Assignment - 6

Ques! What is a Cumulative Distribution function, and how does it work?

Ans CDF of a real-valued Random variable X is the probability that X will take a value less than or equal to x.

CDF used to specify the distribution of Multivariate random variables.

$$F_X(x) = P(X \leq x) \rightarrow \text{For discrete Random variable}$$

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt \rightarrow \text{For continuous Random variable}$$

Property

① $\lim_{X \to -\infty} F_X(X) = 0$ and $\lim_{X \to \infty} F_X(X) = 1$

② For all real no. a & b with continuous Random variable X, then the function F_X is equal to derivative of F_X

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

→ (Plot of addition of all previous points)

Ques 2 When should we use a t-test v/s a z-test?

Ans We use T-test when we donot have population standard deviation or when sample size is less than 30.

Ques 3 How do we examine two categorical characteristics?

Ans ~~For 2 group we use Anova Test (F test) or T test Anova can also be used for more~~

Chi-Square test is performed to examine relationship b/w 2 or more categorical variables

Formula ⇒ $X^2 = \sum \frac{(f_o - f_e)^2}{f_e}$

Ques 4 Explain the concept of chebyshev's Inequality.

Ans It is used to find Emperical formulas for non Gaussian distribution

Formula ⇒ $P\left[(\mu - k\sigma) \leq Y \leq (\mu + k\sigma)\right] \geq \left(1 - \frac{1}{k^2}\right)$

where, $k > 1$ always

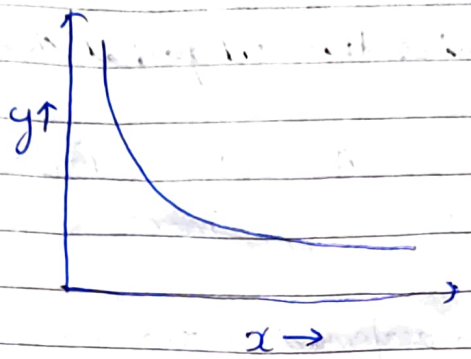
$Y \neq$ Gaussian Distribution

Eg if $k = 2$

$P\left[(\mu - 2\sigma) \leq Y \leq (\mu + 2\sigma)\right] \geq \frac{3}{4} \approx 75\%$

Ques Explain the concept of Pareto Distribution?

Ans Any data which follows 80:20 rule and distribution looks like below is called Pareto Distribution



80:20 Rule ⇒ 80% of Y satisfies 20% of X

Eg: of Pareto Distribution is "Wealth Inequality".
80% of wealth is hold by 20% of People.

We can convert Pareto Distribution to Normal Distribution using Box - Cox transformation.