Data Science and Artificial Intelligence

Probability and Statistics

Introduction to Sampling Distribution

Lecture No.- 03

By- Rahul Sir

Topics to be Covered

Topic	Law of large numbers
Topic	CHI-square distribution

(25 questions) Problem solving

Chi-square distribution

Problems-

Session-

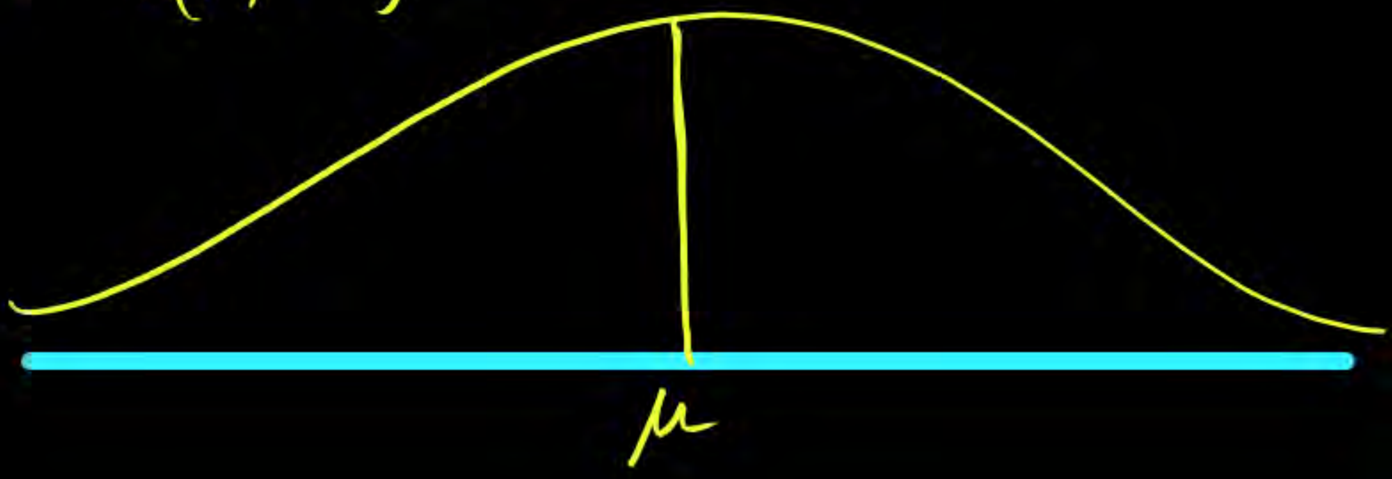
Chi-square

20 questions

Law of Large Number:

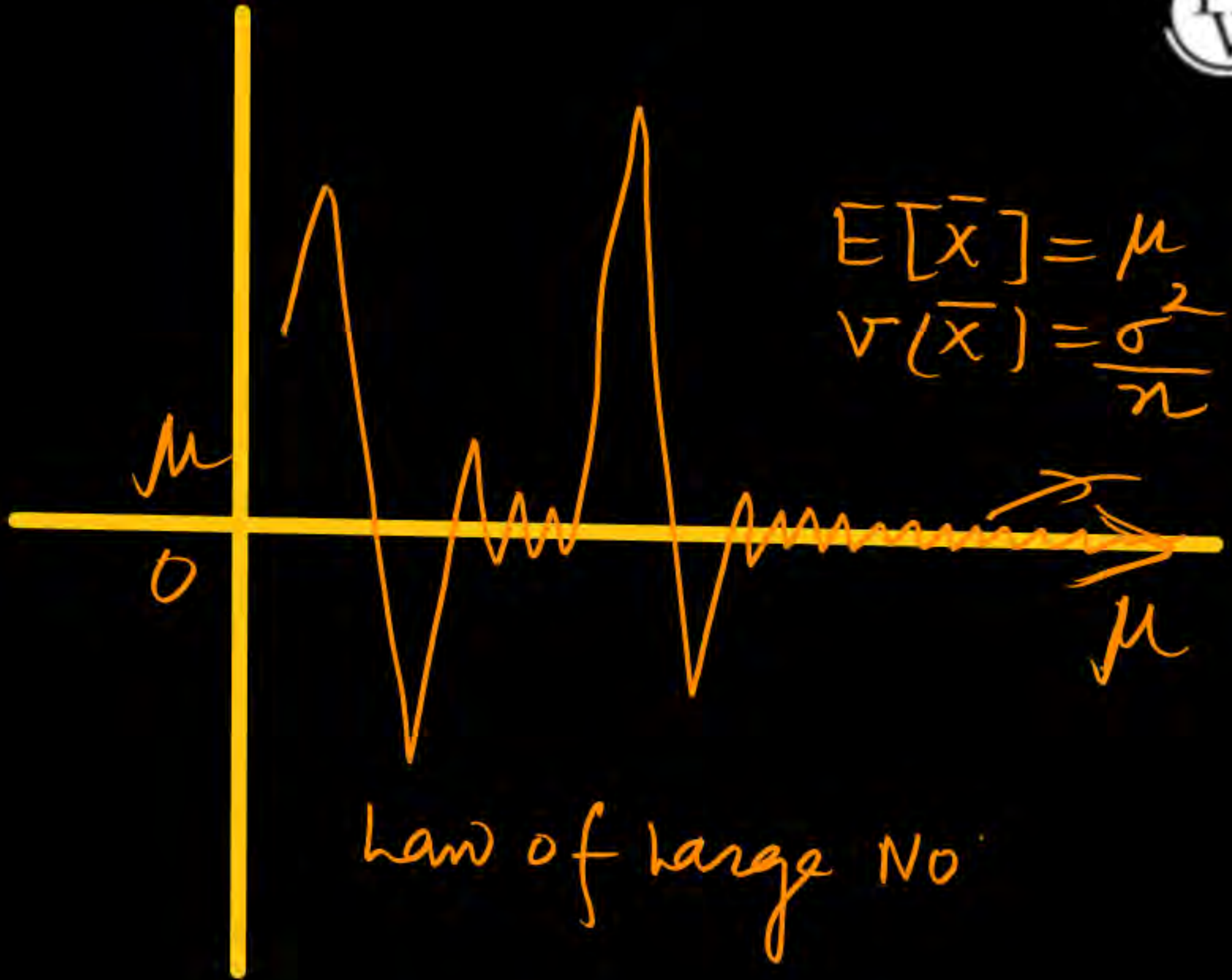
Parent distribution

$X \quad N(\mu, \sigma^2)$



$E[\bar{x}] = \mu$

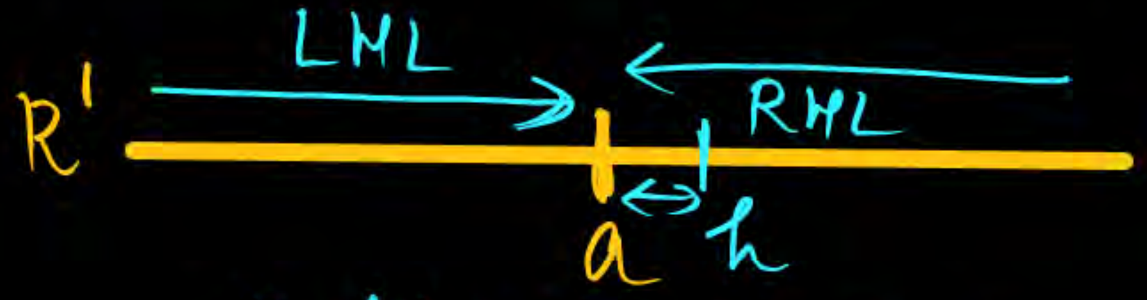
$V(\bar{x}) = \frac{\sigma^2}{n}$



Law of large No.

