Discrete Distribution.

(μ) mean $\rightarrow$ $E[X]$

(σ^2) variance $\rightarrow$ $E[X^2] - (E[X])^2 \rightarrow E[X-\mu]^2$

Binomial / Bernoulli Dist $\rightarrow$ $\boxed{b(x; n, p) = {}^{n}C_{x}\, p^{x} \cdot q^{n-x}}$

don't know when success will come.

$n \rightarrow$ small.

$\boxed{\mu = np}$ $\boxed{\sigma^2 = npq}$

-ve Binomial Dist $\rightarrow$ $\boxed{b^{*}(x; K, p) = {}^{x-1}C_{K-1} \cdot p^{K} \cdot q^{x-K}}$

until a fix number of success occurs.

K^{th} success in x^{th} trial from n total trials.

no. of trials required

$\boxed{\mu = K/p}$

$\boxed{\sigma^2 = Kq/p^2}$

Geometric distribution $\rightarrow$ $\boxed{g(x; p) = p(q^{x-1})}$

no. of trials for 1st success

$\boxed{\mu = 1/p}$ $\boxed{\sigma^2 = q/p^2}$

Poisson distribution $\rightarrow$ $\boxed{T = P(x; \lambda t) = \frac{e^{-\lambda t} \cdot (\lambda t)^{x}}{x!}}$

outcome occuring in a given time interval or region.

outcome needed.

Mean = Var = λt

for $n \rightarrow \infty$, $p \rightarrow 0$, $\mu = \text{const}$

★ $b(x; n, p) \xrightarrow{n \rightarrow \infty} P(x; \lambda t)$

$\boxed{P(x; \mu) = \frac{e^{-\mu} \cdot (\mu)^{x}}{x!}}$

Mean = Variance

↳ in Poission Distri

Mean > Var

↳ in B.D

Variance > Mean

↳ GD

Continuous Distribution.

Normal Dist.	Gamma Dist	Exponential Dist.

Normal Dist.

↓

$n \to$ large $\to \infty$

$$\boxed{f(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}}$$

↳ denoted as

$$\boxed{X \sim N(\mu,\sigma^2) \text{ or } n(x;\mu,\sigma)}$$

Gamma Dist

$$\boxed{\Gamma\alpha = \int x^{\alpha-1} \cdot e^{-\alpha} \cdot dx}$$

$x > 0$

$\Gamma\tfrac{1}{2} = \sqrt{x}$

K^{th} distribution occurs in what time?

$\Gamma n = (n-1)!$ → n = +ve int

$\Gamma\tfrac{1}{2} = \sqrt{x}$ $\quad \Gamma 1 = 0! = 1$

Exponential Dist.

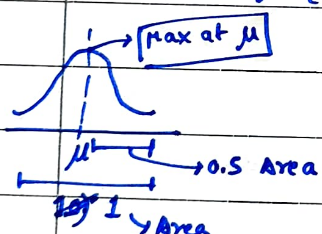
↳ for $\alpha = 1$ in Gamma

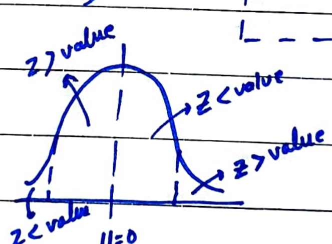
after K^{th} event the time period for $(K+1)^{th}$ Event.

Property

↳ mean = mode = median.

↳ symmetric about origin (μ = mean)*

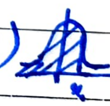




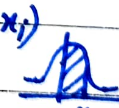
If $\boxed{\text{mean} = 0, \ \text{var} = 1}$ → $Z = \frac{X-\mu}{\sigma}$

↳ standard normal variate.

1st method Table → $(Z < x_i)$



2nd Table → $(0 < Z < x_i)$



Area under curve.

$$\boxed{\frac{1}{\sigma\sqrt{2\pi}} \int_{x_1}^{x_2} \exp\left\{\frac{-(x-\mu)^2}{2\sigma^2}\right\} \cdot dx}$$

if $\mu = 12, \sigma = 4$

$X \geq 20$ → $x = 20$

$Z = \frac{X-\mu}{\sigma} = \frac{20-12}{4} = 2$

Mean = μ ★ $\quad$ Var = $\sigma^2 + \mu^2$

__/__/____

Normal Dist as special case of Binomial. dist.

↳ just like possion it is another limiting case.

