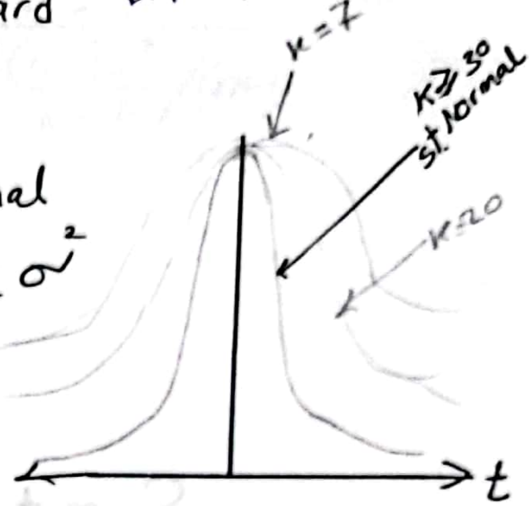
Lec 4

The t- distribution

Similar to normal with its bell shape but has heavier tails

t- distr. $\xrightarrow{n \geq 30}$ standard normal



If $X_1, X_2, \ldots X_n$ is a R.S. from normal population with mean μ, variance σ^2

$\overline{X} \sim N(\mu, \sigma^2)$

$\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$

$T \sim t_{\alpha, k} \Rightarrow E(T) = 0$

Most of time σ^2 is unknown

S^2 is a good estimator for σ^2 $(\sigma^2 \to S^2)$

$$Z = \frac{\overline{X} - \mu}{\frac{S}{\sqrt{n}}} \sim$$

if $n \geq 30$ $\rightarrow$ $Z = \frac{\overline{X} - \mu}{S/\sqrt{n}} \sim N(0,1)$

$n < 30$, σ is unknown $\rightarrow$ $T = \frac{\overline{X} - \mu}{S/\sqrt{n}}$ no longer $N(0,1)$ $\sim t_{n-1}$

Let Z, U are two indep R.V.s

$Z \sim N(0,1)$, $U \sim \chi^2_r$

$$T = \frac{Z}{\sqrt{\frac{U}{r}}} \sim t_r$$

<u>Since</u> $Z = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$

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

Z, U are indep.

Hence $T = \frac{Z}{\sqrt{\frac{U}{r}}} = \frac{\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}}{\sqrt{\frac{\cancel{(n-1)}S^2}{\sigma^2}/\cancel{(n-1)}}} = \frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}}$

$$\therefore \boxed{\frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}} \sim t_{n-1}}$$

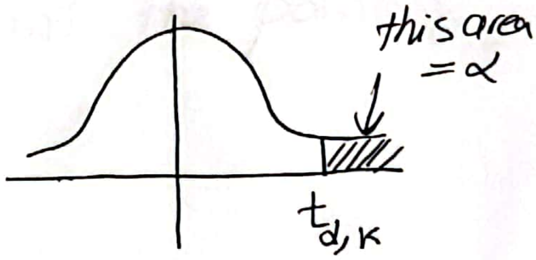
When we use t distribution?

When $n < 30$ and σ^2 is Unknown

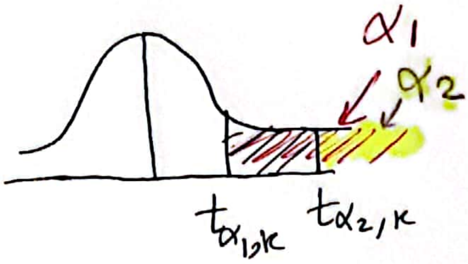
Notes on t-distr. Curve

$\Rightarrow$ Symetric Curve with heavier tails

$Pr(T > t_{\alpha,k}) = \alpha$

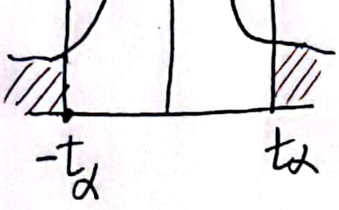


$Pr(t_{\alpha_1,k} < T < t_{\alpha_2,k}) = \alpha_1 - \alpha_2$



$\underset{CDF}{F}(t_{\alpha,k}) = 1 - \alpha$

-ve points of t



$F(-t_\alpha) = 1 - F(t_\alpha) = 1 - (1-\alpha) = \alpha$

To get $t_{0.025,5}$ $\Rightarrow$ $\underset{CDF}{F}(t_{0.025,5}) = 1 - 0.025 = 0.975$

