Prob-Stats

Post-MidSem
Anshul

Joint Distribution Function :-

{Joint cumulative probability distribution function}

$$F(a,b) = P\{X \le a, Y \le b\}$$
$$\forall\ -\infty < a, b < \infty$$

$$F_X(a) = P(X \le a) \equiv F(a, \infty)$$
$$F_Y(b) = P(Y \le b) \equiv F(\infty, b)$$

Marginal Distributions of X & Y.

* $P\{X > a, Y > b\} = 1 - F_X(a) - F_Y(b) + F(a,b)$
* $P\{a_1 < x \le a_2,\ b_1 < y \le b_2\} =$
$$F(a_2,b_2) + F(a_1,b_1) - F(a_1,b_2) - F(a_2,b_1)$$

Joint Probability Mass function :-

{for discrete Random Variables}

$$p(x,y) = P(X = x, Y = y)$$

$$P_X(x) = P(X = x) = \sum_{y:\, p(x,y) > 0} p(x,y)$$
$$P_Y(y) = P(Y = y) = \sum_{x:\, p(x,y) > 0} p(x,y)$$

→ Marginal probability mass function

Note :- X and Y are said to be jointly Continuous if there exists $f(x,y)$, for all real x & y, having property that for every set C in 2D plane,

$$P\{(X,Y) \in C\} = \iint_{(x,y) \in C} f(x,y)\, dx\, dy$$

~~Jointly Continuous~~

Jointly probability density function :-

$$C = \{(x,y) : x \in A, y \in B\}$$

$$P\{X \in A, Y \in B\} = \int_B \int_A f(x,y)\, dx\, dy$$

Note :-

* $F(a,b) = P\{X \in (-\infty, a], Y \in (-\infty, b]\}$
$$= \int_{-\infty}^{a} \int_{-\infty}^{b} f(x,y)\, dy\, dx$$

* $f(a,b) = \dfrac{\partial^2}{\partial a\, \partial b} F(a,b)$

* X & Y jointly Continuous ⟹ they are individually Continuous

* $f_X(x) = \int_{-\infty}^{\infty} f(x,y)\, dy$
$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\, dx$

→ Marginal density functions

Generalisation :- {n Random var.}

$$F(a_1, a_2, \cdots a_n) = P\{X_1 \le a_1, X_2 \le a_2, \cdots X_n \le a_n\}$$

$$P\{(X_1, X_2, \cdots X_n) \in C\} = \iint \cdots \int_{(x_1, x_2, \cdots, x_n) \in C} f(x_1, \cdots, x_n)\, dx_1\, dx_2 \cdots dx_n$$

$$P\{X_1 \in A_1, X_2 \in A_2, \ldots, X_n \in A_n\}$$

$$= \int_{A_n}\int_{A_{n-1}} \cdots \int_{A_1} f(x_1, \ldots x_n)\, dx_1 dx_2 \cdots dx_n$$

The multinomial distribution :-

Sequence of n independent and identicle experiments is performed.

$r \to$ No. of possible outcomes.

$X_i \to$ No. of experiments result in outcome number i. $(1 \le i \le r)$

$P_1, P_2, \ldots, P_r \to$ respective probabilities

$$\sum_{i=1}^{r} P_i = 1$$

$$P\{X_1 = n_1, X_2 = n_2, \ldots, X_r = n_r\}$$

$$= \frac{n!}{n_1! n_2! \cdots n_r!} P_1^{n_1} P_2^{n_2} \cdots P_r^{n_r}$$

$$\forall \sum_{i=1}^{r} n_i = n$$

Independent Random Variable :-

$$P\{X \in A, Y \in B\} = P\{X \in A\} P\{Y \in B\}$$

$$P\{X \le a, Y \le b\} = P\{X \le a\} P\{Y \le b\}$$

$$F(a,b) = F_X(a) F_Y(b)$$

$$P(x,y) = P_X(x) P_Y(y)$$

$$f(x,y) = f_X(x) f_Y(y)$$

$\to \rho \to$ Correlation Coefficient

$$= \frac{Cov(X,Y)}{\sigma_x \sigma_y}$$

Sums of Independent Random Variables :-

$X+Y \to$ New Random variable

$$F_{X+Y}(a) = P\{X+Y \le a\}$$

$\downarrow$

Convolution

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{a-y} f_X(x) f_Y(y)\, dx\, dy$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{a-y} f_X(x)\, dx\, f_Y(y)\, dy$$

$$= \int_{-\infty}^{\infty} F_X(a-y) f_Y(y)\, dy$$

$$f_{X+Y}(a) = \frac{d}{da} F_{X+Y}(a)$$

$$= \int_{-\infty}^{\infty} f_X(a-y) f_Y(y)\, dy$$

$$F_n(x) = P\{X_1 + \cdots + X_n \le x\}$$

$\to$ Uniform Random variables $(0,1)$

for $x \le 1$,

$$F_n(x) = \frac{x^n}{n!}, \quad 0 \le x \le 1$$

$\to$ Can be proved by induction

(A) Sum of two independent uniform random variables :-

$$f_X(a) = f_Y(b) = \begin{cases} 1 & ; 0 < a < 1 \\ 0 & ; \text{o.w.} \end{cases}$$

$$f_{X+Y}(a) = \begin{cases} a & , 0 \le a \le 1 \\ 2-a & , 1 < a < 2 \\ 0 & , \text{ow} \end{cases}$$

$\to$ Triangular Distribution

(B) Gamma Random Variables :-

$$f(x) = \frac{\lambda^{\alpha} x^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)} \quad ; 0<x<\infty$$