SAMPling Distribution

$\bar{x} \quad N\left(\mu, \frac{\sigma^2}{n}\right)$



Right Hand $= a+h$

$h =$ distance Between Two Pts $h>0$

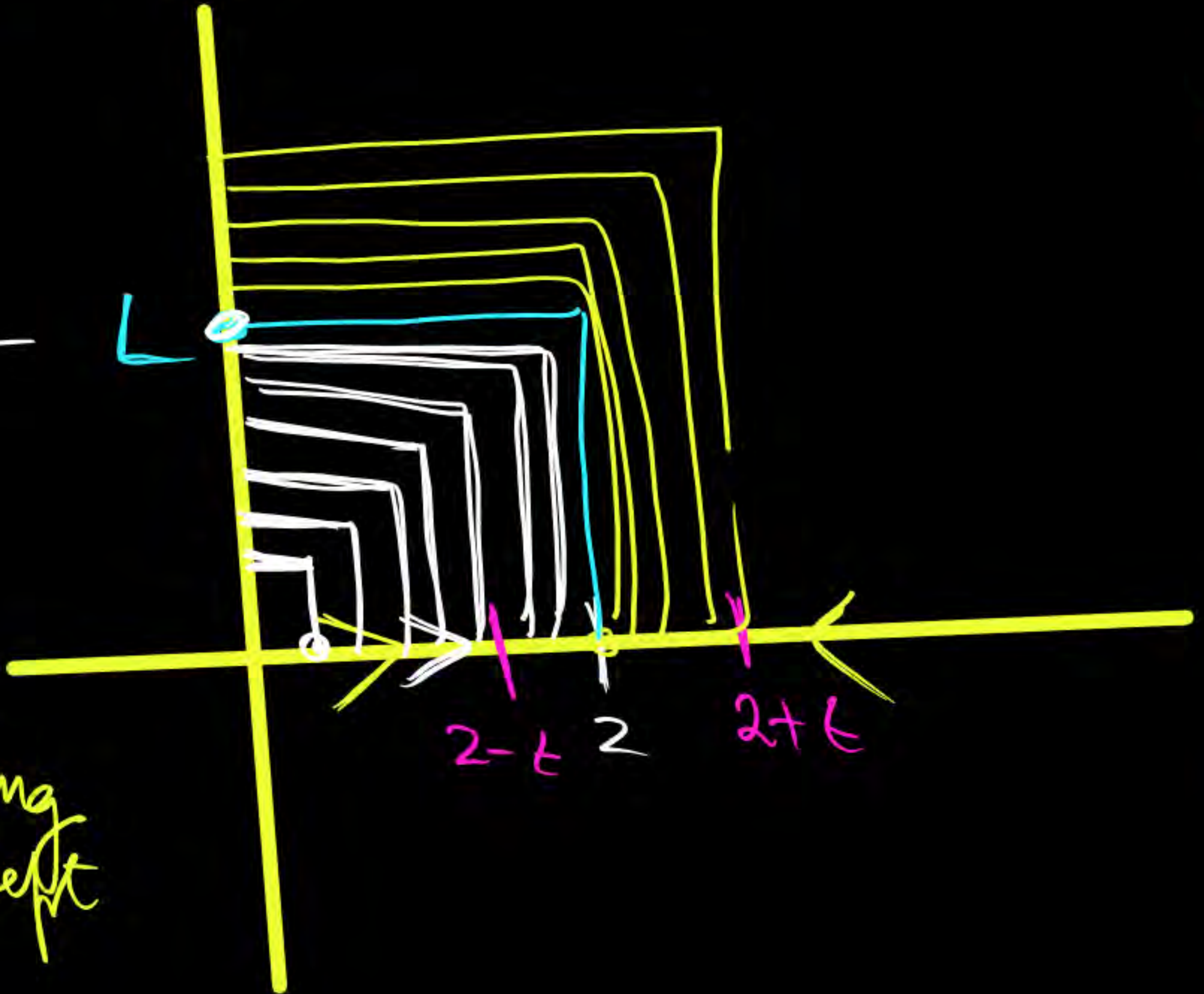
Left Hand $= a-h$

$h = \varepsilon$

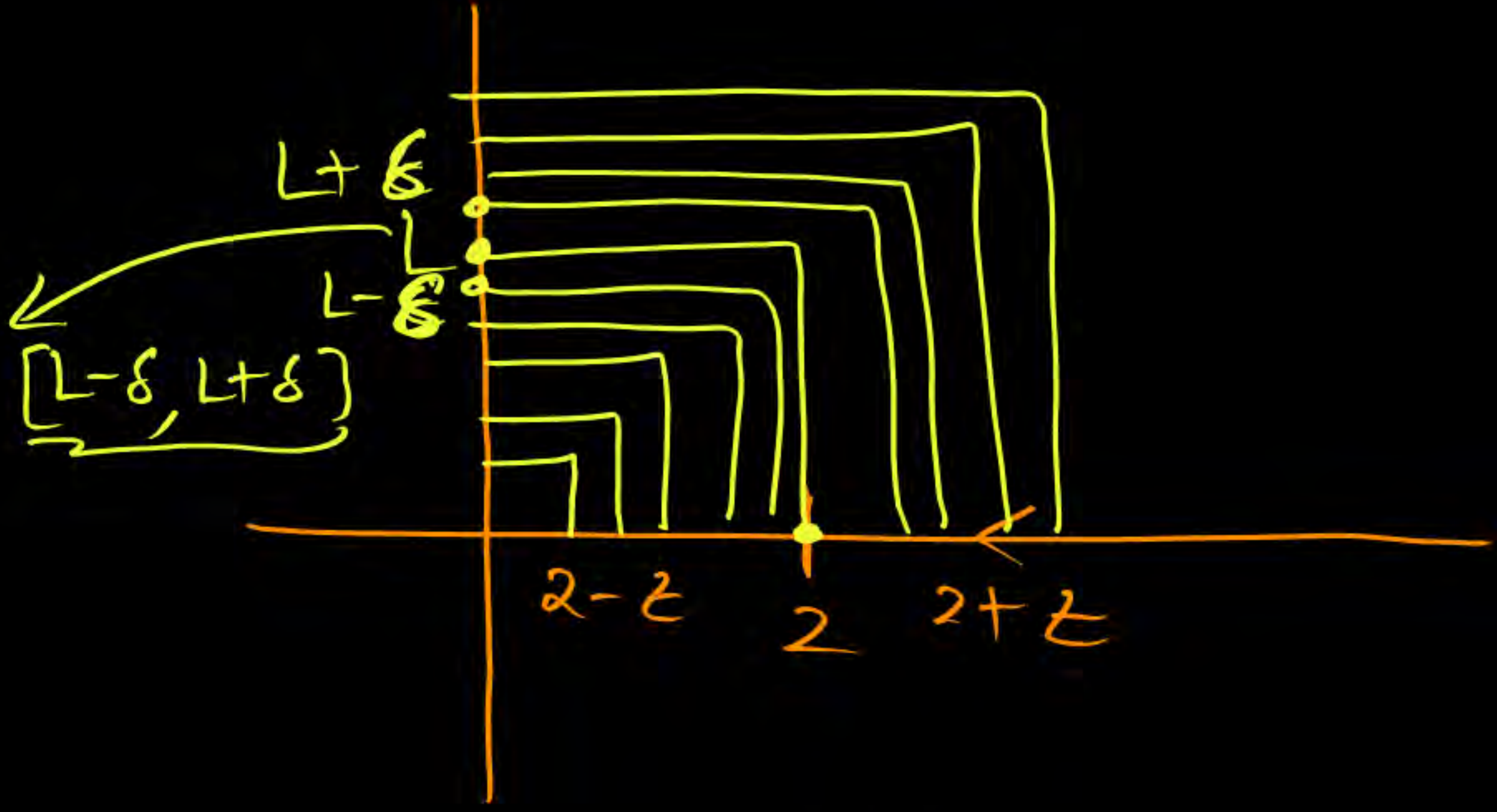
right $= a+\varepsilon$ small Distance $\varepsilon>0$

