Characteristic of Distribution

Let X be a r.v. defined on a probability space $(\Omega, \mathcal{S}, P)$ associated with a random experiement $\mathcal{E}$.

$F_X(\cdot)$: Distribution function

$f_X(\cdot)$: pmf / pdf of X.

It may be desirable to have a set of numerical measures that provide a summary of the prominent features of the prob dist^n of X.

(1) Measure of Central Tendency or location

(a) Mean Gives us the idea about central value of the prob. dist^n around which the value of r.v. X clustered. commonly used measure of central tendency are:

ⓐ Mean :
$$\mu = \mu_1' = E(X) = \begin{cases} \int_{-\infty}^{\infty} x f_X(x)\,dx \\ \sum_{x \in S_X} x f_X(x) \end{cases}$$

Whenever it exists it gives us the idea about average observed value of X when $\mathcal{E}$ is repeated a large number of times.

Note that if dist^n of X is symmetric about μ then

$E(X) = \mu$.

<u>Median</u>: First we define Quantiles of a dist^n.

A number Q_p satisfying

$$P(X \leq Q_p) \geq p \quad \text{and} \quad P(X \geq Q_p) \geq 1-p.$$

$0<p<1$, is called p^{th} Quantile (or quantile of order p) of the dist^n of X.

If F is continuous cdf. then $F(Q_p) = p$ (i.e. $\exists$ a unique quantiles).

$$Q_{\frac{1}{2}} = \text{median of } X = Me.$$

$$Q_{\frac{1}{4}},\ Q_{\frac{1}{2}},\ Q_{3/4} \rightarrow \text{quartiles of } X$$

<u>Example</u>: $f_X(x) = \dfrac{1}{\pi}\dfrac{1}{1+x^2}, \quad -\infty < x < \infty$

$$F_X(x) = \frac{1}{\pi}\left(\tan^{-1} x + \pi/2\right), \quad -\infty < x < \infty$$

$F_X(M) = \frac{1}{2} \Rightarrow Me = 0. \longrightarrow$ Hence median is 0.

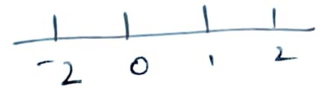
i.e. $Q_{1/2} = 0$.

Ex: $P(X=-2) = P(X=0) = 1/4, \; P(X=1) = 1/3, \; P(X=2) = \frac{1}{6}$

$$P(X \le Me) \ge 1/2 \quad \& \quad P(X \ge Me) \ge \frac{1}{2}$$

we have

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



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

So any Me s.t $0 \le Me \le 1$ satisfies the two conditions. Hence $Me \in [0,1]$ is a median.

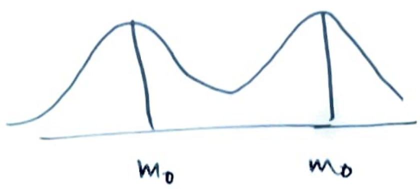
Here median is not unique.

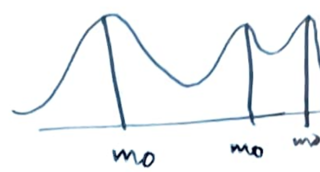
c) Mode: Roughly speaking mode m_0 of a prob. dist^n is the value that occurs with highest probability is defined by

$$f_X(m_0) = \sup\{ f_X(x) : x \in S \}.$$

If the random experiment $\mathcal{E}$ is repeated a large number of times then either mode m_0 or a value in the neighbourhood of m_0 is observed with maximum frequency.

Note that mode of a dist^n may not be unique. A dist^n having single / double / triple / multiple modes is called a unimodal / bimodal / trimodal / multimodal dist^n





Measure of Dispersion

Measure of dispersion give the idea about the scatter (cluster / dispersion) of probability mass of the dist^n about a measure of central tendency

(a) Mean Deviation :

$MD(A) = E(|X - A|)$ → provided it exists
called the mean deviation of X about A

$MD(\mu) = E(|X - \mu|)$ → mean deviation about $\mu = E(X)$

$MD(Me) = E(|X - Me|)$ → mean deviation about median.

(b) Standard Deviation :

$\sigma = \sqrt{Var(X)} = \sqrt{E(X-\mu)^2}$

standard deviation σ gives us the idea of average spread of values of X around the mean μ.

(c) Coefficient of variation :

Coefficient of variation is defined as

$$CV = \frac{\sigma}{\mu}, \quad \mu \neq 0.$$

where $\mu = E(X)$, $\sigma = \sqrt{Var(X)}$.

CV measures variation per unit of mean

CV does not depend of unit of measurements of r.v. X.

CV is very sensitive to small changes in μ when μ is near 0.

③ Measure of Skewness :

~~We define the mean~~ Skewness of a probability distribution is a measure of its assymetry (lack of symmetry).

We have already define the symmetric distn.

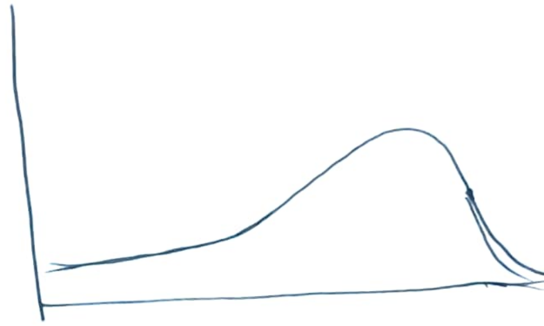
Positively skewed : Have more ~~porob~~ probability mass to the right side of pdf/pmf.

~~the~~ Have longer tails on the right side of pdf.

Negatively skewed distribution

- Have more prob mass to the left side of the pdf / pmf.

Have longer tails on the left side of pdf.



Let $X \sim F(\cdot)$, $E(X) = \mu$, $Var(X) = \sigma^2$

Define $Z = \frac{X - \mu}{\sigma}$: standardized variable (independent of units)

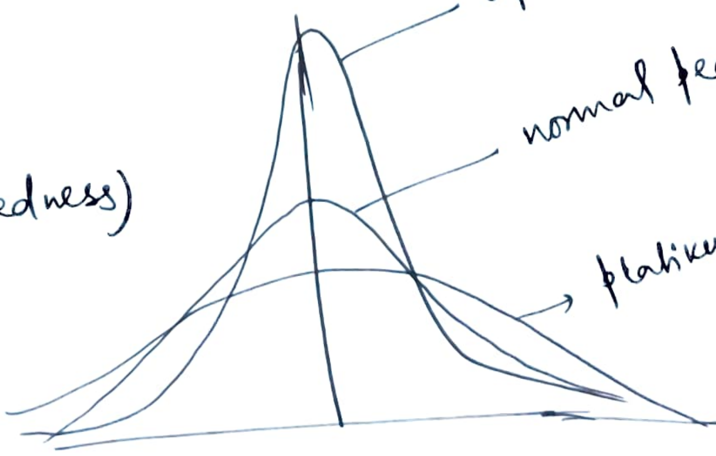
$$\beta_1 = E(Z^3) = \frac{E(X-\mu)^3}{\sigma^3} = \frac{\mu_3}{\sigma^3}$$

$\beta_1 = 0$ symmetic

> 0 positively skewed

< 0 (−)vely skewed.

4) Measure of Kurtosis :



Measure of Kurtosis (Peakedness)

$$\beta_2 = \left(\frac{\mu_4}{\mu_2^2} - 3\right)$$

$\beta = 0$ Normal peak

> 0 Leptokurtic

< 0 platikurtic.