$X \sim \text{gamma}(s, \lambda)$

$Y \sim \text{gamma}(t, \lambda)$

$X+Y \sim \text{gamma}(s+t, \lambda)$

$$f_{X+Y}(a) = \frac{\lambda^{s+t} a^{s+t-1} e^{-\lambda a}}{\Gamma(s+t)}$$

$X_1 + X_2 \cdots + X_n \sim \text{gamma}(\sum^n t, \lambda)$

(C) Normal Random Variables :-

$$N\left(\sum_1^n \mu_i, \sum_1^n \sigma_i^2\right)$$

$$Z = \frac{\sum X_i - \sum \mu_i}{\sqrt{\sum \sigma_i^2}}$$

(D) Poisson Random Variables :-

$$\text{Poisson}\left(\sum_1^n \lambda_i\right)$$

$$f_{X+Y}(a) = \frac{e^{-(\lambda_1+\lambda_2)} (\lambda_1+\lambda_2)^a}{a!}$$

(E) Binomial Random Variables :-

$$\text{Binomial}(n+m, p)$$

$$P\{X+Y=k\} = p^k q^{n+m-k} \sum_{i=0}^{n} \binom{n}{i}\binom{m}{k-i}$$

(F) Geometric Random Variables :-

$$\overset{S_n}{\overset{\shortparallel}{P\{X_1+X_2+\cdots+X_n=k\}}} = \binom{k-1}{n-1} p^n (1-p)^{k-n}, \quad k \geq n$$

p is same for all

(This is basically converted into Negative Binomial) $\sim NB(n,p)$

if p is different for all,

$$P(S_n = k) = \sum_{i=1}^{n} p_i q_i^{k-1} \prod_{j \neq i} \frac{p_j}{p_j - p_i}$$

(G) $\text{Exp}(\lambda) \longrightarrow \text{Gamma}(n, \lambda)$ No. of r.v.

Conditional Distribution :-

(A) Discrete Case :-

$$p_{X|Y}(x|y) = \frac{p(x,y)}{p_Y(y)}$$

$$F_{X|Y}(x|y) = P\{X \leq x \mid Y=y\}$$
$$= \sum_{a \leq x} p_{X|Y}(a|y)$$

(B) Continuous Case :-

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)}$$

$$P(X \in A \mid Y=y) = \int_A f_{X|Y}(x|y)\,dx$$

$$F_{X|Y}(a|y) = \int_{-\infty}^{a} f_{X|Y}(x|y)\,dx$$

The Bivariate Normal Distribution

if $\mu_x, \mu_y, \sigma_x > 0, \sigma_y > 0, -1 < \rho < 1$, then

$$f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} e^{\left\{\frac{-1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 + \left(\frac{y-\mu_y}{\sigma_y}\right)^2 - 2\rho\frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}\right]\right\}}$$

* Conditional

if $\boxed{Y=y}$ then X is normally distributed

with $\mu = \mu_x + \rho\frac{\sigma_x}{\sigma_y}(y-\mu_y)$

$$\sigma^2 = \sigma_x^2(1-\rho^2)$$

* $f_X(x) \sim N(\mu_x, \sigma_x^2)$

Properties of Expectation :-

$g(X,Y) \rightarrow$ function of two random variables X & Y.

$$E[g(X,Y)] = \sum_y \sum_x g(x,y)\, p(x,y)$$

(for Discrete Rand. var.)

$$E[g(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)\, f(x,y)\, dx\, dy$$

or

$$\{\text{if } g(x,y) \geq 0\} \rightarrow \int_y \int_x g(x,y)\, f(x,y)\, dx\, dy$$

(for Continuous Rand. var.)

$$\rightarrow E(X+Y) = E(X) + E(Y)$$

The Sample Mean :-

$X_1, \ldots, X_n \rightarrow$ independent and identically distributed r.v.'s having distribution fn F and expected value μ.

$$\boxed{\bar{X} = \sum_{i=1}^{n} \frac{X_i}{n}}$$

$$\Rightarrow E(X) = E(\bar{X}) = \mu = E[X_i]$$

(A) Binomial random variables,

$$E(X) = E(X_1) + E(X_2) + \cdots + E(X_n) = np$$

$$\forall\ X = X_1 + X_2 + \cdots + X_n$$

(B) Negative Binomial R.V.,

$$X = X_1 + X_2 + \cdots + X_r$$

$$E(X) = \frac{r}{p}$$

(C) Hypergeometric R.V.,

$$X = X_1 + X_2 + \cdots + X_m$$

$X_m \rightarrow$ No. of white balls selected

$n \rightarrow$ balls selected

$N \rightarrow$ Total balls

$$E(X) = \frac{mn}{N}$$

~~Moments of the Number of Events that Occur~~ :-

Method of Transformation :-

$$\boxed{f_{Y_1,Y_2}(y_1,y_2) = f_{X_1,X_2}(x_1,x_2)\ |J|}$$

$$\forall\ J = \frac{\partial(x_1,x_2)}{\partial(y_1,y_2)} = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \\ \frac{\partial x_1}{\partial y_2} & \frac{\partial x_2}{\partial y_1} \end{vmatrix}$$

Product Moments :-

$$\boxed{\mu'_{r,s} = E(X^r Y^s)}$$

Central product Moments :-

$$\boxed{\mu_{r,s} = E[(X-\mu_x)^r (Y-\mu_y)^s]}$$

Covariance :-

$$Cov(X,Y) = E((X-\mu_x)(Y-\mu_y))$$

$$\boxed{Cov(X,Y) = E(XY) - E(X)E(Y)}$$

Note:- if X and Y are independent then,

$$Cov(X,Y) = 0$$

since, $E(XY) = E(X)E(Y)$

Correlation :-

$\rho \rightarrow$ Correlation Coefficient

$$\boxed{\rho_{X,Y} = \frac{Cov(X,Y)}{\sigma_x \sigma_y}}$$

$$-1 \le \rho_{X,Y} \le 1$$

Note:- For transformation,

$$U = \frac{X-a}{h} \qquad V = \frac{Y-b}{K}$$

$$\boxed{Cov(X,Y) = hK\, Cov(U,V)}$$

$$\boxed{\rho_{X,Y} = \rho_{U,V}}$$

$\rightarrow$ independent of change.

