

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

provided the integral on the right is absolutely convergent.

Scanned with CamScanner

Eg :- Sum on two dice

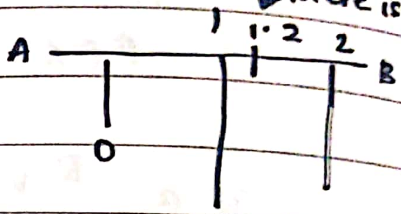
$$E(X) = 2\cdot\frac{1}{36} + 3\cdot\frac{2}{36} + 4\cdot\frac{3}{36} + 5\cdot\frac{4}{36} + 6\cdot\frac{5}{36}$$
$$+ 7\cdot\frac{6}{36} + 8\cdot\frac{5}{36} + 9\cdot\frac{4}{36} + \dots + 12\cdot\frac{1}{36}$$
$$= \frac{252}{36} = 7$$

↑ Avg sum

Working ATM eg.

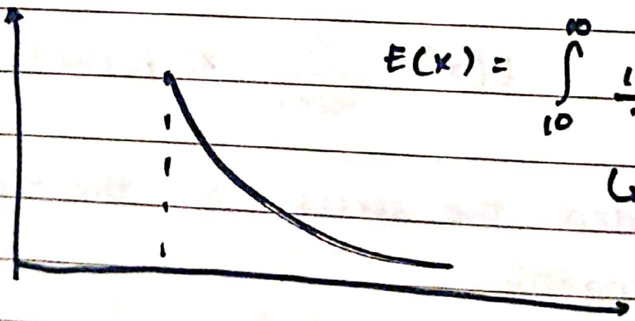
$$E(X) = 0\cdot\frac{1}{10} + 1\cdot\frac{6}{10} + 2\cdot\frac{3}{10}$$
$$= \frac{6}{5} = 1.2$$

At this point there is bal



$$f_X(x) = \begin{cases} 10/x^2 & x > 10 \\ 0 & \text{otherwise} \end{cases}$$

$E(X) \rightarrow$ doesn't exist



$$E(X) = \int_{10}^{\infty} \frac{10}{x}\, dx$$

↳ doesn't converge

Mixed Random Variable

$X \rightarrow$ waiting time at a traffic signal

$$P(X = 0) = \frac{1}{4}$$

$$f_X(x) = \begin{cases} 3/4 & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

$$F_X(x) = \begin{cases} 0 & x < 0 \\ 1/4 & x = 0 \\ 1/4 + \frac{3}{4}\int_0^x dt & 0 < x < 1 \\ 1 & x \geq 1 \end{cases}$$

$$= \begin{cases} 0 \\ \frac{1}{4} + \frac{3}{4}x & 0 \leq x < 1 \\ 1 & x \geq 1 \end{cases}$$

$$E(X) = 0 \cdot \frac{1}{4} + \int_0^1 \frac{3}{4} x \, dx = \frac{3}{8}$$

Symmetric Distribution

A R.V X is symmetric about a point c if

$$P(X \geq x + c) = P(X \leq c - x) \quad \forall x \in R$$

Scanned with CamScanner

If $c = 0$,

$$P(X \geq x) = P(X \leq -x)$$

→ Expectation is a linear operator

$$E(aX + b) = aE(X) + b)$$

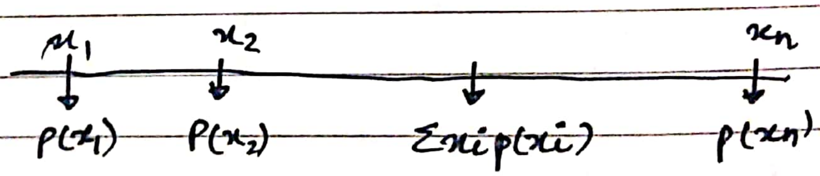
→ $$E(g(X)) = \sum_{x_i \in X} g(x_i)\, P_X(x_i) \quad \text{if } X \text{ is discrete}$$