↳ $n \to \infty$

↳ p and q $\not\to$ small.

if binomial ⇄ Normal then $\boxed{Z = \frac{X - \mu}{\sigma}}$ ($\mu \to np$, $\sigma \to \sqrt{npq}$)

as $n \to \infty$

$n(x;\mu,\sigma) \Rightarrow n(z;0,1)$

Gamma and Exponential dist.

$$\boxed{f(x;\alpha,\beta) = \begin{cases} \frac{1}{\beta^{\alpha}\Gamma\alpha} \cdot x^{\alpha-1} \cdot e^{-x/\beta} & , x>0 \\ 0, \text{elsewhere.} \end{cases}}$$

where $\to \alpha>0$, $\beta>0$

$\lambda = 1/\beta$ → Possion

λ : Avg no. of unit per time

α = no. of events. → if $\alpha = 1$ → Expo; $\alpha \neq 1$ → Gamma

β = Avg time b/w no. of events

Expo

↳ occurs when $\boxed{\alpha = 1}$ →
$$\boxed{f(x;\beta) = \begin{cases} 1/\beta \cdot e^{-x/\beta} & , x>0 \\ 0 & , \text{else where} \end{cases}}$$

	Mean	Variance
Gamma	$\alpha\beta$	$\alpha\beta^2$
Expo	β	β^2

$\boxed{\text{NOTE}}$ –
↳ for solving Gamma Question
↳ we need to know the function by part

+, -, +, -, + ★

$$\int I.II = I\int II - \int\left[\frac{d}{dx}(I)\int II\right].dx$$

Unit - 5

Central Limit theorem

↳ states that normal distribution is the limiting distribution to the sum of the $\boxed{\text{independent}}$ random variables with finite variance as the numbers of random variables get indefinitely large.

⇒ if $X_i = \{i = 1, 2, 3, \ldots, n\}$ be independent distributed random Variables.

such that $E(X_i) = \mu_i$ and $Var(X_i) = \sigma_i^2$

then as $n \to \infty$, the distribution of the sum of these random variables,

$$\{S_n = X_1 + X_2 + X_3 + \ldots + X_n\}$$

$$\Rightarrow \mu = \sum_{i=1}^{n} \mu_i \quad ; \quad \sigma^2 = \sum_{i=1}^{n} \sigma_i^2$$

	Mean	Var	S.D	Size
pop →	μ	σ^2	σ	N
Sample →	$\bar{x}$	s^2	s	n

when n is very large.

↳ we can we bionomial probability

or

limited value as Normal distribution probability

for that Mean $\mu = np$, Variance $\sigma^2 = npq$.

$$\boxed{Z = \frac{X - \mu}{\sigma}}$$

for poisson Variable

↳ $T = P(x; \lambda t) = \dfrac{e^{-\lambda t} \cdot (\lambda t)^x}{x!}$

$\{Mean = Var = \lambda t\}$

$$Z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

{ we need to convert $\bar{x} \Rightarrow Z$ for graphic value.

CLT {Liapounoff's form}

↳ independent R.V

{ for $n \to \infty$

pop $\begin{cases} mean = \sum_{i=1}^{n} \mu_i \\ var = \sum_{i=1}^{n} \sigma_i^2 \end{cases}$

★ Sample $\begin{cases} \mu_i \ (i = 1 \to n) \\ \sigma_i^2 \ (i = 1 \to n) \end{cases}$

* CLT {Lindberg - Lavy's Theorem}

↳ independent and identically distributed r.v

pop $\begin{cases} mean \Rightarrow n\mu \\ var \Rightarrow n\sigma^2 \end{cases}$

★ Sample $\begin{cases} \mu \text{ (for every R.V)} \\ \dfrac{\sigma^2}{n} \text{ (for every R.V)} \end{cases}$

Estimator -

↳ using Estimator we find info of population with the help of sample.

any function of the random sample $x_1, x_2, \dots x_n$ say $T_n(x_1, x_2, \dots, x_n)$ is called statistic.

Characteristic of Estimator

↳ 1) Unbaisedness
2) Consistency
3) Efficiency
4) Sufficiency

unbaised → if $E[T_n] = f(\theta)$

statistic (T_n), parameter (θ)

sample info, pop info.

↳ else

$E[\bar{x}] \neq \mu \longrightarrow$ baised

* +vely baised $E[T_n] > \theta$
-vely baised $E[T_n] < \theta$

$Var = E[x^2] - (E[x])^2$

NOTE

↳ variance is always greater than 0.

~~$Var(ax+b)$~~
~~$= a^2 (x)$~~
~~$E(ax+b)$~~
~~$= a(x)+b$~~

$E[ax+b] = aE(x)+b$

$Var(ax+b) = a^2 Var(x)$

T_n tends to θ in probability.