Note:-

$$Var(X_1+X_2+\dots+X_n)$$
$$= Var(X_1)+\dots+Var(X_n) + 2\sum_{i=1}^{n}\sum_{j=i+1}^{n} Cov(X_i, X_j)$$

Weibull Distribution :-

$X \sim W(\alpha, \beta)$

$$f_X(x) = \begin{cases} \alpha\beta x^{\beta-1} e^{-\alpha x^{\beta}} & , \forall x>0, \alpha>0, \beta>0 \\ 0 & , o.w \end{cases}$$

$$F_X(x) = \begin{cases} 1-e^{-\alpha x^{\beta}} & , x>0 \\ 0 & , x \le 0 \end{cases}$$

$$\text{Mean} = \mu = \frac{\Gamma\left(\frac{\beta+1}{\beta}\right)}{\alpha^{1/\beta}}$$

$$\mu_K = \frac{\Gamma\left(\frac{\beta+K}{\beta}\right)}{\alpha^{K/\beta}}$$

$$\text{Variance} = \sigma^2 = \alpha^{-2/\beta}\left[\Gamma\left(\frac{\beta+2}{2}\right) - \left(\Gamma\left(\frac{\beta+1}{\beta}\right)\right)^2\right]$$

Sampling Distribution

Population $\rightarrow$ group of objects under study

Sample $\rightarrow$ A finite subset of statistical individuals

Sampling $\rightarrow$ The process of selecting a sample s.t.,

- to get max info. with minimum inputs
- find limits of parameters and degree of confidence

Statistic $\rightarrow$ Statistical measures computed from sample observation

Sampling Distribution $\rightarrow$ Probability distribution of statistic. (eg:- $\bar{x} \sim N(\mu, \frac{\sigma^2}{n})$)

Central Limit Theorem :- (CLT)

$X_1, X_2, \dots, X_n \rightarrow$ identicle, independent RVs (iid)

$\rightarrow$ each with mean μ and var σ^2

let, $S_n = (X_1+X_2+\dots+X_n)/n = \bar{X}$

mean $= \sum_{1}^{n} \mu_i = \mu$

Var $= \sigma^2/n$

$$\boxed{\text{if } Z_n = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}, \text{ then } \lim_{n\to\infty} \frac{F_n(z)}{\phi(z)} = 1}$$

i.e. if we have a large number of iid random variables then the Distribution for the sum of these random variables is approximately the same as Standard Normal distribution

$$Z_n \sim N(0,1)$$

$$\forall \; Z_n = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}, \; n \to \infty$$

→ Generally $n \geq 30$.

Random Sampling :-

$X_1, X_2, X_3, \ldots, X_n$ → n iid r.v's each with pdf $f(x)$

then $(x_1, x_2, \ldots, x_n)$ is a random sample if the joint distribution is,

$$\boxed{f(x_1, x_2, \ldots, x_n) = f(x_1) f(x_2) \cdots f(x_n)}$$

Statistic :- a function of random sample.

eg:- $\bar{X} = \frac{1}{n}(X_1 + X_2 + \cdots + X_n)$ (Sample Mean)

$S^2 = \frac{1}{(n-1)} \sum (X_i - \bar{X})^2$ (Sample Variance)

$X_{(\frac{n+1}{2})}$ if n is odd

$\frac{X_{(\frac{n}{2})} + X_{(\frac{n}{2}+1)}}{2}$ if n is even

(Sample Median)

Sample Range : $X_{(n)} - X_{(1)}$

☆ iid → identically and independently distributed

Sampling distribution of Mean :-

we assume that the concerned population has a mean μ and variance σ^2.

if $(x_1, x_2, \ldots, x_n)$ is a random sample then,

Sample Mean $= \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$

$$\boxed{E(\bar{x}) = \frac{1}{n} \sum E(x) = \mu}$$

$$\boxed{Var(\bar{x}) = \frac{\sigma^2}{n}}$$

→ This is known as Standard Error of $\bar{x}$ (Sample Mean).

let $X_{11}, X_{12}, \ldots, X_{1n}$ → iid RV's with mean μ_1, and var. σ_1^2 and $X_{21}, X_{22}, \ldots, X_{2n}$ → iid RV's with mean μ_2 and var σ_2^2.

if $\bar{X}_1 = \frac{\sum x_{1i}}{n}$ & $\bar{X}_2 = \frac{\sum x_{2i}}{n}$ then,

$$\boxed{\frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{n}}} \to N(0,1) \text{ as } n_1 \to \infty, \; n_2 \to \infty}$$

Distributions arising from Normal Distribution :-

Chi-Square Distribution :-

$Z_1, Z_2, \ldots, Z_K$ → normally and iid r.v's with $\mu = 0, \sigma^2 = 1$ each.

$$\chi^2 = Z_1^2 + Z_2^2 + \cdots + Z_K^2$$

$$\chi^2 = \sum_{1}^{n} \left(\frac{X_i - \mu_i}{\sigma_i}\right)^2$$

Chi Square Variable

$$f(x) = \begin{cases} \dfrac{x^{K/2-1} \cdot e^{-x/2}}{2^{K/2}\,\Gamma(K/2)} & , x > 0 \\ 0 & , \text{o.w.} \end{cases}$$

$K \to$ degree of freedom.