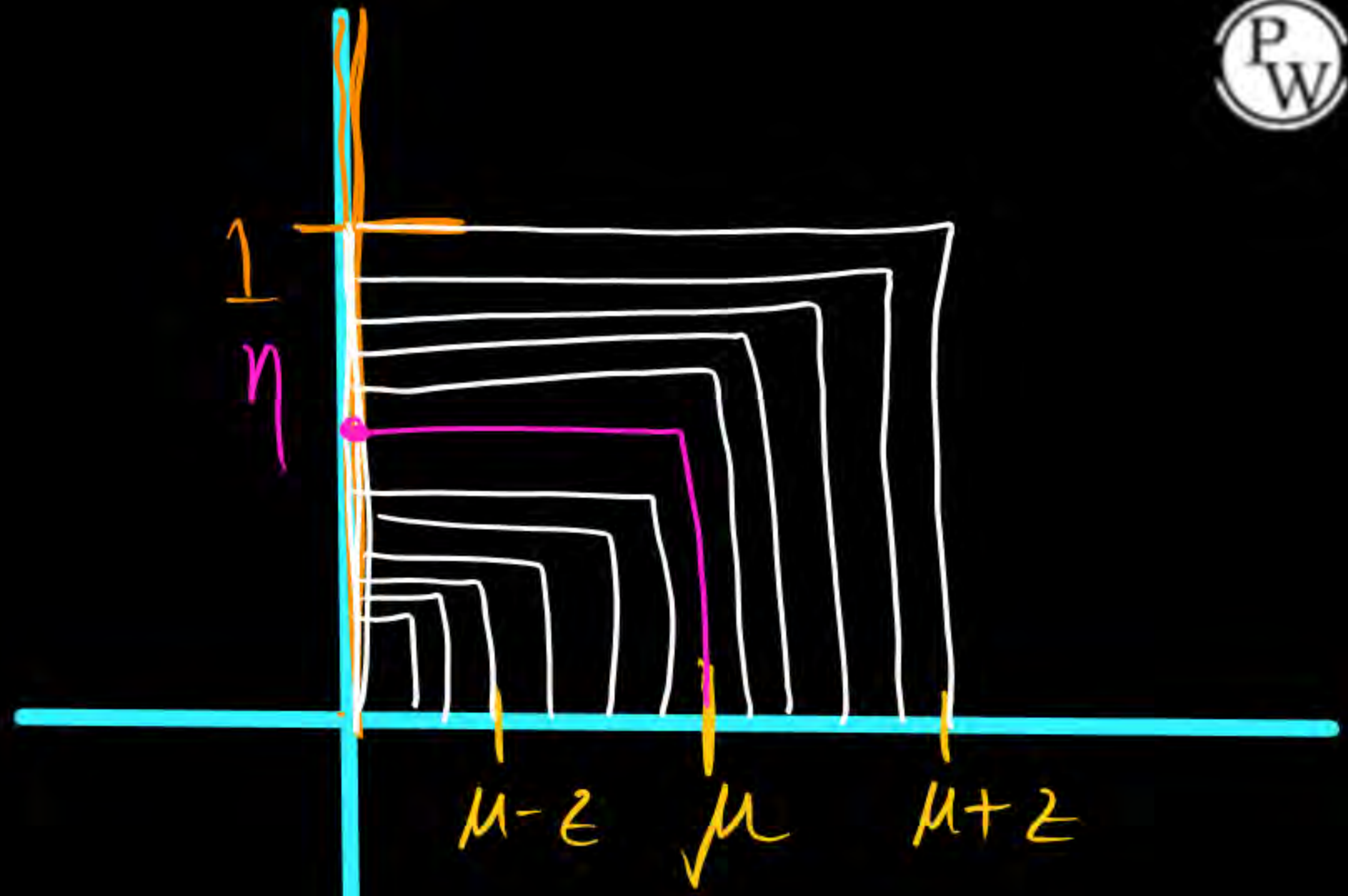
Left $= a-\varepsilon$

for R^2



Limiting concept





$$|X-\mu| < \epsilon$$

$$-\epsilon < X-\mu < \epsilon$$

$$\mu-\epsilon < X < \mu+\epsilon$$

$$P[|X-\mu| \leq \epsilon] \geq 1-\eta$$

ϵ = given μ = given

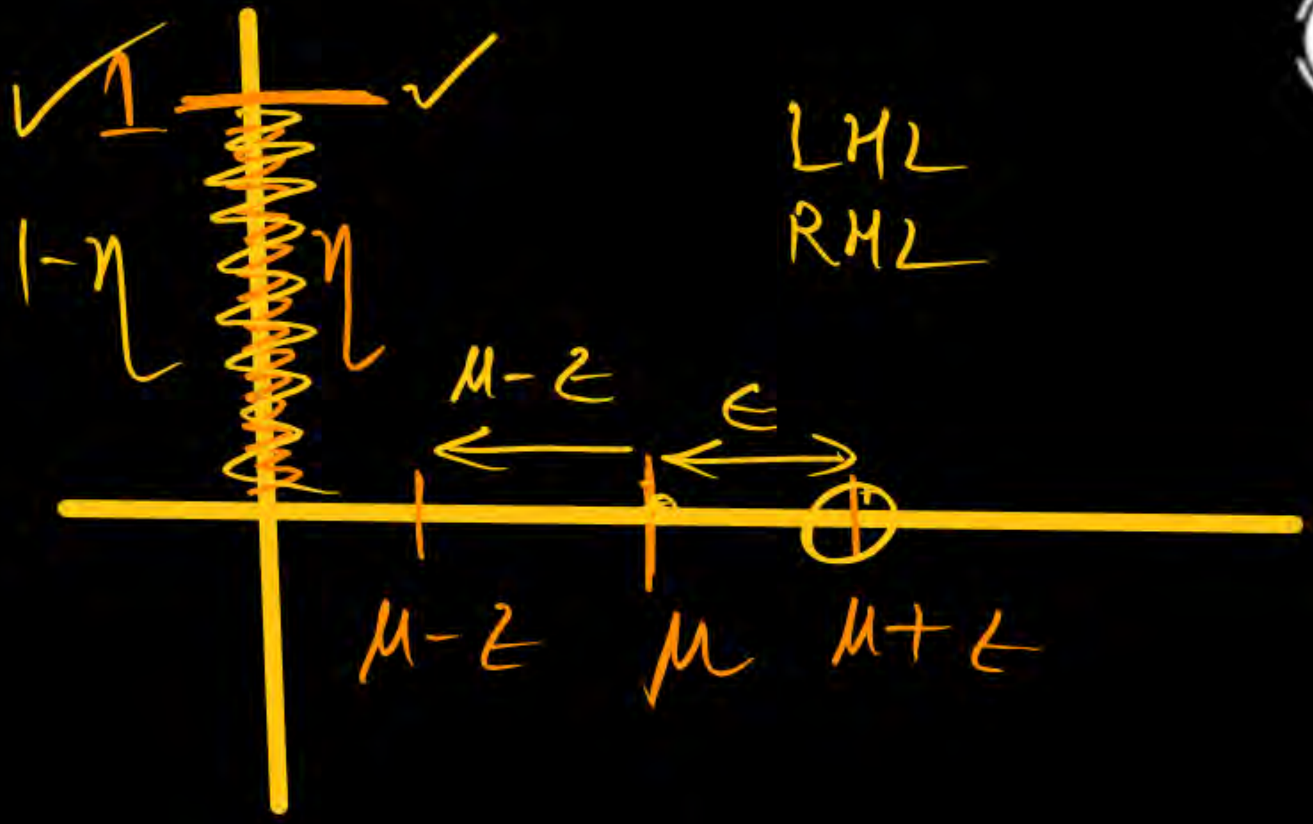
η = prob.

$|x| < a$

$-a \le x \le a$

$|X - \mu| < \epsilon$

$|\text{Random var} - \text{mean}| < \epsilon$



LHL
RHL

for SAMPLE mean

$|\bar{X} - \mu| < \epsilon$

$$\boxed{\mu - \epsilon \le \bar{X} \le \mu + \epsilon}$$

$$\boxed{P\left[|\bar{X} - \mu| < \epsilon\right] \ge 1 - \eta} \longrightarrow \text{Chebsher inequality}$$

$\left.\begin{matrix}\epsilon \\ \eta\end{matrix}\right]$ Two arbitrary

μ = mean

$\bar{X}$ = SAMPLE mean

$$P\left[|\bar{X}-\mu| < \epsilon\right] \geq 1-\eta$$

No. of Trials Are required $n \geq \dfrac{\sigma^2}{\epsilon^2 \eta}$

Chebshev Inequality

Law of Large No.

required trials

No. of trials $\geq \dfrac{(\text{standard})^2}{\epsilon^2 \cdot \eta}$

$$P\left[|\bar{X}-\mu| < \epsilon\right] \geq 1-\eta$$

$1-\eta = 0.99$

Q8. A pathologist wants to estimate the mean time required to complete a certain analysis on the basis of sample study so that he may be 99% confident that the mean time may remain with ±2 days of the mean. As per the available records, the population variance is 5 days^2. What must be the size of the sample for this study?