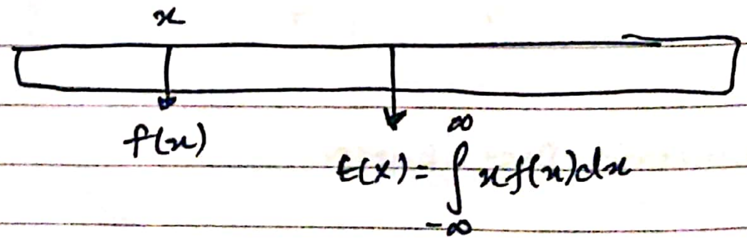
$$= \int_{-\infty}^{\infty} g(x)\, f_X(x)\, dx \quad \text{if } X \text{ is cont.}$$

provided right hand series/ integral is absolutely convergent

Moments: let $g(x) = x^k$

$\mu_k' = E(X^k)$ → k^{th} moment about origin (or) k^{th} known central moment.





We usually denote $\mu_1' = E(X) = \mu$ called mean of R.V X

$\mu_k = E(x-\mu)^k$

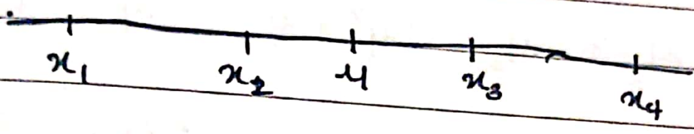
→ k^{th} about mean

→ k^{th} central moment

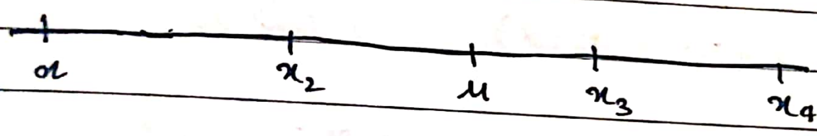
$\mu_1 = E(x-\mu) = E(x) - \mu = 0$

So the first central moment is always 0.

①



②



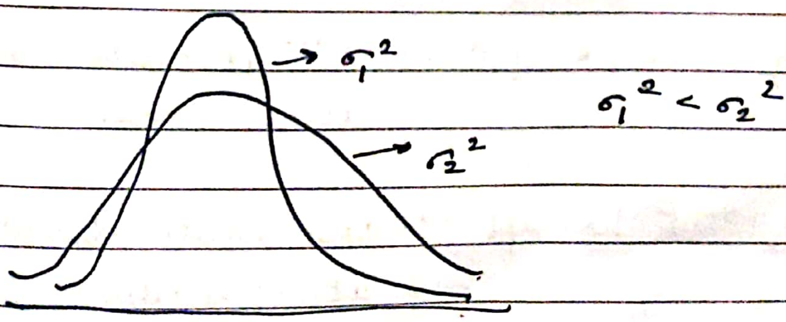
$E|x-\mu|$ → Abs. mean deviation from mean

$\mu_2 = E(x-\mu)^2$ → variance of x

$= Var(x)$

$= \sigma^2$

$\sigma = \sqrt{var(x)}$ = Std. Deviation of x



$\sigma_1^2 < \sigma_2^2$

Relationship between μ_k and μ_k'

$\mu_k = E(X-\mu)^k$ (k +ve integers)

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

$= \mu_k' - \binom{k}{1}\mu_{k-1}'\mu + \binom{k}{2}\mu_{k-2}'\mu^2 \dots + (-1)^{k+1}\mu^k$

$\mu_2 = \mu_2' - 2\mu^2 + \mu^2 = \mu_2' - \mu^2$

$var(x) = E(x^2) - \{E(x)\}^2 \geq 0$

$\therefore E(x)^2 \geq \{E(x)\}^2$

$\mu_k' = E(x^k) = E(\overline{x-\mu} + \mu)^k$

$= E\left[(x-\mu)^k + \binom{k}{1}(x-\mu)^{k-1}\mu + \dots \mu^k\right]$

$= \mu_k + \binom{k}{1}\mu_{k-1}\mu + \dots \mu^k$

$\beta_k' = E|x|^k \rightarrow k^{th}$ abs. moment about origin.

