PROBABILITY AND STATISTICS (UCS401)

Lecture-29

(Theory of Estimation and consistency with illustrations)

Sampling Distributions and Theory of Estimation (Unit –V & VI)

THAPAR INSTITUTE
OF ENGINEERING & TECHNOLOGY
(Deemed to be University)

Dr. Rajanish Rai
Assistant Professor
School of Mathematics
Thapar Institute of Engineering and Technology, Patiala

Student's t distribution

(i) Let $X \sim N(0,1)$, $Y \sim \chi^2_n$ $\xRightarrow{\text{indep}}$ $\dfrac{X}{\sqrt{Y/n}} \sim T_n$

T_n : T on 'n' degrees of freedom

(ii) p.d.f. of T:

$$f_T(t) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{2^{\frac{n+1}{2}}\sqrt{\pi n}\,\Gamma\left(n/2\right)\left(1+\frac{t^2}{n}\right)^{\frac{n+1}{2}}} \qquad -\infty < t < \infty$$

Observation : $f_T(t) = f_T(-t)$ $\Rightarrow$ symmetric about $t=0$

$\text{Med}(T) = 0$ and also $\mathbb{E}(T) = 0$

How to make T variable (Assuming Normal population to derive)

We need a standard Normal : $\dfrac{\bar{X}-\mu}{\sqrt{\sigma^2/n}} \sim N(0,1)$

a Chi-square : $\dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$

$\xRightarrow{\text{ind}}$ $\dfrac{\left(\dfrac{\bar{X}-\mu}{\sqrt{\sigma^2/n}}\right)}{\sqrt{\dfrac{(n-1)S^2}{\sigma^2(n-1)}}} = \dfrac{\bar{X}-\mu}{\sqrt{S^2/n}}$

Thus, $\boxed{\dfrac{\bar{X}-\mu}{\sqrt{S^2/n}} \sim t_{n-1}}$

F- distribution $F_{m,n}$

Let

$X \sim \chi^2_m$, $Y \sim \chi^2_n$ $\xrightarrow{\text{indep}}$ $\dfrac{X/m}{Y/n} \sim F_{m,n}$

How to make a F- variable

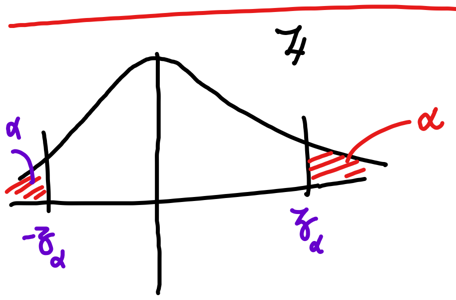
- $\dfrac{(n_1-1)S_1^2}{\sigma_1^2} \sim \chi^2_{n_1-1}$
- $\dfrac{(n_2-1)S_2^2}{\sigma_2^2} \sim \chi^2_{n_2-1}$

$\xRightarrow{\text{indep}}$ $\dfrac{\left(\dfrac{(n_1-1)S_1^2}{\sigma_1^2(n_1-1)}\right)}{\left(\dfrac{(n_2-1)S_2^2}{\sigma_2^2(n_2-1)}\right)} = \dfrac{\sigma_2^2 S_1^2}{\sigma_1^2 S_2^2} \sim F_{n_1-1,\, n_2-1}$

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

Assuming Normal population, then a General Rule.

Symm				
✓	Z	(i)	$\dfrac{\bar{X}-\mu}{\sqrt{\sigma^2/n}} \sim N(0,1)$	Used when some claim about μ is to be done when σ^2 is known. Based on sample
✗	χ^2	(ii)	$\dfrac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}$	Used when some claim about σ^2 is to be done when μ is unknown. Based on sample
✓	t	(iii)	$\dfrac{\bar{X}-\mu}{\sqrt{s^2/n}} \sim t_{n-1}$	used when some claim about μ is to be done when σ^2 is unknown. Based on sample
✗	F	(iv)	$\dfrac{\sigma_2^2 s_1^2}{\sigma_1^2 s_2^2} \sim F_{n_1-1,n_2-1}$	used when some claim about the ratio of two variances of populations is to be done. Based on sample.



Observations heavily used while doing statistical inference, testing of Hypothesis.