$\therefore t_{0.025,5} = \underset{\text{Inv Cumulative}}{F^{-1}}(0.975)$

Scanned with CamScanner

using **Minitab**

Calc/ prob. distr. /t

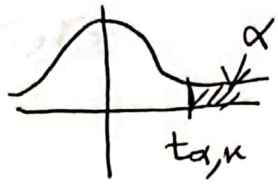
- Inverse Cum distr

degree of free = 5 , input Constant = 0.975 ↵

$t_{0.025,5} = 2.57$

two cases here

① given $t_{\alpha,k}$ (t-value) point, want area



use Cumulative fn $F(t_{\alpha,k}) = 1-\alpha$

$\therefore \underset{\text{area}}{\alpha} = 1 - \underset{\text{CDF}}{F(t_{\alpha,k})}$

$Pr(T > t_{\alpha,k}) = 1 - P(T < t_{\alpha,k})$

② given area α and you want the point $t_{\alpha,k}$

$F(t_{\alpha,k}) = 1-\alpha$

$\underset{\text{t-value}}{t_{\alpha,k}} = \underset{\text{inv Cumulative}}{F^{-1}}(1-\alpha)$

The F- distribution

Let U_1, U_2 are two indep. R.V's

$U_1 \sim \chi^2_{r_1}$ $\quad U_2 \sim \chi^2_{r_2}$

then $F = \dfrac{U_1/r_1}{U_2/r_2} \sim F_{r_1, r_2}$

Let $X_1, X_2, \cdots, X_{n_1}$ be indep random samples
$Y_1, Y_2, \cdots, Y_{n_2}$

from Population with $X_i \sim N(\mu_1, \sigma_1^2)$
$Y_i \sim N(\mu_2, \sigma_2^2)$

if $U_1 = \dfrac{(n_1-1)S_1^2}{\sigma_1^2} \sim \chi^2_{n_1-1}$

$U_2 = \dfrac{(n_2-1)S_2^2}{\sigma_2^2} \sim \chi^2_{n_2-1}$

Hence $F = \dfrac{U_1/n_1-1}{U_2/n_2-1} = \dfrac{\dfrac{(n_1-1)S_1^2}{\sigma_1^2(n-1)}}{\dfrac{(n_2-1)S_2^2}{\sigma_2^2(n_2-1)}} =$

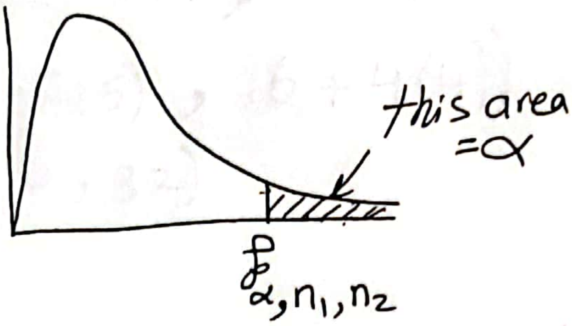
$$= \frac{S_1^2\,\sigma_2^2}{S_2^2\,\sigma_1^2} \sim F_{n_1-1, n_2-1}$$