$1-\eta = 0.99$

$\epsilon = \text{mean} = 2$

$\sigma = 5$ Days

No. of sample size required $n \geq \dfrac{\sigma^2}{\epsilon^2 \eta}$

$P[|\bar{x}-\mu| \leq \epsilon] \geq 1-\eta$

mean

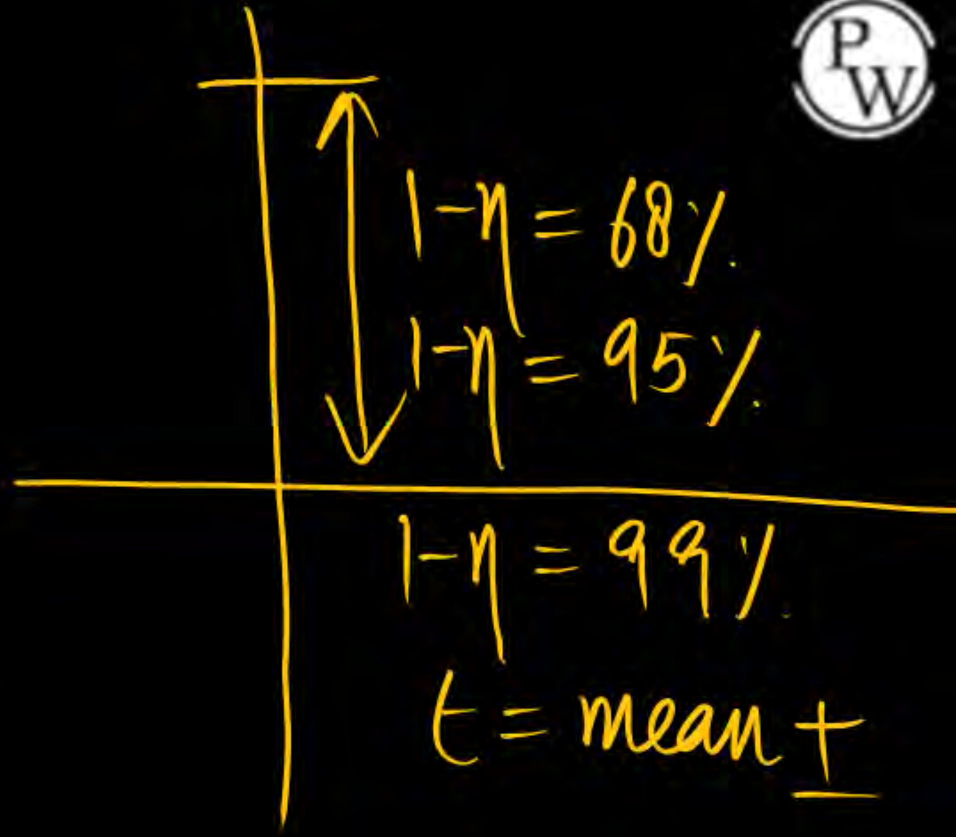
prob. $\eta = 1-0.99 = 0.01$

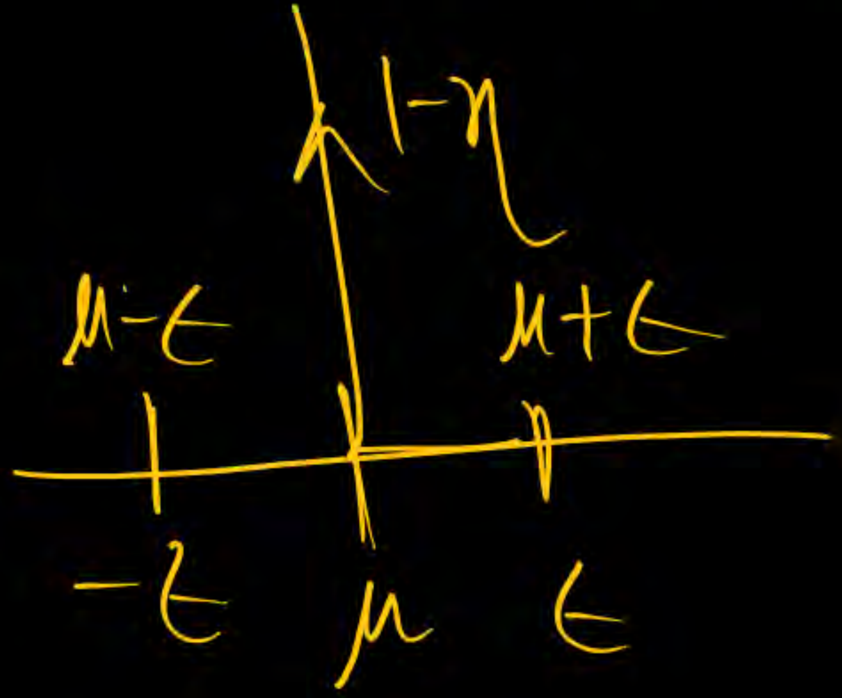
$$n > \frac{\sigma^2}{\epsilon^2 \eta} = n > \frac{(5)}{(2)^2 \times 0.01}$$

number of trials $\geq \frac{5}{4 \times 0.01}$ =

No. of trials are required $\geq \frac{5}{0.04}$

≥ 125 trials





Introduction to Sampling Distribution

Q9. An investigator wishes to estimate the mean of a population using a sample large enough that the probability will be 0.95 that the sample mean will not differ from the population mean by more than 25 percent of the standard deviation. How large a sample should be taken?

$1-\eta = 0.95$

$\eta = 1 - 0.95$

$\eta = 0.05$

$\epsilon = 0.25\sigma$

Using Large of trials

$n \geq \dfrac{\sigma^2}{\epsilon^2 \cdot \eta}$ $\quad n \geq \dfrac{\sigma^2}{(0.25\sigma)^2 \cdot 0.05} = n \geq 320$

Sample

Introduction to Sampling Distribution

[TEST – Distribution (Discrete)
15 questions

Q10. The mean of a population is unknown and having a variance equal to 2. Find out that how large a sample must be taken, so that the probability will be at least 0.95 that the sample mean will lie within the range of 0.5 of the population mean?

variance $\sigma^2 = 2$

$1 - \eta = 0.95$

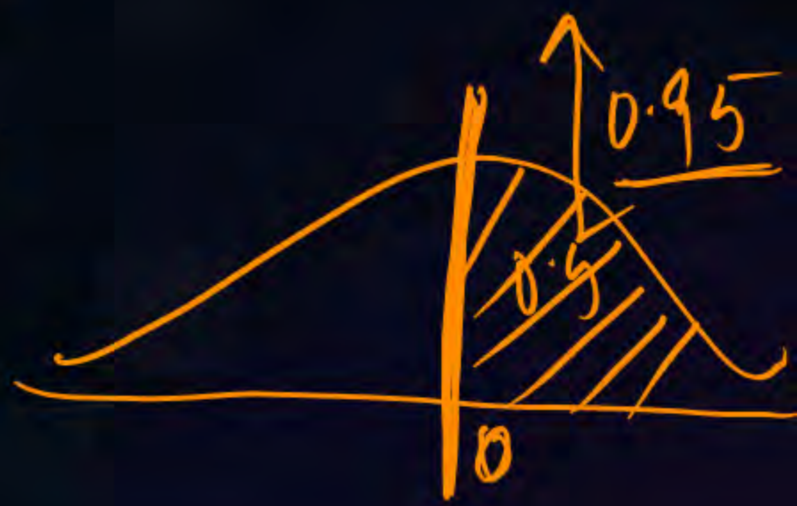
$\eta = 0.05$

$\epsilon = 0.5$

No. of trials Are required

$$n \geq \frac{\sigma^2}{\epsilon^2 \eta} \Rightarrow n \geq \frac{2}{(0.5)^2 \times 0.05}$$

$$\Rightarrow \underline{n \geq 160 \text{ SAMPLE}}$$



CHI-Square Distribution :

Gamma Distribution → (λ, α)