$\beta_k = E|x-\mu|^k \rightarrow k^{th}$ absolute moment about mean

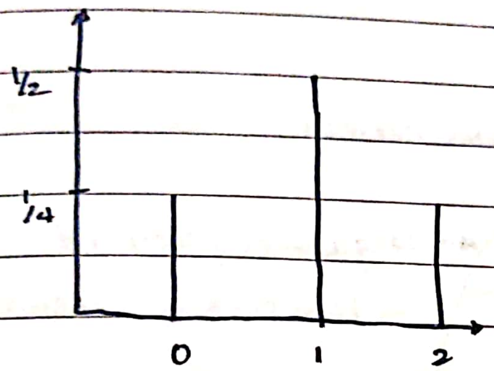
$\beta_1 = E|x-\mu| \rightarrow$ abs. mean deviation about mean.

Factorical Moments

$$\alpha_k = E[x(x-1)\cdots(x-k+1)],\ k=1,2\ldots$$

$x \to$ no. of heads

$P(x=0) = \frac{1}{4}$, $P(x=1) = \frac{1}{2}$, $P(x=2) = \frac{1}{4}$



$E(x) = 0\cdot\frac{1}{4} + 1\cdot\frac{1}{2} + 2\cdot\frac{1}{4}$

$= 1$

$E(x^2) = \frac{1}{2} + 1 = \frac{3}{2}$, $V(x) = \frac{3}{2} - 1 = \frac{1}{2}$

$\sigma = \frac{1}{\sqrt{2}}$

~~x → no. of defectives in his selection~~

Suppose a store has 10 AC's and 3 are defective.
A consumer buys 2 at random

$x \to$ no. of defectives in his selection

$x \to \{0,1,2\}$

$P_x(0) = \frac{{}^7C_2}{{}^{10}C_2} = \frac{7}{15}$ $P_x(1) = \frac{7}{15}$

$P_x(2) = \frac{1}{15}$

$E(x) = 0 \cdot \frac{7}{15} + 1 \cdot \frac{7}{15} + 2 \cdot \frac{1}{15}$

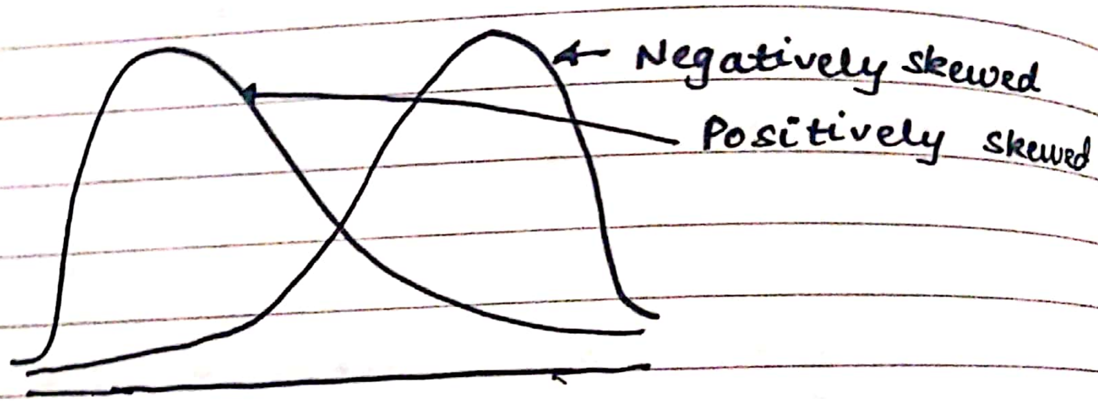
$= \frac{9}{15} = 0.6$

$E(x^2) = \frac{7}{15} + 1 \cdot \frac{7}{15} + 4 \cdot \frac{1}{15} = \frac{11}{15}$