F distr. is used with two sample variances

F-distr Curve

$\Rightarrow$ Not symmetric

$\Rightarrow$ alway +ve



$$\boxed{\underset{CDF}{F}\left(f_{\alpha,n_1,n_2}\right) = 1-\alpha}$$

Note $\quad f_{1-\alpha,n_1,n_2} = \dfrac{1}{f_{\alpha,n_2,n_1}}$

e.g. $\quad f_{0.95,6,10} = \dfrac{1}{f_{0.05,10,6}} = \dfrac{1}{4.06} = 0.246$

Ex① Consider the four indep R.V.^{s}

X, Y, U, V such that

$X \sim N(0,16) \Rightarrow \dfrac{X-0}{4} \sim N(0,1)$

$Y \sim N(5,4) \Rightarrow \dfrac{Y-5}{4} \sim N(0,1)$

$U \sim \chi^2_4$

$V \sim \chi^2_{16}$

State the distribution of the following Variables

① $X+2Y \sim N\left(E(X+2Y),\ V(X+2Y)\right)$

$N\left(E(X)+2E(Y),\ V(X)+4V(Y)\right)$

$N(10,\ 32)$

$\boxed{2}$ $X - 2Y \sim N\left(E(X-2Y), \operatorname{var}(X-2Y)\right)$

$$\sim N\left(0 - 2(5), 16 + 4(4)\right)$$

$$\sim N(-10, 32)$$

$\boxed{3}$ $\frac{X^2}{16} + \frac{(Y-5)^2}{4} \sim \chi^2_2$

$$\frac{X-0}{4} \sim N(0,1) \quad , \quad \frac{Y-5}{4} \sim N(0,1)$$

$\boxed{4}$ $\frac{X}{\sqrt{V}} \Rightarrow \frac{\frac{X}{\cancel{4}}}{\sqrt{\frac{V}{\cancel{16}}}}$ $\frac{\text{st. normal}}{\sqrt{\chi^2_{16}/16}}$

$$= \frac{X}{\sqrt{V}} \sim t_{16}$$

$\boxed{5}$ $\frac{4U}{V} \Rightarrow \begin{matrix} U \sim \chi^2_4 \\ V \sim \chi^2_{16} \end{matrix} \Rightarrow \frac{U/4}{V/16} \sim F(4,16)$

$$\Rightarrow 4\,\frac{U}{V} \sim F(4,16)$$

Ch. 1 Exercises

(6) If $X_1, X_2 \cdots X_n$ are i.i.d $N(0,\sigma^2)$

State the distribution of each of the following variables

(a) $U = 3X_1 - 5X_2 + 8$

$$U \sim N\left(E(U), V(U)\right)$$

$$E(3X_1 - 5X_2 + 8) = 3E(X_1) - 5E(X_2) + 8$$
$$= 0 - 0 + 8 = 8$$

$$V(3X_1 - 5X_2 + 8) = 9V(X_1) + 25V(X_2)$$
$$= 9\sigma^2 + 25\sigma^2 = 34\sigma^2$$

$$U \sim N(8, 34\sigma^2)$$

(b) $V = \sum_{i=1}^{n} X_i \sim N(0, n\sigma^2)$

(c) $W = \dfrac{\left(\sum_{i=1}^{n} X_i\right)^2}{n\sigma^2}$

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} \sim N\left(0, \frac{\sigma^2}{n}\right)$$

$$\frac{\bar{X} - 0}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

$$\frac{\frac{\sum X_i}{n}}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1) \Rightarrow \frac{\sum X_i}{\sqrt{n}\,\sigma} \sim N(0,1)$$

$$\left(\frac{\Sigma x_i}{\sqrt{n}\,\sigma}\right)^2 \sim \chi^2_1$$

$$\frac{(\Sigma x_i)^2}{n\sigma^2} \sim \chi^2_1$$

(d) $Y = \dfrac{2X_1^2}{X_2^2 + X_3^2}$

$$X_1 \sim N(0, \sigma^2)$$

$$\frac{X_1}{\sigma} \sim N(0,1)$$

$$\left(\frac{X_1}{\sigma}\right)^2 \sim \chi^2_1$$

$$\frac{X_1^2}{\sigma^2} \sim \chi^2_2 \;,\; \frac{X_2^2}{\sigma^2} + \frac{X_3^2}{\sigma^2} \sim \chi^2_2$$

$$Y = \frac{\frac{X_1^2}{\cancel{\sigma^2}}}{\frac{X_2^2 + X_3^2}{\cancel{\sigma^2}}/2} = \frac{2X_1^2}{X_2^2 + X_3^2} = F(1, 2)$$