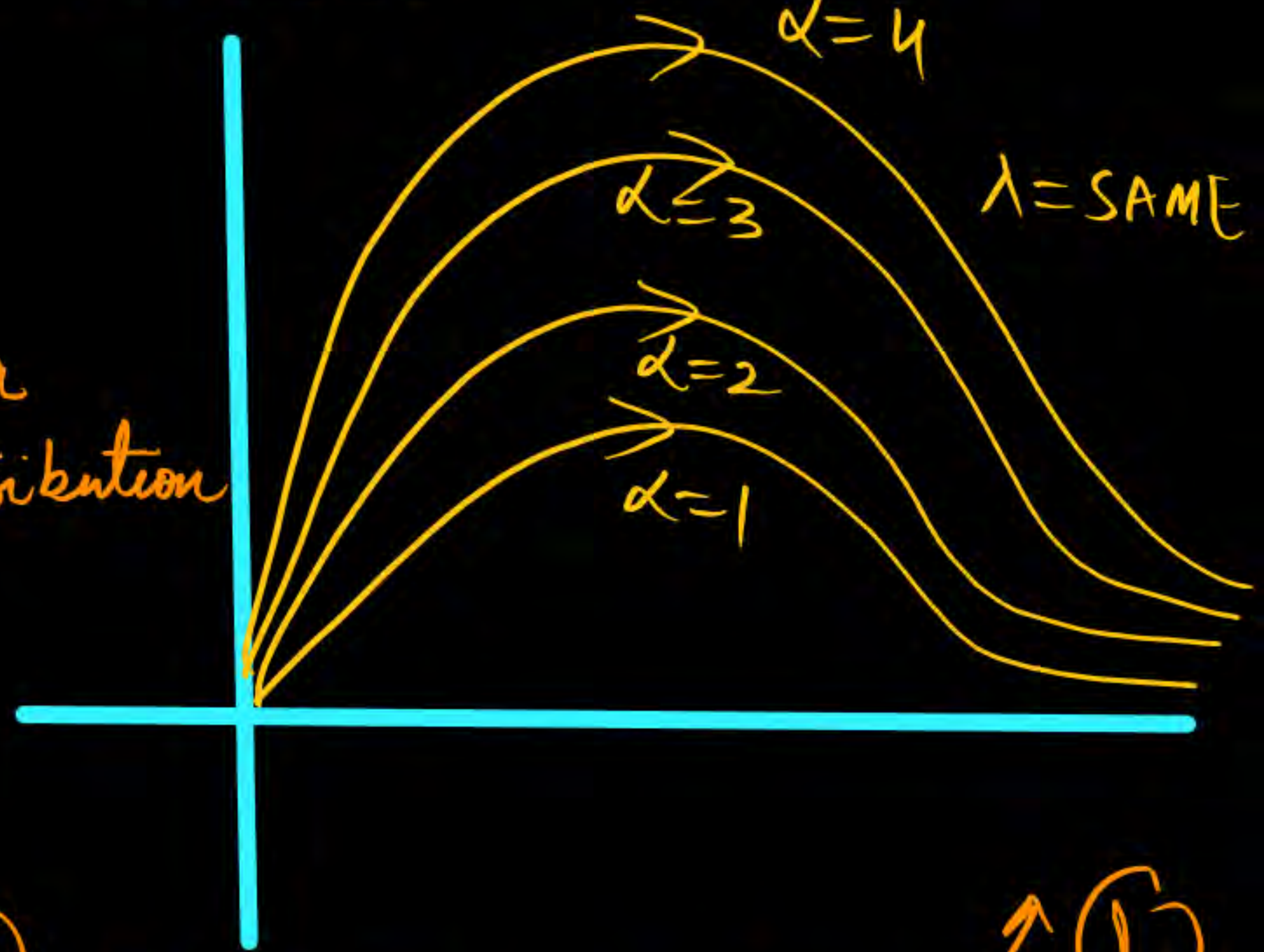
α = claim sizes

λ = Parameter

$$\boxed{f(x) = \frac{\lambda^{\alpha}}{\Gamma\alpha} e^{-\lambda x} x^{\alpha-1}}$$

$\lambda > 0$

$\alpha > 0$



→ Using cumulative Prob. of gamma Distribution

→ Chi square Distribution

Degree of Freedom

1, 2, 3, 4, 5, 6 → 3.5

70 student → mean marks 45

50 mean

$E[X] = 45$ — Condition → 1

DEGREE of freedom = 69 →

If sample size $= n$ Degree of freedom $\nu = (n-1)$

69 Dependent variable + 1 variable (Independent)

Gamma distribution

$$f(x) = \frac{\lambda^{\alpha}}{\Gamma\alpha} e^{-\lambda x} x^{\alpha-1}$$

$$\boxed{\begin{array}{ll} \alpha = \nu/2 \quad \lambda = \frac{1}{2} & \text{Chi square} \\ \alpha = \frac{(n-1)}{2} \quad \lambda = \frac{1}{2} & \text{Distribution} \end{array}}$$

Chi square Distribution