$Var(x) = \frac{11}{15} - \frac{9}{25} \simeq 0.37$

$\sigma = 0.6$

Asymmetry or Skeweness



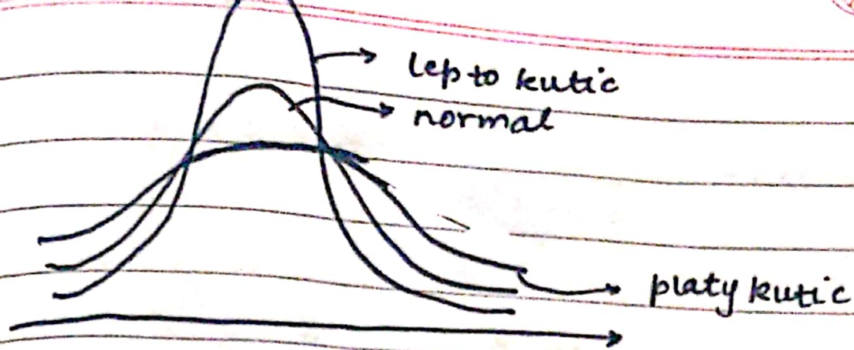
Measure of skewness

$\beta_1 = \frac{\mu_3}{\sigma^3} = \frac{\mu_3}{\mu_2^{3/2}}$

$\beta_1 = 0 \rightarrow$ Symmetric distribution

$> 0 \rightarrow$ for +vely skewed

$< 0 \rightarrow$ for -vely skewed



Measure of kutosis → peaked ness

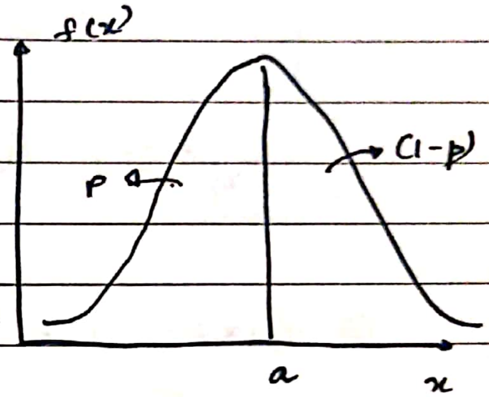
$$\beta_2 = \frac{\mu_4}{\mu_2^2} - 3 \quad \begin{cases} = 0 & \text{Normal} \\ > 0 & \text{Leptokutic} \\ < 0 & \text{Platykutic} \end{cases}$$

Quantiles

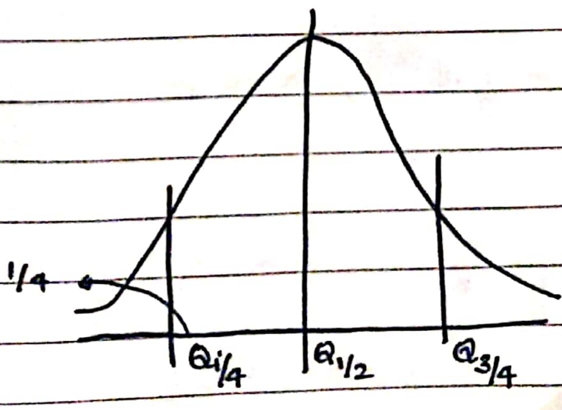
$P(x \leq a) = p$

A real no. Q_p satisfying

$P(x \leq Q_p) \geq p$ and



$P(x \geq Q_p) \geq 1-p \quad 0<p<1$ is called p^{th} quartil (or quartile of order p) of distribution of X. If F is a absolutely cont. then $F(Q_p) = p$ i.e there is a unique quartile



If $p = 1/2$, then its the median

Eg :- ① $f_x(x) = \dfrac{1}{\pi(1+x^2)}$ $\quad -\infty < x < \infty$

$F_x(x) = \dfrac{1}{2} + \dfrac{1}{\pi}\,\mathrm{Tan}^{-1}x$ $\quad x \in R$

CDF

$F_x(0) = 0$

So $H = 0$ is the median

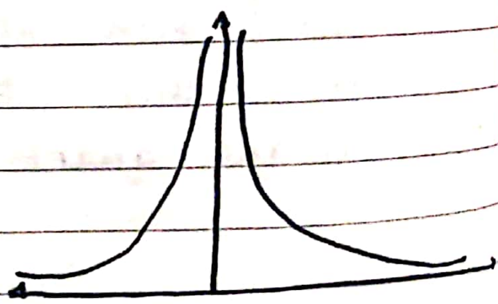
$E(x) = \displaystyle\int_{-\infty}^{\infty} \frac{1}{\pi} \frac{x}{1+x^2}\,dx$

doesn't exist

$Q_{1/4} = -1$, $Q_{3/4} = 1$ as $F(-1) = 1/4$, $F(1) = 3/4$

② $f_x(x) = \dfrac{1}{2} e^{-|x|}$, $\quad -\infty < x < \infty$

double exponential or laplace distn.