(e) $Y = \dfrac{\Sigma X_i}{\sqrt{\Sigma X_i^2}}$

$$X_i \sim N(0, \sigma^2)$$

$$\frac{X_i}{\sigma} \sim N(0,1)$$

$$\Sigma X_i \sim N(0, n\sigma^2)$$

$$\frac{\Sigma X_i}{\sqrt{n}\,\sigma} \sim N(0,1)$$

$$\sum_{i=1}^{n}\left(\frac{X_i}{\sigma}\right)^2 \sim \chi^2_n$$

$$\frac{\frac{\Sigma X_i}{\sqrt{n}\,\sigma}}{\sqrt{\Sigma \frac{X_i^2}{\sigma^2}/n}} \sim t_n$$

Summary

$\Rightarrow$ if $X_1, X_2, \ldots X_n$ i.i.d $N(\mu, \sigma^2)$

① $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$ is a R.v. $\sim N(\mu, \frac{\sigma^2}{n})$

$$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

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

③ $Z_i = \frac{X_i-\mu}{\sigma} \sim N(0,1)$

④ $\sum_{i=1}^{n} x_i \sim N(n\mu, n\sigma^2)$

⑤ $U = \sum_{i=1}^{n} (Z_i)^2 \sim \chi^2_n$

$\Rightarrow$ If $U_i \sim \chi^2_{n_i}$ $\therefore \sum_{i=1}^{k} U_i \sim \chi^2_{\sum_{i=1}^{k} n_i}$

$\Rightarrow$ If $T = \frac{Z}{\sqrt{\frac{U}{n}}} \sim t_n$

$\frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}} \sim t_{n-1}$ if $n<30$, σ is unknown

$\Rightarrow$ If $F = \frac{U_1/n_1}{U_2/n_2} \sim F(n_1, n_2)$

Chapter two

Point Estimation

statistical inference

To infer is to make some conclusion or evaluation based on information that is not really complete

Statistical inference is the Process by which information from sample data is used to draw conclusion about the population from which the sample was selected

Parameter estimation : to estimate population parameters

Hypothesis testing

Point estimation Ch.2

$\hat{\mu} = \bar{X}$

are never perfect
they always have an error

Interval estimation (Confidence interval) Ch.3

$L < \mu < U$ with Confidence 95%

let $X_1, X_2, \dots, X_n$ random Sample i.i.d (indep. identically distri)

$x_1, x_2, \dots, x_n$ observed data

$T = t(X_1, X_2, \dots X_n)$

a function of random Sample that doesnot depend on any unknown Parameter is called statistic e.g $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ Called Statistic

$S^2 = \frac{1}{n-1} \sum_{i=1}^{n}(X_i-\overline{X})^2$ " "

A statistic is also a random variable has a sampling distribution

e.g. If $X_1, X_2, --- X_n$ R.S. $\sim N(\mu, \sigma^2)$

$\overline{X}$ is a statistic, R.V.

S^2 " " , R.V.

e.g In Binomial (n, p)

p: prob of success if unknown

we need to estimate $\hat{p}$ ← hat

e.g. In exponential dist. (λ)

if λ is unknow $\Rightarrow$ we need $\hat{\lambda}$

$\hat{\lambda}$ Called estimator of λ

(R.V.)