Mean $= \mu = K$

Variance $= \sigma^2 = 2K$

$$\boxed{P\{\chi_K^2 \geq \chi^2_{\alpha,K}\} = \int_{\chi^2_{\alpha,K}}^{\infty} f(x)\,dx = \alpha}$$

$$M_{\chi^2}(t) = MGF = E(e^{tx}) = (1-2t)^{-K/2}$$

<u>Note</u>:- • χ^2 is identicle to Gamma$(\frac{K}{2}, \frac{1}{2})$

$$f(x) = \frac{\frac{1}{2} e^{-\frac{x}{2}} \left(\frac{x}{2}\right)^{\frac{K}{2}-1}}{\Gamma(\frac{K}{2})}, \quad x > 0$$

• Sum of chi-sq. random var.,

$$\chi^2_{\Sigma K_i}$$

t-Distribution :-

$Z \sim N(0,1)$

$V \to$ chi-sq. r.v. with K deg. of freedom. (χ^2_K)

if Z and V are independent then,

$$\boxed{t_K = \frac{Z}{\sqrt{V/K}}}$$

$$\boxed{f(t) = \frac{\Gamma\left(\frac{K+1}{2}\right)}{\sqrt{K\pi}\,\Gamma\left(\frac{K}{2}\right)} \times \frac{1}{\left(\frac{t^2}{K}+1\right)^{\left(\frac{K+1}{2}\right)}}}$$

with K degree of freedom

Mean $= \mu = 0$

Var $= \sigma^2 = \dfrac{K}{K-2}$

Student's t-distribution :-

$X \sim N(0,1)$, $Y \sim \chi^2_n$ } independent then

$T = \dfrac{X}{\sqrt{Y/K}}$ is said to be Student's t-distribution,

transformation,

$T = \dfrac{X}{\sqrt{Y/K}}$, $U = Y \Rightarrow X = \sqrt{\dfrac{U}{K}}\,T$, $y = u$

$$\boxed{f_{T,U}(t,u) = \frac{1}{2^{\frac{K+1}{2}}\sqrt{K\pi}\,\Gamma\left(\frac{K}{2}\right)} e^{-\frac{u}{2}\left(1+\frac{t^2}{K}\right)} \cdot u^{\frac{K+1}{2}-1}}$$

$\forall\ -\infty < t < \infty$, $u > 0$

$\to$ T is symmetric at $t=0$

$\to$ odd order moments vanish.

Marginal pdf

F-Distribution :-

$W, Y \to$ Chi-Square r.v. with u, v degree of freedom.

$$F = \frac{W/u}{Y/v}$$

$$h(f) = \frac{\Gamma\left(\frac{u+v}{2}\right)\left(\frac{u}{v}\right)^{u/2} f^{\left(\frac{u}{2}\right)-1}}{\Gamma\left(\frac{u}{2}\right)\Gamma\left(\frac{v}{2}\right)\left[\frac{u}{v}f+1\right]^{\frac{u+v}{2}}}, \quad 0 < f < \infty$$

Mean $= \mu = \dfrac{v}{v-2}$, $v > 2$

Var $= \sigma^2 = \dfrac{2v^2(u+v-2)}{u(v-2)^2(v-4)}$, $v > 4$

$$\boxed{P\{F \geq F_{\alpha,u,v}\} = \int_{F_{\alpha,u,v}}^{\infty} h(f)\,df = \alpha}$$

Parameter Estimation

Point Estimation :-

a single numerical value of a statistic that corresponds to that parameter.

Eg :-

① Mean μ , $\hat{\mu} = \bar{X}$ → Sample Mean

↓ estimator of μ

② Var. σ^2 , $\hat{\sigma}^2 = S^2$ → Sample variance

③ $\mu_1 - \mu_2$, $\hat{\mu}_1 - \hat{\mu}_2 = \bar{X}_1 - \bar{X}_2$

④ Proportion P , $\hat{P} = X/n$

(A) Properties of Estimator :-

$\hat{\theta}$ is unbiased estimator of θ if,

$$E(\hat{\theta}) = \theta$$

→ Mean Square Error of estimator,

$$\boxed{MSE(\hat{\theta}) = E(\hat{\theta} - \theta)^2 = Var(\hat{\theta}) + \underbrace{(\theta - E(\hat{\theta}))}_{\text{bias}}^2}$$

↓

important when comparing two estimators,

The relative efficiency of $\hat{\theta}_2$ to $\hat{\theta}_1$ is,

$$= \frac{MSE(\hat{\theta}_1)}{MSE(\hat{\theta}_2)}$$

→ Minimum Variance of unbiased Estimator $\hat{\theta}$ is,

$$V(\hat{\theta}) \geq \frac{1}{n E\left(\frac{d}{d\theta} \ln(f(x,\theta))\right)^2}$$

↳ Cramer - Rao Lower Bound

(B) Method of Maximum Likelihood :-

$x_1, x_2, \ldots, x_n$ → observed value in a random sample of size n.

$f(x, \theta)$ → Prob. dist. fn

↳ Unknown parameter.

$$\boxed{L(\theta) = f(x_1, \theta) \cdot f(x_2, \theta) \ldots f(x_n, \theta)}$$

The value of θ that maximises $L(\theta)$ is called Maximum likelihood Estimator (MLE) of θ.

Eg :-

For mean μ, MLE $= \bar{X}$

(C) Method of Moments :-

$\theta_1, \theta_2, \ldots, \theta_k$ → K Unknown Parameters

$X_1, X_2, \ldots, X_n$ → Random Sample of size n. from X

$f(x; \theta_1, \theta_2, \ldots, \theta_k)$ } Cont.
$P(x; \theta_1, \theta_2, \ldots, \theta_k)$ } Disc. → Probability functions

$$\boxed{m'_t = \frac{1}{n} \sum_{i=1}^{n} X_i^t} \quad \forall\, t = 1, 2, \ldots, K$$

↳ first K sample moments about origin.

$$\boxed{\begin{aligned}\mu_t' = E(X^t) &= \int_{-\infty}^{\infty} x^t f(x;\theta_1,\dots\theta_k)\,dx \quad \text{Cont.}\\ &= \sum_{x\in R_x} x^t P(x;\theta_1,\dots\theta_k) \quad \text{Disc.}\end{aligned}}$$

↳ first K population moments about the origin.

$$\boxed{\mu_t' = m_t'} \quad \forall t = 1,2,\dots,K$$

↓ Solving these K equations gives,

$\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_K$

(Moment estimators of $\theta_1, \theta_2, \dots \theta_K$)