$F_x(x) = \displaystyle\int_{-\infty}^{x} \frac{1}{2} e^{t}\,dt$

$= 1 - \dfrac{1}{2}.e^{-x}$, $\quad x > 0$

$$F_x(x) = \begin{cases} \dfrac{e^x}{2} & x < 0 \\ 1 - \dfrac{1}{2} e^{-x} & x \geq 0 \end{cases}$$

Scanned with CamScanner

$F_X(0) = \frac{1}{2}$, so median is zero

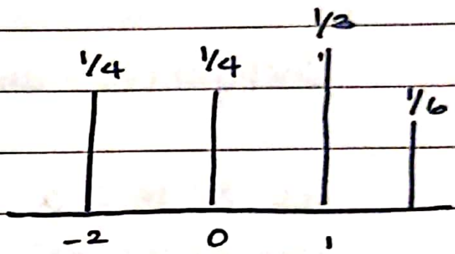
$\frac{e^{x}}{2} = \frac{1}{4} \Rightarrow x = -\log_e 2 \rightarrow Q_{1/4}$

$F_X(x) = \frac{3}{4} \Rightarrow 1 - \frac{1}{2}e^{-x} = \frac{3}{4}$

$\Rightarrow x = \ln 2 \rightarrow Q_{3/4}$

Eg :- $P(X=-2) = P(X=0) = 1/4$
$P(X=1) = 1/3 \quad P(X=2)\ 1/6$

$0 \leq M \leq 1$
↓
Median



first quartile
$-2 \leq Q_{1/4} \leq 0$

$M_X(t) = E(e^{tX})$, $t \in R$
is called mgf of X if it exists

Eg :- $f_X(x) = \begin{cases} 2x & 0 < x < 1 \\ 0 & \text{else} \end{cases}$

$M_X(t) = E(e^{tX}) = \int_0^1 f(x)\, e^{tx}\, dx$

$= \Big[$

Date
Page

$\rightarrow \frac{d^k M_x(t)}{dt^k}\Big|_{t=0} = \mu_k'$

Theorem : ① If the moment of order $t\,(>0)$ exists then the moment of orders $(0<s<t)$ exists for a given R.V x if the moment of order $s\,(>0)$ doesn't exist, then the moment of order $t\ (t>s)$ doesn't exist.

② The mgf uniquly determines a cdf

Chebyshev's Inequality

Let x be a R.V with mean μ and variance σ^2. Then for any $k>0$, then

$$P(|x-\mu| \geq k) \leq \frac{\sigma^2}{k^2}$$

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

$$= \int_{-\infty}^{\infty} (x-\mu)^2 f_x(x)\,dx$$

$$P(|x-\mu| < k) \geq 1-\frac{\sigma^2}{k^2}$$

Other forms

$$P(|x-\mu| < k\sigma) \geq 1-\frac{1}{k^2}$$

$$P(|x-\mu| \geq k\sigma) \leq \frac{1}{k^2}$$

Let us take $k = 1$

k

$$P(|x-\mu| < \sigma) \geq 0$$

Trivial Statement (Obvio)

$k = 2$

$$P(|x-\mu| < 2\sigma) \geq 3/4$$

Suppose $\mu = 18$, $\sigma = 2.5$. With what probability can we assert that there will be b/w 8 – 28 costly purchases

$$P(8 \leq x \leq 28) = P(-10 \leq x - 18 \leq 10)$$

$$= P(|x-18| \leq 10) \geq 1 - \frac{\sigma^2}{100}$$

$$\geq 15/16$$

Independent observations are available from a popⁿ with mean μ, variance 1. How many observations are needed in order that probability is atleast 0.9 that mean of observations differs from μ by not more than 1?