* ⇒ (i) $E[T_n] \to f(\theta)$ as $n \to \infty$
⇒ (ii) $Var[T_n] \to 0$ as $n \to \infty$

Efficiency

↳ Var[statistics] ⟶ lesser

{less Var → more Efficient}

In Variance property

if T is consistent estimator of θ
and f is conti function of θ
then $f(T)$ is C.E of $f(\theta)$.

Sufficiency

↳ meaning: If we can get all info of parameter in sample using estimator, then that estimator is called S.E.

$$N(\mu, \sigma^2) \quad \boxed{f(x; \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}}$$

↳ if $T_n = T(x_1, x_2, x_3, \dots x_n)$ is independent of θ, and is an estimator of θ.

⇒ ie $p\left(\frac{x_1, x_2, \dots, x_n}{T_n}\right) = \frac{p((x_1, x_2, \dots, x_n) \cap T_n)}{P(T_n)}$

↳ factorisation Theorem :

↳ $T = t(x)$ is sufficient only if joint density function (L) of the sample values can be used in form of

$\boxed{L = g_\theta[t(x)] \cdot h(x)}$ or $\boxed{L = \prod_{i=1}^{n} f(x, \theta)}$ multiplication.

$g_\theta[t(x)]$ → depends on θ and x and $h(x)$ is independent of θ.

$f(x, \theta)$ function of x and θ

this function could be N.D, B.D, P.D, Gamma, Expo, anything or any other functions.

Maximum Likelihood Estimator (MLE)

is a technique used for estimating the parameters of a given distribution, i.e. (θ), using some observed data

↳ Likelihood function is given by :- joint density function.

$$L = f(x_1, x_2, x_3, \dots x_n / \theta) = \prod_{i=1}^{n} f(x_i / \theta) \quad \text{— (1)}$$

↳ for finding maximum likelihood estimator we have to maximize the likelihood function (1)

(i) $\frac{dL}{d\theta} = 0$

(ii) $\frac{d^2L}{d\theta^2} < 0$

θ is the unknown parameter

- for N.D → μ, σ^2
- for B.D → n, p
- for G.D → p (Geometric)
- for P.D → λ

in Normal Distri

↳ we have two variables μ & σ^2

$\Rightarrow \left.\frac{dL}{d\mu}\right|_{\sigma^2 = \text{const}}$ and $\left.\frac{dL}{d\sigma^2}\right|_{\mu = \text{const}}$

→ In any Question we will first find likelihood function the by maximizing it we will get (MLE).

* any where in math, if nothing is given about sample character then, by default they are independent and identically distributed

[NOTE] -

↳ Logarithm is a non-decreasing function, so for maximizing L, it is equivalently correct to maximize $\log L$

$$\frac{1}{L}\frac{dL}{d\theta} = 0 \rightarrow \frac{d\log L}{d\theta} = 0$$

Log-likelihood function is easier to work

$$\frac{d(\log x)}{dx} = \frac{1}{x}$$

$\log a^b = b \log a$

$\log 1 = 0$

$\log 0 = \infty$

$\Rightarrow L = \prod_{i=1}^{n} f(x_i / \theta)$ → will be easier if → $\boxed{\log L = \sum_{i=1}^{n} \log f(x_i / \theta)}$

if its discrete

- ↳ B.D
- ↳ P.D
- ↳ Geo. D

If its conti

- ↳ N.D
- ↳ Gamma. D
- ↳ Expo. D

if function is discrete then we will get some fixed value to satisfy that funcn the by maximizing it and then the MLE will selected from that those.

Unit-6

(i) $H_0 \to$ Null Hypo

$H_1 \to$ alternate of Null

we will choose

level of Significance → risk factor

↳ 1% 5% 10%

↳ called α

(ii) Z-test → $n \geq 30$

t-test → $n < 30$

χ^2-test → Goodness of fit

F-test → variance

$H_0 \Rightarrow \mu = \mu_0$

$H_1 \Rightarrow \mu_1 \neq \mu_0$ [Two-Tailed Test]

$\mu_1 > \mu_0$ [left-Tailed Test]

$\mu_1 < \mu_0$ [Right-Tailed Test]

one Tailed

(iii) Comparing calculated and Tabulated

↳ then Accept H_0

and Reject H_0

Error in Sampling / creating Samples.

Level of Significance

	Decision: H_0 is True	H_0 is false
Reject H_0	Type I - Error [Prob = α]	Correct decision
Accept H_0	Correct decision	Type-II-Error [prob = β]

level of Confidence

Type I ⇒ Reject H_0, when it is True.

Type II ⇒ Accept H_0, when it is false.