Precision of Estimation :-

$$\sigma_{\hat{\theta}} = \sqrt{V(\hat{\theta})}$$

↳ Standard error of $\hat{\theta}$ = Standard deviation of $\hat{\theta}$

Single-Sample Confidence interval Estimation :-

$$P\{L \le \theta \le U\} = 1-\alpha$$

$\theta \to$ Unknown parameter

L & $U \to$ Lower- and Upper-Confidence limits

$1-\alpha \to$ Confidence Coefficient

→ One-sided $100(1-\alpha)\%$ lower Confidence interval on θ,

$$P(L \le \theta) = 1-\alpha$$

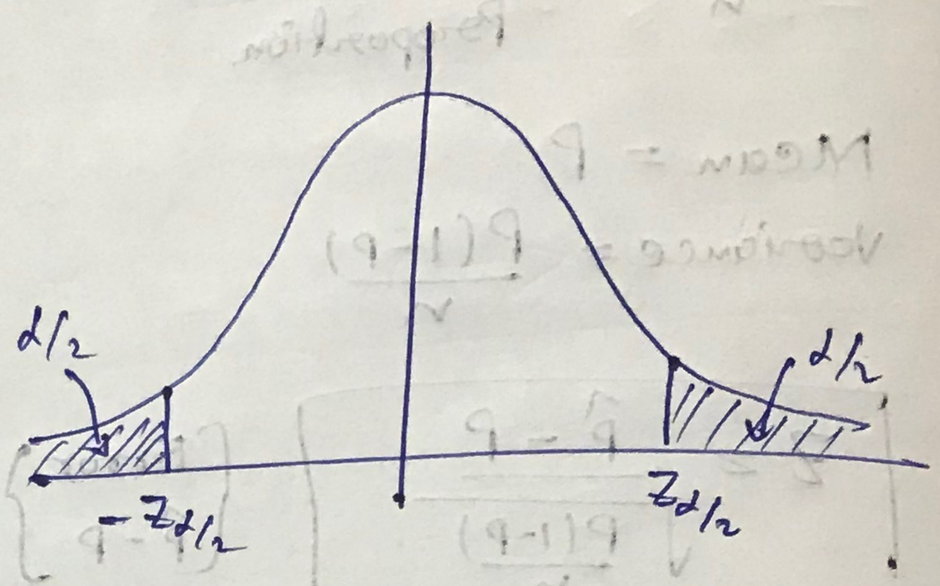
Similary upper.

→ $\theta - L$ or $U - \theta \to$ accuracy of the estimator

(A) Confidence Interval on the Mean of Normal Distribution :-

$$Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \quad \left\{\begin{matrix}\text{error}\\ \bar{X}-\mu\end{matrix}\right\}$$

$$P\{-Z_{\alpha/2} \le Z \le Z_{\alpha/2}\} = 1-\alpha$$



So $100(1-\alpha)\%$ two-sided Confidence interval on μ is,

$$\boxed{\bar{X} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{X} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}}$$

if Variance is Unknown,

$$t = \frac{\bar{X}-\mu}{S/\sqrt{n}}$$

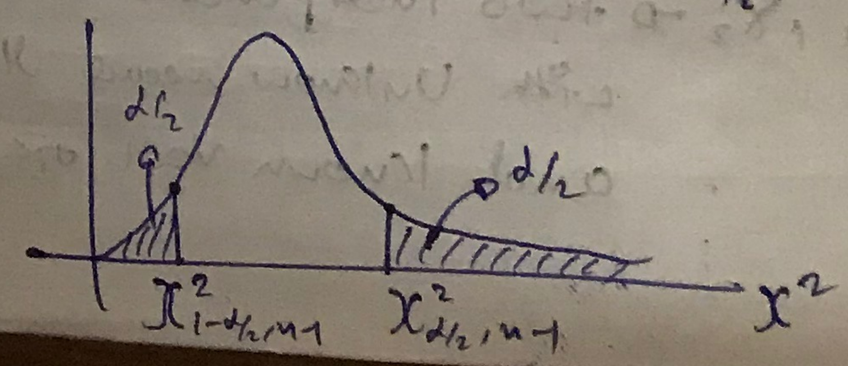
$$P\{-t_{\alpha/2,n-1} \le t \le t_{\alpha/2,n-1}\} = 1-\alpha$$

$$\boxed{\bar{X} - t_{\alpha/2,n-1}\, S/\sqrt{n} \le \mu \le \bar{X} + t_{\alpha/2,n-1}\, S/\sqrt{n}}$$

(B) Confidence interval on Variance of Normal distribution :-

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2}$$

$$P\{\chi^2_{1-\alpha/2,n-1} \le \chi^2 \le \chi^2_{\alpha/2,n-1}\} = 1-\alpha$$



$$\boxed{\frac{(n-1)S^2}{\chi^2_{\alpha/2,n-1}} \le \sigma^2 \le \frac{(n-1)S^2}{\chi^2_{1-\alpha/2,n-1}}}$$

(C) Confidence Interval on a Proportion :-

$\hat{p} = \frac{X}{n}$ → Point estimator of Proportion

Mean = P

Variance = $\frac{P(1-P)}{n}$

$$\boxed{Z = \frac{\hat{P} - P}{\sqrt{\frac{P(1-P)}{n}}}}$$

$\left\{\begin{matrix}\text{Error} \\ \hat{P} - P\end{matrix}\right\}$

$\forall\, n$ → sufficiently large

→ approximately Standard Normal

$$P\left\{-Z_{\alpha/2} \le Z \le Z_{\alpha/2}\right\} = 1-\alpha$$

$100(1-\alpha)\%$ two-sided Confidence interval on P is :-

$$\boxed{\hat{p} - Z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \le P \le \hat{p} + Z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}$$

Two-Sample Confidence Interval Estimation :-

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

X_1, X_2 → two independent r.v's with Unknown means μ_1, μ_2 and known var. σ_1^2, σ_2^2.