Suppose observations are $x_1, x_2 \cdots x_n$

$$\bar{x} = \frac{1}{n} \sum x_i = \mu$$

$$v(\bar{x}) = \frac{1}{n^2} \sum v(x_i) = \frac{1}{n}$$

$$1 - \frac{1}{n} > 0.9$$

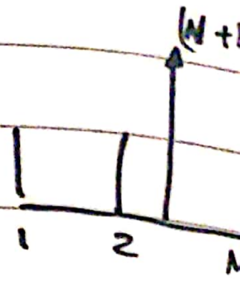
$$P(|\bar{x}-\mu| < 1) \geq 1 - \frac{1}{n}$$

$$n > 10$$

① **Discrete Uniform distn.**

$x \to 1, 2 \dots N \qquad P(x = i) = \frac{1}{N} \qquad i = 1, \dots N$

$$\mu_1' = \sum_{i=1}^{N} \frac{i}{N} = \frac{N+1}{2}$$

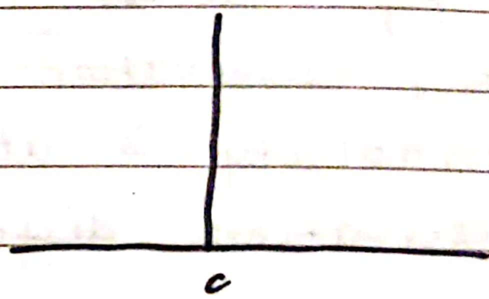


$$\mu_2' = E(x^2) = \sum \frac{i^2}{N} = \frac{(N+1)(2N+1)}{6}$$

$$V(x) = (N^2 - 1)/12$$

② **Degenrate Dist.**

$$P(x = c) = 1$$



$$\mu_k' = c^k$$

$$\mu_1' = c, \quad \mu_2' = c^2$$

MGF : $M_x(t) = E(e^{tx})$

$$= \sum_{i=1}^{N} \frac{e^{2t}}{N} = \begin{cases} \dfrac{e^t(e^{Nt} - 1)}{N(e^t - 1)} & t \neq 0 \\ 1 & t = 0 \end{cases}$$

(3) Bernoulli Trials

In a random expt., if we have two possible outcomes we associate them with success (s) and failure (f)

$\Omega = \{s, f\}$ $\quad x \to$ no. of success

$x(s) = 1, \quad x(f) = 0$

$P(x=1) = p \qquad p(x=0) = 1-p$
$= (q)$

$\mu_1' = p, \quad \mu_2' = p \; \dots \; \mu_k' = p$

$\mu_2 = \mu_2' - \mu_1'^2 = p - p^2$
$= pq$

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

(4) Binomial Dist^{n}:

Suppose independent Bernoulli trails have been conducted under identical cond., with probability of succes p.

Let $x \to$ no. of successes in n-trails

$P_x(k) = P(x=k) = \binom{n}{k} p^k q^{n-k}$

$k = 0, 1, \dots, n \quad , \quad 0 < p < 1$

$$\sum_{k=0}^{n} P_X(k) = (q+p)^n =$$

$$\mu_1' = E(X) = \sum_{k=0}^{n} k \binom{n}{k} p^k q^{n-k}$$

$$= \sum_{k=1}^{n} \frac{n!}{(k-1)!\,(n-k)!} p^k q^{n-k}$$

$$= np \leftarrow \text{Mean}$$

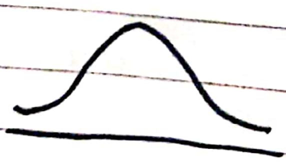
$$\mu_2' = n(n-1)p^2$$

$$V(X) = \mu_2' - \mu_1'^2$$
$$= np(1-p)$$
$$= npq$$

$$\mu_3 = np(1-p)(1-2p)$$
$$\mu_4 = 3npq^2 + npq(1-6pq)$$

Measure of skewness

$$\beta_1 = \frac{\mu_3}{\sigma^3} = \frac{1-2p}{(npq)^{1/2}} = 0, \quad p = 1/2$$