$$\boxed{f(x) = \frac{\left(\frac{1}{2}\right)^{\nu/2}}{\Gamma\frac{\nu}{2}} e^{-\frac{x}{2}} x^{\frac{\nu}{2}-1}}$$

Sample size 1

$\chi^2_{(1)}$

$$f(x) = \frac{\lambda^{\alpha}}{\Gamma\alpha} e^{-\lambda x} x^{\alpha-1}$$

$$f(x) = \frac{\left(\frac{1}{2}\right)^{\frac{\nu}{2}}}{\Gamma\frac{\nu}{2}} e^{-\frac{x}{2}} x^{\frac{\nu}{2}-1}$$

$\alpha = \nu/2 \qquad \lambda = \frac{1}{2}$

ν = degree of Freedom

Mean $= \frac{\alpha}{\lambda}$

Var $= \frac{\alpha}{\lambda^2}$

$\chi^2_{(1)}$ MEAN $= \frac{\nu/2}{\frac{1}{2}} = \nu$ degree of Freedom $\boxed{\text{mean}}$

$\chi^2_{(1)}$ variance $= \frac{\nu/2}{\left(\frac{1}{2}\right)^2} = 2\nu$

Different forms Chi Square Distribution

$N(\mu, \sigma^2)$

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

$\chi^2_{(1)} = Z^2$

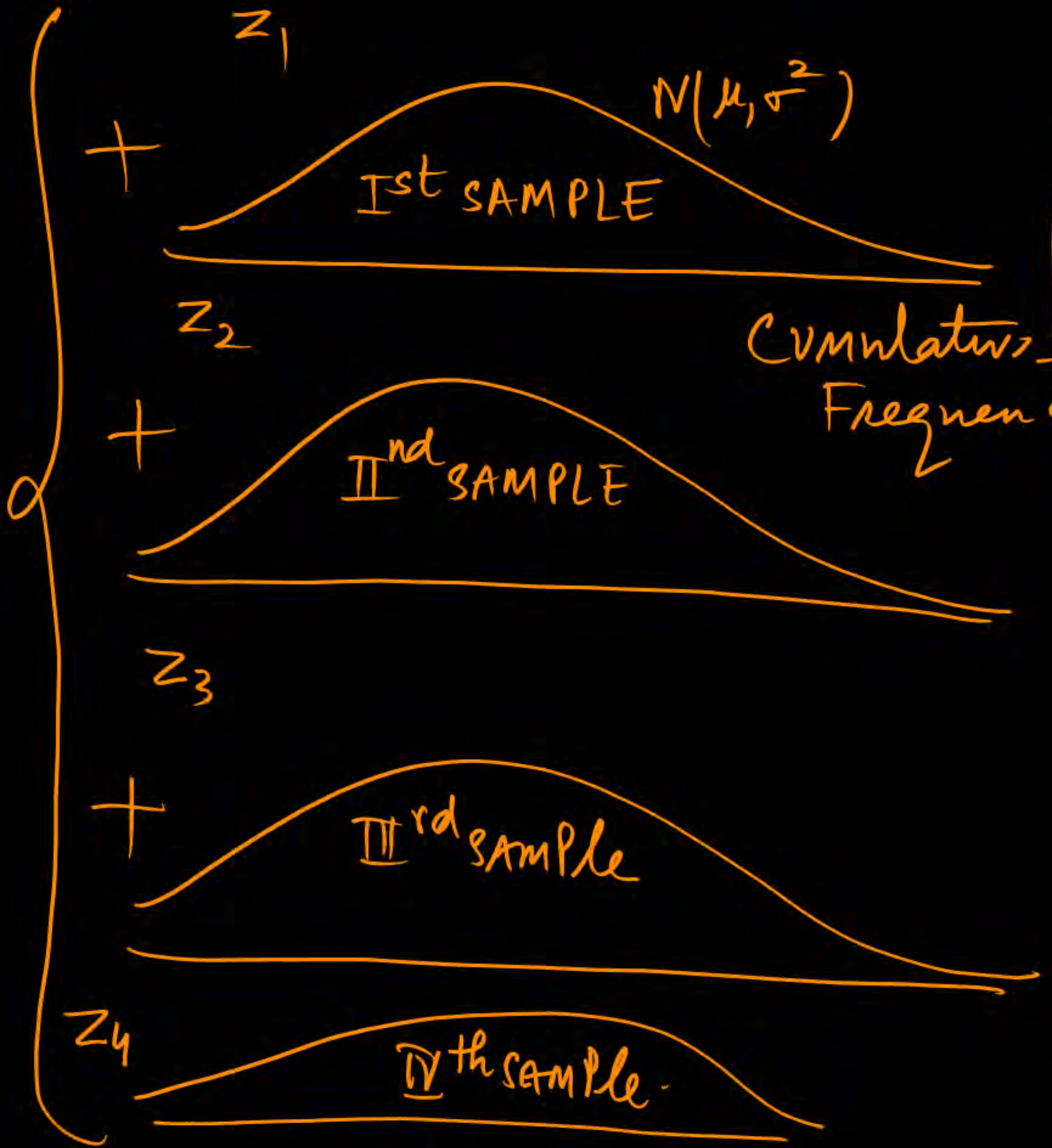
$\chi^2_{(1)} = \sum\left[\frac{X-\mu}{\sigma}\right]^2$