$$P\left\{-Z_{\alpha/2} \le Z \le Z_{\alpha/2}\right\} = 1-\alpha$$

$$\boxed{\bar{X}_1 - \bar{X}_2 - Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \le \mu_1 - \mu_2 \le \bar{X}_1 - \bar{X}_2 + Z_{\alpha/2}\sqrt{\text{—}}}$$

(A) Confidence interval on the difference b/w means of two normal distributions, Variance Unknown :-

$\{\sigma_1^2 = \sigma_2^2 = \sigma\}$

$$S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$$

$$t = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$P\left\{-t_{\frac{\alpha}{2},n_1+n_2-2} \le t \le t_{\alpha/2,n_1+n_2-2}\right\} = 1-\alpha$$

$$\boxed{\bar{X}_1 - \bar{X}_2 - t_{\alpha/2,n_1+n_2-2}\, S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \le \mu_1 - \mu_2 \le \text{—} + \text{—}}$$

(B) Confidence Interval on the ratio of variance of two Normal distributions :-

$$F = \frac{S_2^2/\sigma_2^2}{S_1^2/\sigma_1^2}$$

$$\boxed{\frac{S_1^2}{S_2^2} F_{1-\frac{\alpha}{2},n_2-1,n_1-1} \le \frac{\sigma_1^2}{\sigma_2^2} \le \frac{S_1^2}{S_2^2} F_{\frac{\alpha}{2},n_2-1,n_1-1}}$$

Tests of Hypotheses :-

Statistical Hypotheses :-

eg :-

$H_0 : \mu = 2500$ psi

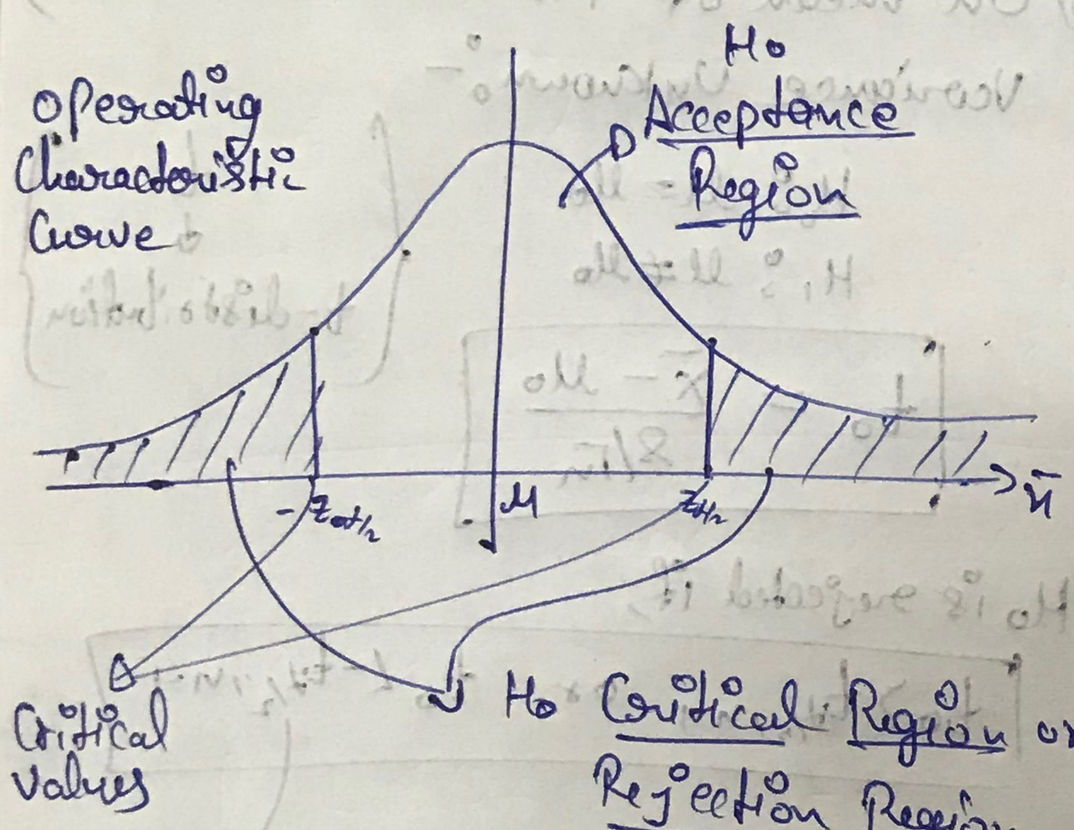
$H_1 : \mu \neq 2500$ psi

$H_0 \rightarrow$ Null Hypotheses.

$H_1 \rightarrow$ alternative hypothesis.

$H_1 : \mu \neq 2500$ psi $\rightarrow$ two-sided alt. hypo. ($\mu > 2500$ or $\mu < 2500$)

$H_1 : \mu > 2500$ psi $\rightarrow$ One-sided alt. hypo.



Type I and II errors :-

$\alpha = P(\text{type I error}) = P\left\{\text{reject } H_0 / H_0 \text{ is true}\right\}$