$\beta_1 = 0$
$p = 1/2$

Perfectly symmetric

$\beta_1 > 0$
$p < 1/2$

Postively Skewed

$p > 1/2$ Negatively skewed

$\beta_1 < 0$

$$\beta_2 = \frac{\mu_4}{\mu_2^2} - 3 = \frac{1 - 6pq}{npq} = 0 \quad , \quad pq = 1/6$$

$> 0 \qquad pq < 1/6$

$< 0 \qquad pq > 1/6$

Moment generating fn:

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

$$= \sum_{k=0}^{n} e^{tk} \binom{n}{k} p^k q^{n-k}$$

$$= \sum_{k=0}^{n} \binom{n}{k} (pe^t)^k q^{n-k}$$

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

An airline knows that 5% of people making reservations dont turn up for the flight. So it sells 52 tickets for a 50 seat flight. What is the probb. that every passenger who turns up, gets a seat?

A) $x \rightarrow$ no. of passengers who turn up for the flight.

$x \sim$ Bin $(52, 0.95)$

$$P(x \leq 50) = 1 - P(x=51) - P(x=52)$$

$$= 1 - \binom{52}{51} (0.95)^{51} (0.05) - \binom{52}{52} (0.95)^{52}$$

$$= 0.74$$

Geometric Distribution

Suppose Bernoulli trails are conducted independently under identical cond. till first success is achieved.
let x denote no. of trails needed for first success

$$P(x=k) = q^{k-1} p$$

$$\sum_{k=1}^{\infty} P_x(k) = \sum_{k=1}^{\infty} q^{k-1} p$$

$$= p(1+q+q^2+\dots)$$

$$= \frac{p}{1-q} = 1$$

$$\mu_1' = E(x) = \sum_{k=1}^{\infty} k\, q^{k-1} p$$

$$= p + 2pq + 3pq^2 + \dots$$

$$= p(1+2q+3q^2+\dots)$$

$$= \frac{1}{p}$$

So as to calculate $E(x)(x-1)$, $E(x)(x-1)(x-2)$ etc. we can use

$$\sum_{j=k}^{\infty} \binom{j}{k} r^{j-k} = \sum_{i=0}^{\infty} \binom{k+i}{k} r^i$$

$$= \frac{1}{(1-r)^{k+1}} \qquad 0<r<1$$

$\mu_2' = \frac{1+q}{p^2}$ $\quad$ $var(x) = \frac{q}{p^2}$

→ Always +vely skewed

Scanned with CamScanner

$$M_X(t) = E(e^{tx}) = \sum_{k=1}^{\infty} e^{tk} q^{k-1} p$$

$$= pe^t \sum_{k=1}^{\infty} (qe^t)^{k-1}$$

$$= \frac{pe^t}{1-qe^t}$$

If $qe^t < 1 \Rightarrow t < -\log q$

Eg :- suppose independent tests are conducted on mice while developing a vaccine. If they prob. of success in 0.2 in each trail. what is the probability that atleast 5 trails are needed to get first success

A) $X \to$ no. of trails needed to get first success

$X \sim$ Geo (0.2)

$$P(X \geq 5) = \cancel{1-P} \sum_{k=5}^{\infty} \left(\frac{4}{5}\right)^{k-1}\left(\frac{1}{5}\right)$$

$$= \left(\frac{4}{5}\right)^4 \left(\frac{1}{5}\right)\left[1 + \frac{4}{5} + \frac{4}{5}^2 + \cdots\right]$$

$$= \frac{\left(\frac{4}{5}\right)^4 \cancel{\left(\frac{1}{5}\right)}}{\cancel{\left(1-\frac{4}{5}\right)}} = \left(\frac{4}{5}\right)^4$$

$$= 0.4096$$

$\to$ $X \sim (\text{Geo}(p))$

$$P(X > m) = \sum_{k=m+1}^{\infty} q^{k-1} p = \frac{q^m p}{1-q} = q^m$$

$$P(X > m+n \mid X > n) = \frac{P(A \cap B)}{P(B)}$$

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

$$= \frac{q^{m+n}}{q^n} = q^m = P(x >$$

Memoryless property of geometric distn:

Negative Binomial Distn:-

Suppose Bernoulli trails are conducted independently under identical cond. until the r^{th} success is observed

$x \to$ no. of trails needed for r^{th} success

$$P_X(k) = \underbrace{\binom{k-1}{r-1} q^{k-r} p^{r-1}}_{\substack{r-1 \text{ success} \\ \text{in } (k-1) \text{ trails}}} \underbrace{(p)}_{\substack{k^{th} \text{ trail} \\ \text{is success}}}$$

$$= \binom{k-1}{r-1} q^{k-r} p^r$$

$\mu_1' = E(x) = \frac{r}{p}$, $\mu_2 = \text{var}(x) = \frac{rq}{p^2}$

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

$$= \sum_{k=r}^{\infty} e^{tk} \binom{k-1}{r-1} q^{k-r} p^r$$

$$= \left(\frac{pe^t}{1-qe^t}\right)^r, \quad 0 < t < -\log q$$

Q) suppose an airplane fails 2 of it's engines fail where the prob. of failure of each engine is p.

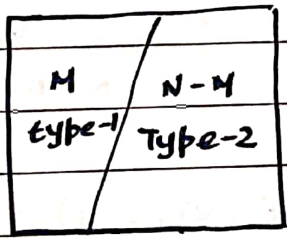
$X \rightarrow$ no. of engines failing

$$P_X(t) = \binom{k-1}{2-1} q^{k-2} p^2$$

$$= (k-1) q^{k-2} p^2, \quad k = 2, 3 \ldots$$

Hypergeometric Distribution

suppose a population has N-elements



N

$X \rightarrow$ no. of items of typle I in selected sample

$$P_X(k) = P(X=k) = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}} \quad k = 0, 1, \ldots N$$

$k \leq M$,

$n-k \leq N-M$

~~tk~~

$$(1+x)^N = (1+x)^M \; (1+x)^{N-M}$$

Co-eff of x^n on both sides

$$\binom{N}{n} = \sum_{k=0}^{n} \binom{M}{k} \binom{n-M}{n-k}$$

$$\mu_1' = E(X) = \frac{\sum_{k=1}^{n} k \binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}}$$

$$= \frac{nM}{N} \sum_{j=0}^{n-1} \binom{M-1}{j} \binom{N-1-M-1}{n-1-j} \Big/ \binom{N-1}{n-1}$$

$$= \frac{nM}{N}$$

$$E\,X(X-1) = \frac{n(n-1)\,M(M-1)}{N(n-1)}$$

$$E(X^2) = E(X)(X-1) + E(X)$$

$$= \left(\frac{N-n}{N-1}\right) \frac{nM}{N} \left(1 - \frac{M}{N}\right)$$

Theorem : let $X \sim$ hypergeo (M, N, n)
If $M \to \infty$, $N \to \infty \ni \frac{M}{N} \to p$, then

$$p_X(k) \to \binom{n}{k} p^k q^{n-k}, \quad k = 0, 1 \ldots n$$

suppose popn. size N is unknown

suppose M (type I) is unknown

$$\frac{X}{N} \to \frac{M}{N}$$

Scanned with CamScanner

$$M \propto \frac{Nx}{n}$$

$$N \propto \frac{Mn}{x}$$

Poisson process

We are observing events happening over time / area / space etc.
We say that the occurances / happenings observed in a time scale follow a poisson process they satisfy following assumptions

① No. of occurances in disjoint time int. are independent

② The probability of a single occurance during a small interval is prop. to length of interval

$X(h) \rightarrow$ no. of occurcerences in interval of length (h)

$P(X(h) = 1) = \lambda h$

③ Prob. of more than one occurance in a small time interval is negligable

$X(t) \rightarrow$ no. of occurances in an interval of length t

Under assumption (1) - (3),

$$P_n(t) = \frac{e^{-\lambda t}(\lambda t)^n}{n!}, \quad n = 0, 1, 2 \ldots$$