$\hat{\theta} = t(X_1, X_2, --- X_n)$ is the estimator of θ

estimate $\Rightarrow$ value

e.g. In Chi-Squ $\chi^2_{\nu} \Rightarrow \hat{\nu}$ estimator of ν

$\therefore X_1, X_2, ... X_n$ is R.S from Population

PdF/PMF $f(x, \theta)$

$T = t(X_1, X_2, --- X_n)$ statistic

$\hat{\theta} = t(X_1, X_2, --- X_n)$ estimator of θ

two question :-

1. How to estimate
   How to get point estimate < Method of Moment / Method of Max Likelihood MLE

2. Properties of estimators

(Mean-squared-error)

unbiased estimators | Minimum variance unbiased Estimator MVUE | efficiency | Consistency

Methods of Point Estimation

$\boxed{1}$ Method of Moment

If the parameter only is unknown

$X_1, X_2, X_n, \dots X_n$ R.S. i.i.d from a given Population distr.

Population	Sample
$\mu_1 = E(X)$ 1st moment	$m_1 = \frac{1}{n}\sum_{i=1}^{n} X_i$ 1st. sample m
$\mu_2 = E(X^2)$ 2nd moment	$m_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$ 2nd S.moment
$\vdots$	$\vdots$
$\mu_K = E(X^K)$ kth moment	$m_K = \frac{1}{n}\sum_{i=1}^{n} X_i^K$ kth S.moment
computed by pdf/PMF $< \int / \sum$ in terms of population parameter	$m_1, m_2, \dots m_K$ are numbers

by equating sample moments and Population moment

$$\mu_K = m_K$$

will get estimates of estimators

P. moment $\mu_K = E(X^K)$ S. moment $m_K = \frac{\sum X_i^K}{n}$

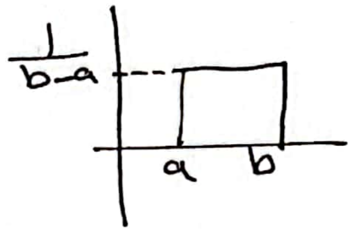
Ex ① let X be uniformly distributed on the interval $(\alpha, 1)$. Given a random sample of size n, use the method of moment to obtain a formula for estimating the Parameter α

Re Call that

for uniform (a, b)

$$f(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{o.w.} \end{cases}$$

pdf



$\Rightarrow E(X) = \frac{a+b}{2}$

then Uniform $(\alpha, 1)$

$$f(x) = \begin{cases} \frac{1}{1-\alpha} & 1 \leq x \leq \alpha \\ 0 & \text{o.w.} \end{cases}$$

get μ_1 = $E(x)$

1st population moment

from pdf

$$\mu_1 = \int_{\alpha}^{1} x \, f(x) \, dx$$

$$\mu_1 = \frac{1}{1-\alpha} \int_{\alpha}^{1} x \, dx = \frac{1}{1-\alpha} \left[\frac{x^2}{2}\right]_{\alpha}^{1}$$

$$\mu_1 = \frac{1-\alpha^2}{2(1-\alpha)}$$

$$\mu_1 = \frac{1+\alpha}{2}$$

$$m_1 = \frac{1}{n} \sum_{i=1}^{n} X_i = \overline{X}$$

1st sample moment

by equating $m_1 = \mu_1$

$$\frac{1+\alpha}{2} = \overline{X} \Rightarrow \boxed{\hat{\alpha} = 2\overline{X} - 1}$$

EX (2) Given a random sample of size n from Poisson Population, use the method of moment to obtain a formula for estimating the Parameter λ

Recall that

for Poisson(λ) $\quad$ X discrete R.V.

$$f(x) = \frac{e^{-\lambda} \lambda^x}{x!} \quad , x = 0, 1, 2 \cdots$$

μ_1 $= \lambda$ (from $\sum x f(x)$)

1st population moment $\quad$ interms of population parameters

$$m_1 = \frac{1}{n} \sum_{i=1}^{n} x_i = \bar{X}$$

1st sample moment

equating $\mu_1 = m_1$

$$\hat{\lambda} = \bar{X}$$

Ex (3) Given a random sample of size n from $N(\mu, \sigma^2)$ Population
Use the method of moment to obtain formula for estimating the parameters μ, σ^2