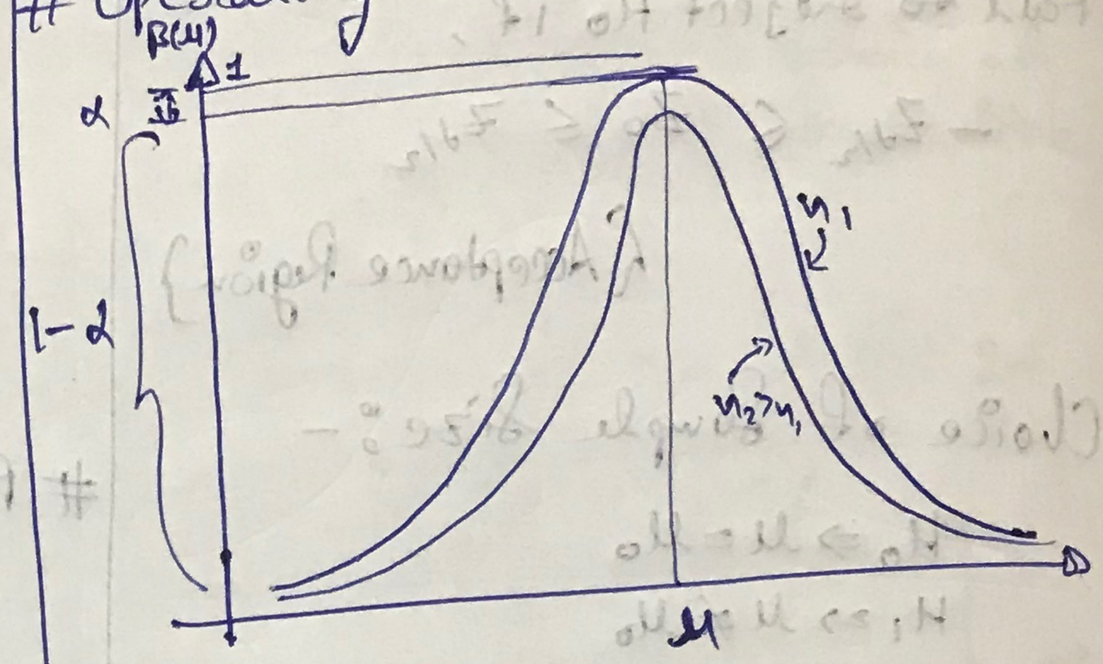
$\beta = P(\text{type II error}) = P\left\{\text{accept } H_0 / H_0 \text{ is false}\right\}$

Power of the test $= 1 - \beta = P\left\{\text{reject } H_0 / H_0 \text{ is false}\right\}$

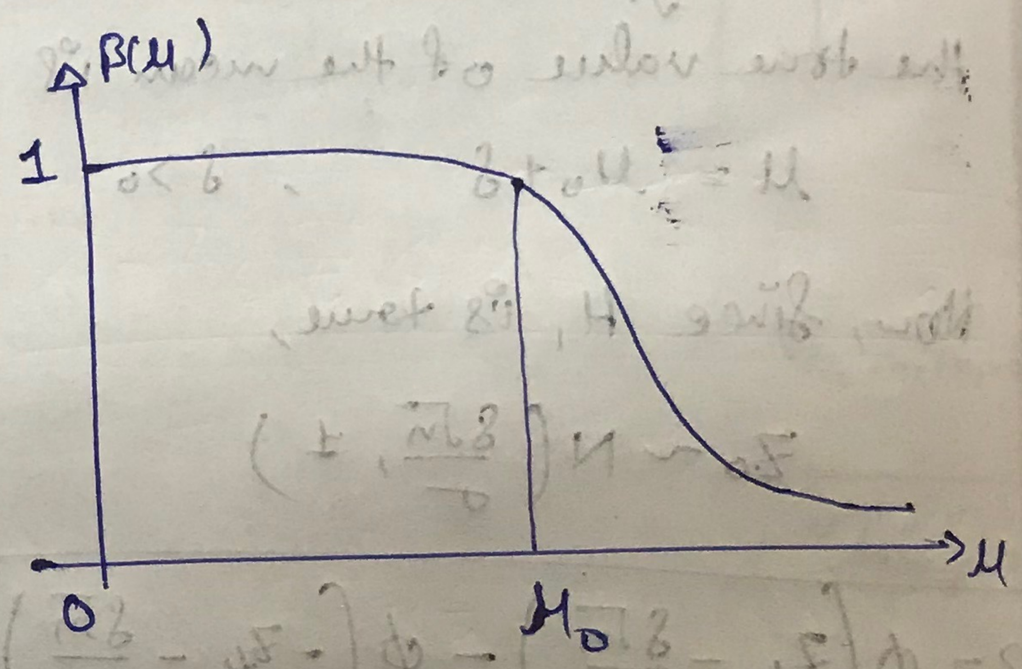
☆ $\boxed{\beta(\mu) = 1 - \alpha(\mu)}$

Note :- The probability of Type I error α is often called the significance level or size of the test.