$\chi^2_{(1)} = \sum\left[\frac{X-\mu}{\sigma}\right]^2$

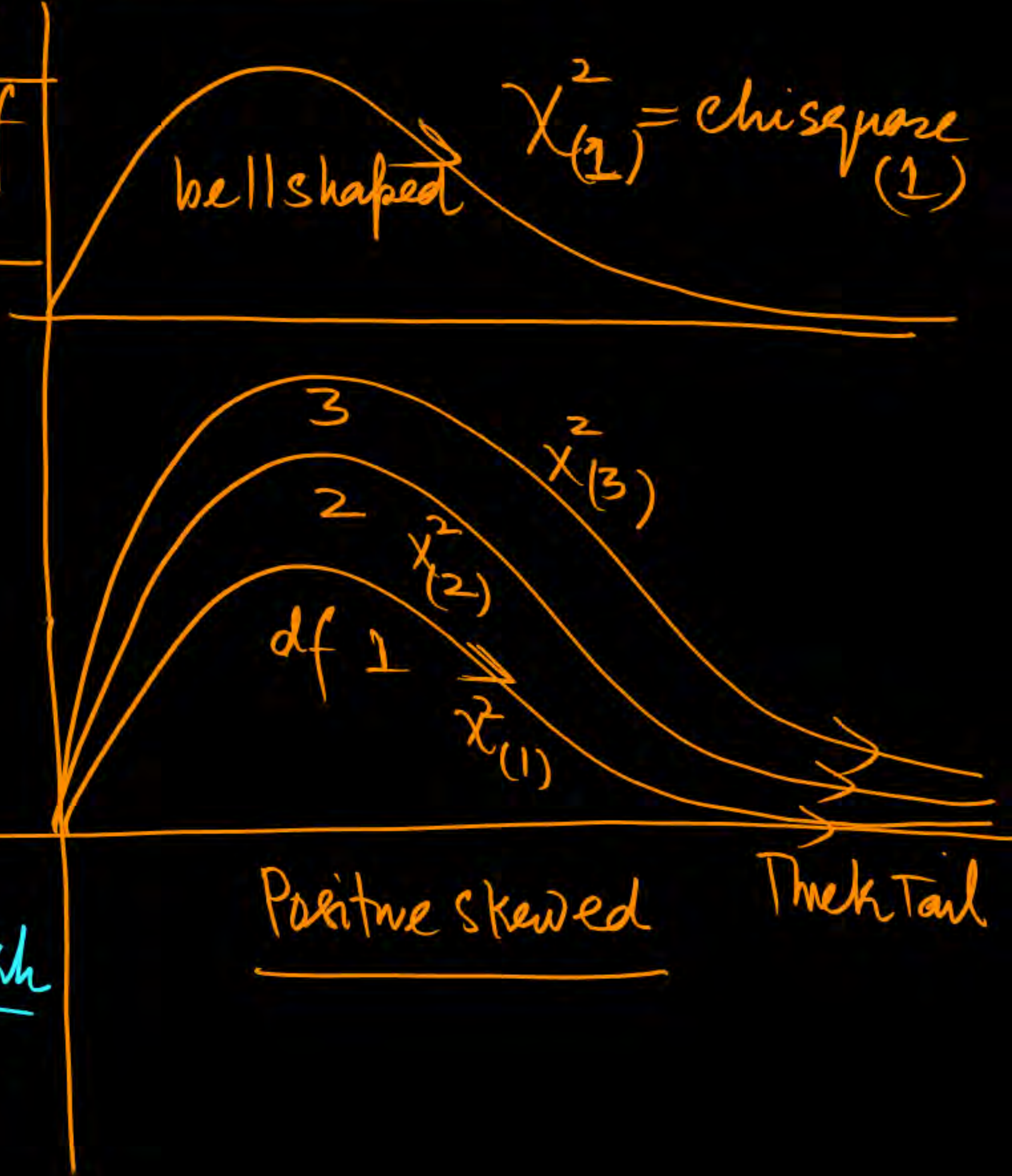
$$\chi^2_{(1)} = \left(z_1^2 + z_2^2 + z_3^2 + z_4^2\right)$$