1^{st} population moment $= 1^{st}$ sample moment

$$\mu_1 = E(X) \quad , \quad m_1 = \frac{\sum X_i}{n}$$

$$\boxed{\hat{\mu} = \bar{X}}$$

2^{nd} p. moment $= 2^{nd}$ sample moment

$$m_2 = \frac{\sum X_i^2}{n}$$

$$\mu_2 = E(X^2)$$

$$\sigma^2 = E(X^2) - (E(X))^2$$

$$E(X^2) = \sigma^2 + \mu^2$$

$$\mu_2 = m_2$$

$$\hat{\sigma}^2 + \hat{\mu}^2 = \frac{\sum X_i^2}{n}$$

$$\hat{\sigma}^2 = \frac{\sum X_i^2}{n} - \bar{X}^2$$

$$\boxed{\hat{\sigma}^2 = \frac{1}{n} \sum (X_i - \bar{X})^2}$$

Note

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

<u>Why</u>

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

$$= \frac{1}{n} [\sum X_i^2 - 2\bar{X} \underbrace{\sum X_i}_{n\bar{X}} + n\bar{X}^2]$$

$$= \frac{1}{n} [\sum X_i^2 - 2n\bar{X}^2 + n\bar{X}^2] = \frac{1}{n} \sum X_i^2 - \bar{X}^2$$

Formula Sheet for
Probability and Statistical II ()

Chapter 1

If $X_1, X_2, \ldots, X_n$ constitute a R.S. from a $N(\mu, \sigma^2)$, then

(1) $Z = \dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$.

(2) $T = \dfrac{\overline{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$

(3) $U = \dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$

(4) $F = \dfrac{U_1/(n_1 - 1)}{U_2/(n_2 - 1)} = \dfrac{S_1^2\,\sigma_2^2}{S_2^2\,\sigma_1^2} \sim F_{n_1-1, n_2-1}$

Chapter 2

Point Estimation

(1) $\mu_k = E(X^k)$, $m_k = \dfrac{1}{n}\sum_{i=1}^{n} X_i^k$, $k = 1, 2, \ldots, r$

(2) $MSE(\hat{\theta}) = E(\hat{\theta} - \theta)^2$

Special Probability Distributions

A- Discrete Distributions

Name of Disn	*PMF*	*parameters*	*Mean*	*Variance*
Bernoulli	$f(x,p) = p^x q^{1-x}$, $x = 0, 1$	p	p	pq
Binomial	$f(x,n,p) = \binom{n}{x} p^x q^{n-x}$, $x = 0, 1, 2, \ldots, n$	n, p	np	npq
Poisson	$f(x;\lambda) = \dfrac{e^{-\lambda}\lambda^x}{x!}$ for $x = 0, 1, 2, \ldots$	λ	λ	λ
Geometric	$f(x;p) = p\,q^{x-1}$ for $x = 1, 2, 3, \ldots$	p	$1/p$	q/p^2
-ve Binomial	$f(x;k,p) = \binom{x-1}{k-1} p^k q^{x-k}$, $x = k, k+1, k+2, \ldots$	k, p	$\dfrac{k}{p}$	$\dfrac{kq}{p^2}$

B- Discrete Distributions

Name of Disn	*PDF*	*parameters*	*Mean*	*Variance*
Uniform	$f(x) = \dfrac{1}{\beta - \alpha}$ for $\alpha < x < \beta$	α, β	$(\alpha+\beta)/2$	$(\beta-\alpha)/12$
Exponential	$f(x;\theta) = \dfrac{1}{\theta} e^{-x/\theta}$ for $x > 0$	θ	θ	θ^2
Gamma	$f(x;\theta) = \dfrac{1}{\theta^n \Gamma(n)} x^{n-1} e^{-x/\theta}$ for $x > 0$ if n is +ve integer, $\Gamma(n) = (n-1)!$	θ	$n\theta$	$n\theta^2$
Normal	$f(x,\mu,\sigma) = \dfrac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$ for $-\infty < x < \infty$	μ, σ^2	μ	σ^2