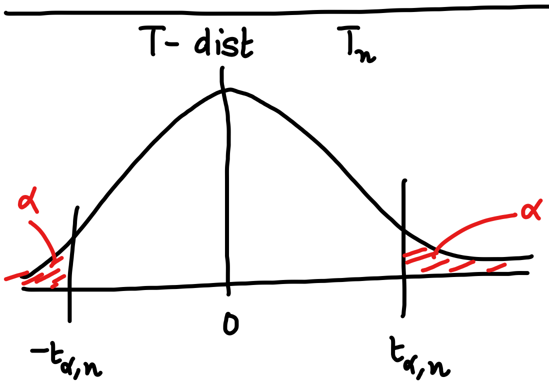
What is 'α' : $P(Z \geq z_\alpha) = \alpha$

and $P(Z \leq -z_\alpha) = \alpha$

$\Rightarrow P(Z \geq -z_\alpha) = 1-\alpha$

$P(Z \geq z_{1-\alpha}) = 1-\alpha$

$\Rightarrow \boxed{-z_\alpha = z_{1-\alpha}}$



$P(T \geq t_{\alpha,n}) = \alpha$

$P(T \leq -t_{\alpha,n}) = \alpha$

using the same argument as above.

$\boxed{-t_{\alpha,n} = t_{1-\alpha,n}}$

Z and t are symmetric about 0.

χ^2 and F are not symmetric distributions

Elementary Statistical inference

X is a R.V with the following Questions

(i) p.m.f. / p.d.f is completely unknown

(ii) The form / family of the pdf is known but the parameters are unknown.

In this estimation problem we will be dealing with second question i.e Parameter estimation. and in specific $\boxed{\text{POINT estimation}}$.

Ex: $X \sim Bin(n, p)$: estimate n, p

$X \sim Pois(\lambda)$: estimate λ.

$X \sim U(0, \theta)$: estimate θ.

$X \sim exp(\lambda)$: estimate λ.

$X \sim N(\mu, \sigma^2)$: estimate μ, σ^2.

Q. How do you make inferences/conclusions for a parameter of populati[on]

(A) By Analysing the sample from that population.

Q. What special properties should your sample STATISTIC have in order to have reasonable estimation/inferences.

A. First of all, there are two terms : ESTIMAND : which is to be estimated

ESTIMATOR : which will estimat[e] the estimand.

Q. So, the statistic should have some special properties in order to be a 'good' ESTIMATOR

What is 'good'.

(i) The first property is that the estimator should be UNBIASED

Unbiased Estimator

Def^n: A Statistic $T(X_1, X_2, X_3 \cdots X_n)$ is said to be unbiased Estimator of $f(\theta)$ if

$$\boxed{\mathbb{E}(T) = f(\theta)} \quad \forall \theta$$

if $\mathbb{E}(T) \neq f(\theta)$ then T is a biased estimator.

Ex $X \sim Bin(n, p)$, n is known, p is unknown

find an unbiased estimator.

Ans My claim: $T(X) = \frac{X}{n}$ is unbiased for p

Verify: $\mathbb{E}(T) = \mathbb{E}\left(\frac{X}{n}\right) = \frac{1}{n} \cdot np = \boxed{p}$

i.e 30 success in 50 trials then $\hat{p} = \frac{30}{50} = .6$ is unbiased estimator for p

Theorem: If $X_1 X_2 X_3 \cdots X_n$ be i.i.d from a distribution with mean μ and variance σ^2.

Then
$$\mathbb{E}(\bar{X}) = \mu$$
$\Rightarrow$ $\bar{X}$ is unbiased for μ.

is s^2 unbiased for σ^2.

Check: $\mathbb{E}(s^2) = \mathbb{E}\left(\frac{1}{n-1}\sum(X_i - \bar{X})^2\right) = \frac{1}{n-1}\left\{\mathbb{E}\left(\sum_{i=1}^{n}(X_i-\bar{X})^2\right)\right\}$

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

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

$$= \frac{1}{n-1}\left\{\mathbb{E}\left(\sum X_i^2 - n\bar{X}^2\right)\right\}$$

$$= \frac{1}{n-1}\left\{\sum \mathbb{E}(X_i^2) - n\,\mathbb{E}(\bar{X}^2)\right\}$$

$$= \frac{1}{n-1}\left\{\sum_{i=1}^{n} V(X_i) + (\mathbb{E}(X_i))^2 - n\left(\frac{\sigma^2}{n} + \mu^2\right)\right\}$$

$$= \frac{1}{n-1}\left\{\sum(\sigma^2 + \mu^2) - n\left(\frac{\sigma^2}{n} + \mu^2\right)\right\}$$

$$= \frac{1}{n-1}\left\{n(\sigma^2 + \mu^2) - \sigma^2 - n\mu^2\right\}$$

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

$\mathbb{E}(X_1^2 + X_2^2)$

$= \sum_{i=1}^{2} \mathbb{E}(X_i^2)$

$V(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$

$V(\bar{X}) = \mathbb{E}(\bar{X}^2) - (\mathbb{E}(\bar{X}))^2$

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

see previous lecture pdf.