Operating Characteristic Curve :-



Two-Sided test



One-sided test

Test of Hypotheses on a single sample

(A) On mean of a Normal Dist., Variance Known :-

$H_0 : \mu = \mu_0$

$H_1 : \mu \neq \mu_0$

now for a random sample of size n,

$$Z_0 = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

(26) → Reject H_0 if either,

$$Z_0 > Z_{\alpha/2}$$

or

$$Z_0 < Z_{\alpha/2}$$

{Critical or Rejection Region}

→ Fail to reject H_0 if,

$$-Z_{\alpha/2} \leq Z_0 \leq Z_{\alpha/2}$$

{Acceptance Region}

Choice of Sample Size :-

$H_0 \Rightarrow \mu = \mu_0$

$H_1 \Rightarrow \mu \neq \mu_0$

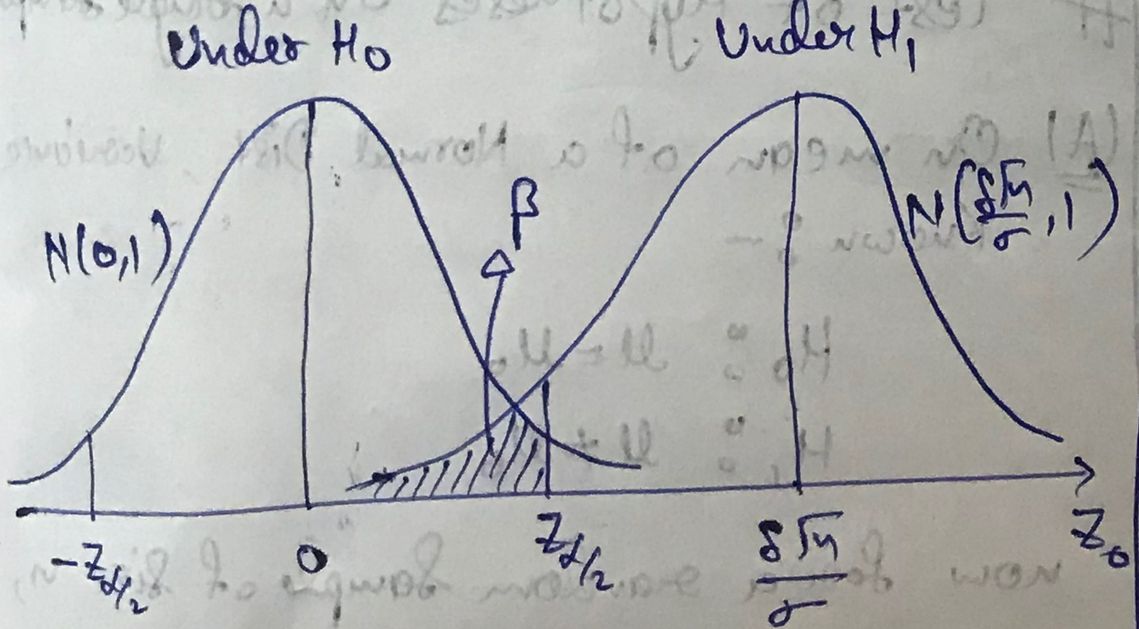
if null hypothesis is false and the true value of the mean is

$$\mu = \mu_0 + \delta \quad , \quad \delta > 0$$

Now, since H_1 is true,

$$Z_0 \sim N\left(\frac{\delta\sqrt{n}}{\sigma}, 1\right)$$

$$\boxed{\beta = \phi\left(Z_{\alpha/2} - \frac{\delta\sqrt{n}}{\sigma}\right) - \phi\left(-Z_{\alpha/2} - \frac{\delta\sqrt{n}}{\sigma}\right)}$$



$$\boxed{\text{Parameter } d = \frac{|\mu - \mu_0|}{\sigma} = \frac{|\delta|}{\sigma}}$$

if $\delta > 0$,

$$\phi\left(-Z_{\alpha/2} - \frac{\delta\sqrt{n}}{\sigma}\right) \approx 0$$

So $\beta \approx \phi\left(Z_{\alpha/2} - \frac{\delta\sqrt{n}}{\sigma}\right)$

$$-Z_\beta \approx Z_{\alpha/2} - \frac{\delta\sqrt{n}}{\sigma}$$

$$\boxed{n \approx \frac{(Z_{\alpha/2} + Z_\beta)^2 \sigma^2}{\delta^2}}$$

P-value → The smallest value of significance that would lead to rejection of H_0.

(B) On mean of Normal Distribution, Variance Unknown :-

$H_0 : \mu = \mu_0$

$H_1 : \mu \neq \mu_0$

{t → t-distribution}

$$\boxed{t_0 = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}}$$

H_0 is rejected if,

$$\boxed{t_0 > t_{\alpha/2, n-1} \text{ or } t_0 < -t_{\alpha/2, n-1}}$$

→ for, $H_0 : \mu = \mu_0$

$H_1 : \mu > \mu_0$

H_0 is rejected if,

$$\boxed{t_0 > t_{\alpha, n-1}}$$

$\alpha/2$

α

(C) On Variance of Normal Dist. :-

{for Normal Population}

$H_0 : \sigma^2 = \sigma_0^2$

$H_1 : \sigma^2 \neq \sigma_1^2$

$$\chi_0^2 = \frac{(n-1)s^2}{\sigma_0^2}$$

H_0 is rejected if,

$$\chi_0^2 > \chi^2_{\alpha/2, n-1} \quad \text{"or"}$$

$$\chi_0^2 < \chi^2_{1-\alpha/2, n-1}$$

→ for large sample :-

$$\boxed{Z_0 = \frac{s-\sigma}{\sigma/\sqrt{2n}}}$$

(D) On a Proportion :-

$H_0 : P = P_0$

$H_1 : P \neq P_0$

$$\boxed{Z_0 = \frac{X - nP_0}{\sqrt{nP_0(1-P_0)}}}$$

H_0 rejected if,

$$\boxed{Z_0 > Z_{\alpha/2} \quad \text{or} \quad Z < -Z_{\alpha/2}}$$

Tests of Hypotheses on Two Samples :-

(A) On Means of two Normal distn, Variance Known :-

$H_0 : \mu_1 = \mu_2$

$H_1 : \mu_1 \neq \mu_2$

$$\bar{X}_1 - \bar{X}_2 \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

$$\boxed{Z_0 = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}}$$

→ Since it is ~~too~~ true H_0 i.e. $\mu_1 = \mu_2$

H_0 is rejected if,

$$\boxed{Z_0 > Z_{\alpha/2} \quad \text{or} \quad Z_0 < Z_{\alpha/2}}$$

$$\boxed{d = \frac{|\delta|}{\sqrt{\sigma_1^2 + \sigma_2^2}}}$$

$$\boxed{n = \frac{\sigma_1^2 + \sigma_2^2}{\sigma_1^2/n_1 + \sigma_2^2/n_2}}$$ → choice of sample size

(B) Variance Unknown in (A) :-

$$s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1 + n_2 - 2}$$

$$\boxed{t_0 = \frac{\bar{X}_1 - \bar{X}_2}{s_p\sqrt{1/n_1 + 1/n_2}}}$$

$\sigma_1^2 = \sigma_2^2$

H_0 is rejected if,

$$t_0 > t_{\alpha/2, n_1+n_2-2} \quad \text{or} \quad t_0 < -t_{\alpha/2, n_1+n_2-2}$$

$$\boxed{t_0^* = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}}$$ → $\{\sigma_1^2 \neq \sigma_2^2\}$

$$\boxed{d = \frac{|\delta|}{2\sigma}}$$