Z-test

↳ $\boxed{Z = \dfrac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}}$ $\sim N(0,1)$ [σ^2 is Known]

used when

$\boxed{Z = \dfrac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}}$ [σ^2 is not known]

$\boxed{s^2 = \dfrac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2}$

↳ $|z| > z_\alpha$ → Reject H_0

$|z| < z_\alpha$ → Accept H_0

Df = max independent values that have freedom to vary in sample data.

i.e → $\boxed{(n-1) \text{ is degree.}}$

mean → single mean

↘ Difference mean

Level of Significance	1%	5%	10%
T.–	2.58	1.96	1.645
R –	2.33	1.645	1.28
L –	–2.33	–1.645	–1.28

T-test

$$\boxed{t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}} \sim t(n-1)\,df$$

__/__/____

$$\boxed{s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

$$\boxed{ns^2 = (n-1)s^2}$$

$$\boxed{s^2 = \left(1 - \frac{1}{n}\right)s^2}$$

$|t| > t_\alpha \rightarrow$ Reject H_0

$|t| < t_\alpha \rightarrow$ Accept H_0

as $n \rightarrow \infty$

$s^2 = S^2$

$$\Rightarrow t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} = \frac{\bar{x} - \mu}{\sqrt{\frac{S^2}{n}}} = \frac{\bar{x} - \mu}{\sqrt{\frac{s^2}{n-1}}} = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n-1}}}$$

★ if not given greater or lesser than (two-tailed)

judicial limit / confidence limit $= \boxed{\bar{x} \pm \frac{\sigma}{\sqrt{n}} \cdot Z_\alpha}$ ★

★ Difference of mean (for two means).

↳ Z-test

let $\bar{x}_1, n_1, \sigma_1^2$ and $\bar{x}_2, n_2, \sigma_2^2$ are given → then under $H_0 \rightarrow \mu_1 = \mu_2$

for large sample

$$\boxed{Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}} \sim N(0,1)$$

or

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1}{n_1} + \frac{s_2}{n_2}}}$$

↳ t-test ↳ when $\{\mu_1 = \mu_2\}$ → is taken as H_0

↳ under assumption → $\sigma_x^2 = \sigma_y^2 = \sigma^2$

$$\boxed{t = \frac{(\bar{x} - \bar{y}) - (\mu_x - \mu_y)}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}}$$

$$\left\{ S^2 = \frac{1}{n_1 + n_2 - 2}\left[\sum_{i=1}^{n}(x_i - \bar{x})^2 + \sum_{i=1}^{n}(y_i - \bar{y})^2\right]\right.$$

★ ↳ with $\boxed{(n_1 + n_2 - 2)\,df}$

★ Paired t-test :-

↳ when sample size $\{n_1 = n_2\}$

↳ the two samples are not independent but the sample observation are paired together.

$$\boxed{\bar{d} = \frac{1}{n}\sum_{i=1}^{n} d_i}$$

sample mean of both obs, reduced to one mean

$$t = \frac{\bar{d}}{\frac{s}{\sqrt{n}}}$$

where, $\boxed{s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(d_i - \bar{d})^2}$ with $(n-1)\,df$

* χ^2-test

Mean → t-test, Z-test
Variance → f-test, χ^2 test

↳ $Df = (n-1)$ *

↳ Let $\sigma^2 = \sigma_0^2$ → specified value.

___/___/______

* → question mei frequency related ho tab.

↳ to test b/w theory and Experimental discrepency.

$$\boxed{\chi^2 = \sum_{i=1}^{n} \frac{(f_i - e_1)^2}{e_1}}$$

f_i → observed frequency

e_1 → expected frequency

$$\frac{\sum f_i}{n} = e_1$$

* f-test

↳ we use f-test for equality of two population Variance

⇒ $H_0 = \sigma_x^2 = \sigma_y^2$

$$F = \frac{\sigma_x^2}{\sigma_y^2}$$

$$\boxed{F = \frac{S_x^2}{S_y^2}}$$

S_x^2 → unbaised, always greater

S_y^2 → smaller than S_x

$$\boxed{S_x^2 = \frac{1}{(n_1-1)} \sum_{i=1}^{n_1} (x_i - \bar{x})^2}$$

$$\boxed{S_y^2 = \frac{1}{(n_2-1)} \sum_{i=1}^{n_2} (y_i - \bar{y})^2}$$

t-test

↳ difference of mean → when $\mu_1 = \mu_2$

↳ Paired t-test → when $n_1 = n_2$

sum of sq. of deviation

↳ $\sum (x_i - \bar{x})^2$