Thus, s^2 is unbiased for σ^2

⊙ If it is kown that $X_1 X_2 \ldots X_n$ are i.i.d $\sim N(\mu, \sigma^2)$

then, we know that

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

$$\Rightarrow \mathbb{E}\left(\frac{(n-1)s^2}{\sigma^2}\right) = n-1$$

$$\Rightarrow \frac{n-1}{\sigma^2}\mathbb{E}(s^2) = n-1$$

$$\Rightarrow \boxed{\mathbb{E}(S^2) = \sigma^2}$$

Since we saw $\bar{X}$ is unbiased for μ

can we say $(\bar{x})^2$ unbiased for μ^2

See: $\mathbb{E}[(\bar{x})^2] = var(\bar{x}) + (\mathbb{E}(\bar{x}))^2$

$= \frac{\sigma^2}{n} + \mu^2 \neq \mu^2$

So in general it does not Hold that if

T is unbiased for θ ✗

then $f(T)$ would be unbiased for $f(\theta)$

for this case, we can see that

$$\mathbb{E}(\bar{X}^2) = \frac{\sigma^2}{n} + \mu^2 \Rightarrow \mathbb{E}(\bar{x}^2) - \frac{\sigma^2}{n} = \mu^2$$

$$\Rightarrow \mathbb{E}(\bar{x}^2) - \frac{\mathbb{E}(S^2)}{n} = \mu^2 \Rightarrow \mathbb{E}\left(\bar{X}^2 - \frac{S^2}{n}\right) = \mu^2$$

$\mathbb{E}(s^2) = \sigma^2$
Shown above
i.e. s^2 is unbiased for σ^2.

$$\Rightarrow \bar{X}^2 - \frac{S^2}{n} \text{ is unbiased for } \mu^2$$

⊙ If $X_1, X_2, X_3, \ldots X_n$ i.i.d $\sim U(0, \theta)$

Find the unbiased estimator of θ

A $\because \mathbb{E}(\bar{X}) = \mu$ = population mean

$$\Rightarrow \mathbb{E}(\bar{x}) = \frac{0+\theta}{2} = \frac{\theta}{2} \Rightarrow \mathbb{E}(\bar{X}) = \frac{\theta}{2} \Rightarrow 2\,\mathbb{E}(\bar{x}) = \theta$$

$$\Rightarrow \mathbb{E}(2\bar{x}) = \theta$$

$\Rightarrow 2\bar{X}$ is unbiased for θ

Ex: $x_1 = 0.1$, $x_2 = 0.2$, $x_3 = 0.4$, $x_4 = 0.6$ is a random sample from $U(0,\theta)$

Unbiased estimator of $\theta = 2\bar{x} = 2\left(\frac{0.1+0.2+0.4+0.6}{4}\right) = \boxed{0.65}$

But also see $\max\{X_1, X_2, X_3 \cdots X_n\} = Y$ (say)

Each $X_i \sim U(0,\theta)$
p.d.f of X_i is
$$f(x) = \begin{cases} \frac{1}{\theta} & 0<x<\theta \\ 0 & \text{else} \end{cases}$$

find p.d.f of Y.

$$F_Y(y) = \mathbb{P}(Y \le y) = \mathbb{P}(\max\{X_1 X_2 \cdots X_n\} \le y)$$
$$= \mathbb{P}(X_1 \le y, X_2 \le y, X_3 \le y, \cdots X_n \le y)$$
$$= \mathbb{P}(X_1 \le y)\,\mathbb{P}(X_2 \le y) \cdots \mathbb{P}(X_n \le y)$$
$$= \int_0^y \frac{1}{\theta} dx_1 \int_0^y \frac{1}{\theta} dx_2 \cdots \int_0^y \frac{1}{\theta} dx_n = \left(\frac{y}{\theta}\right)^n$$

$$F_Y(y) = (y/\theta)^n \Rightarrow f_Y(y) = n(y/\theta)^{n-1}\frac{1}{\theta} = \frac{n}{\theta^n} y^{n-1}$$

$$\Rightarrow f_Y(y) = \frac{n}{\theta^n} y^{n-1} \qquad 0<y<\theta$$

$$\mathbb{E}(Y) = \int_0^\theta y \cdot \frac{n}{\theta^n} y^{n-1}\, dy = \left(\frac{n}{n+1}\right)\theta$$

$$\Rightarrow \mathbb{E}(Y) = \left(\frac{n}{n+1}\right)\theta \Rightarrow \mathbb{E}\left(\left(\frac{n+1}{n}\right)Y\right) = \theta$$

$\Rightarrow \left(\frac{n+1}{n}\right)Y$ is unbiased for θ

$\Rightarrow$ $x_1 = 0.1$, $x_2 = 0.2$, $x_3 = 0.4$, $x_4 = 0.6$ is a r.sample from $U(0,\theta)$

$\Rightarrow$ $\left(\frac{4+1}{4}\right)(0.6) = \boxed{0.75}$ is unbiased estimate for θ

Q $x_1 = 0.5$, $x_2 = 0.6$, $x_3 = 0.7$, $x_4 = 0.1$ random sample from $U(0,\theta)$

Find two different unbiased estimates for θ.