$$\chi^2_{(1)} = \sum_{i=1}^{n} z_i^2$$

$\boxed{\begin{array}{l} \text{Mean} = df \\ \text{vari} = 2df \end{array}}$

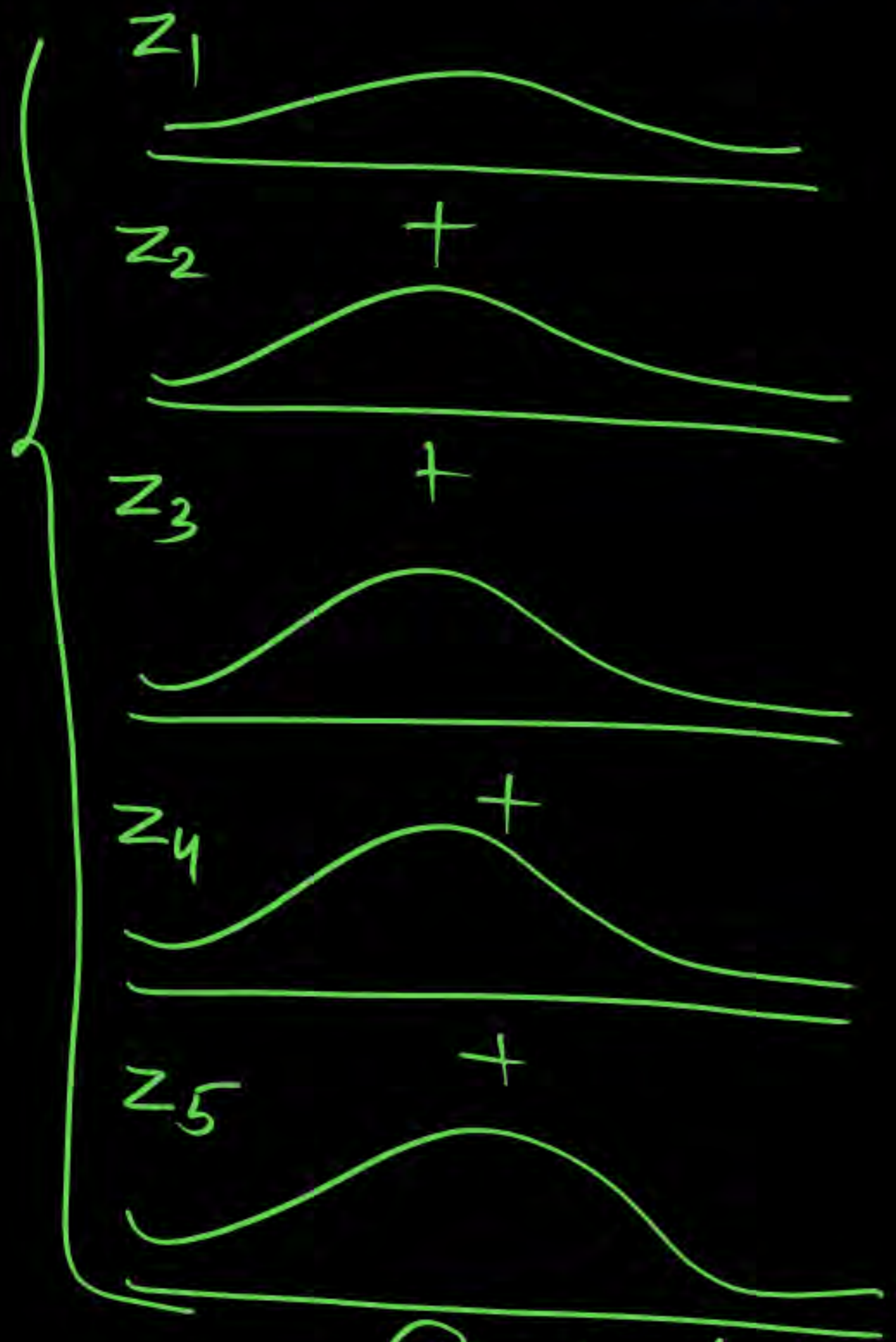


Cumulative Frequency

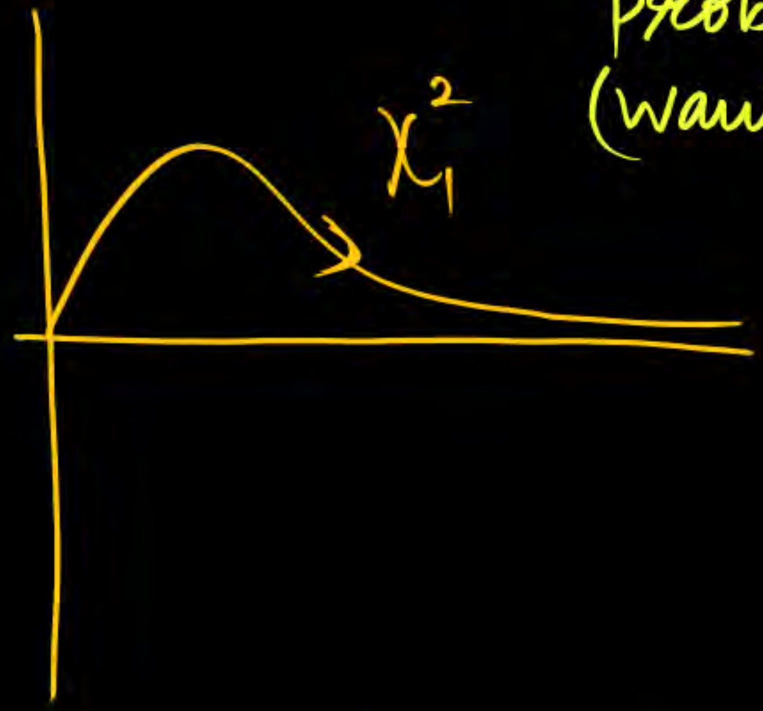


Shahrukh
Khan

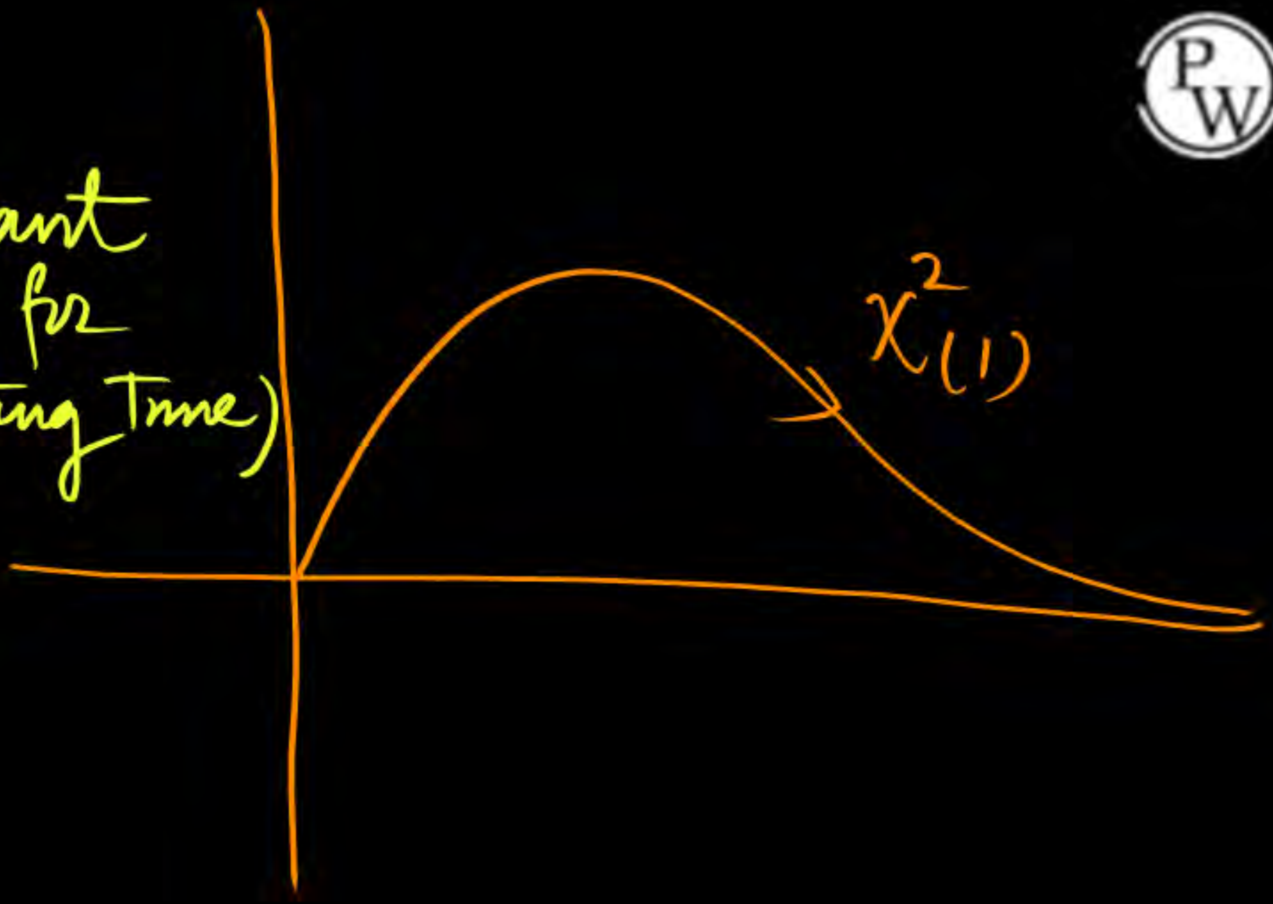
$$\chi^2_{(1)} = \sum_{i=1}^{n} z_i^2$$



=



cumulant prob for (waiting Time)



\# degree of freedom = 10

mean = degree of freedom = $\nu = 10$

variance = $2\nu = 20$

S.D = $\sqrt{20}$

\# Df = $(n-1)$

mean = $(n-1)$

variance = $2(n-1)$

S.D = $\sqrt{2(n-1)}$

α^2 = No. of claims in large Population size

THANK - YOU