A (i) $2\bar{x} = 2\left(\frac{0.5+0.6+0.7+0.1}{4}\right) = \boxed{0.95}$

(ii) $\left(\frac{n+1}{n}\right)\max_i\{X_i\} = \frac{5}{4}(0.7) = \boxed{0.875}$

The above argument shows that unbiased estimators are not UNIQUE.

Consistency (A property for the long run)

Another desirable property of an ESTIMATOR is, that it should be CONSISTENT.

Mathematically. A statistic T_n is a consistent estimator of θ of a given distribution iff $\forall \epsilon > 0$

$$\lim_{n\to\infty} \mathbb{P}(|T_n - \theta| < \epsilon) = 1$$

Ex: $X_1, X_2 \dots X_n$ i.i.d $\sim U(0,\theta)$

then we know that

$$\mathbb{E}\left(\left(\frac{n+1}{n}\right) X_{max}\right) = \theta$$

c.d.f of $Y = X_{max}$ is $F_Y(y) = \begin{cases} 0 & y < 0 \\ \left(\frac{y}{\theta}\right)^n & 0 \le y < \theta \\ 1 & \theta \geqslant y \end{cases}$

$$\begin{aligned} &\mathbb{P}(|Y - \theta| < \epsilon) \\ &= \mathbb{P}(-\epsilon < Y - \theta < \epsilon) \\ &= \mathbb{P}(\theta - \epsilon < Y < \epsilon + \theta) \\ &= F_Y(\theta + \epsilon) - F(\theta - \epsilon) \\ &= 1 - \left(\frac{\theta - \epsilon}{\theta}\right)^n \\ &= \boxed{1} \end{aligned}$$

$\because n \to \infty \Rightarrow \left(\frac{\theta - \epsilon}{\theta}\right)^n \to 0$

$\Rightarrow X_{max}$ is consistent for θ

Ex $x_1 = 0.1$, $x_2 = 0.2$, $x_3 = 0.4$, $x_4 = 0.6$ r. sample from $U(0,\theta)$

$\Rightarrow \boxed{0.6}$ is consistent for θ

$\underline{\underline{Q}}$ How to show in general whether a statistic T_n is consistent for θ

$\underline{\underline{A}}$ If

$$\mathbb{E}(T_n) = \theta_n \longrightarrow \theta \quad \text{as } n \longrightarrow \infty$$
$$V(T_n) = \sigma_n^2 \longrightarrow 0$$

then T_n is consistent for θ

IEx: If $X_1, X_2 \dots X_n$ be i.i.d with $\mathbb{E}(X_i) = \mu$, $V(X_i) = \sigma^2$

then we know $\mathbb{E}(\bar{X}) = \mu \longrightarrow \mu$ as $n \to \infty$

$V(\bar{X}) = \frac{\sigma^2}{n} \longrightarrow 0$ as $n \to \infty$

$\Rightarrow \bar{X}$ is consistent for μ.

If $X_1, X_2, \ldots, X_n$ be i.i.d from $N(\mu, \sigma^2)$ then S^2 is consistent for σ^2

(i) $\mathbb{E}(S^2) = \sigma^2 \longrightarrow \sigma^2$ as $n \to \infty$

(ii) $\underbrace{V(S^2) \longrightarrow 0}_{\text{show}}$ as $n \to \infty$

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

$$\Rightarrow V\left(\frac{n-1}{\sigma^2} . S^2\right) = 2(n-1)$$

$$\Rightarrow \frac{(n-1)^2}{\sigma^4} V(S^2) = 2(n-1)$$

$$\Rightarrow V(S^2) = \frac{2\sigma^4}{(n-1)}$$

$$\Rightarrow V(S^2) \longrightarrow 0 \quad \text{as } n \to \infty$$

$\Rightarrow S^2$ is Consistent for σ^2

Thus $\bar{X}$ is unbiased for μ $\quad$ S^2 is unbiased for σ^2

$\bar{X}$ is consistent for μ $\quad$ S^2 is consistent for σ^2

$\bar{X}$: sample mean.

S^2 : sample variance.

μ : Population mean.

σ^2 : Population variance.

$$\left\{ \begin{aligned} \bar{X} &= \frac{1}{n} \sum_{i=1}^{n} X_i \\ S^2 &= \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2 \end{aligned} \right\}$$

Remember.