STAT 231

October 26, 2016

Tutorial today

3-30 → Cyntha STP 105

6-00 → Banerjee, DC 1351

Video Review → Friday

<u>Roadmap</u>

- All definitions of interval estimation using sampling techniques.
- Chi-Squared.
- T-distribution
- Application.

INTERVAL ESTIMATION

ESTIMATE

ESTIMATOR.

COVERAGE INTERVAL

CONFIDENCE INTERVAL

SAMPLING DISTRIBUTION

PIVOTAL QUANTITY.

Underlying model

$$Y = (Y_1, \ldots Y_n) \sim f(y_i; \theta)$$

Sample: $\{y_1, \ldots y_n\}$ θ = unknown parameter.

ESTIMATE :

An estimate of θ $(\hat{\theta})$ is a function $g(y_1, \ldots y_n)$ which is your "best guess" of θ.

MLE s are estimates:

ESTIMATES are sometimes called POINT ESTIMATES

<u>ESTIMATOR</u> : is a random variable of which your estimate is an outcom

$\tilde{\theta}$ = POINT ESTIMATOR.

Example: $Y_1, .. Y_n \sim G(\mu, \sigma)$ with σ = known. We want to estimate μ = unknown.

$\bar{y}$ = sample mean = estimate

$\bar{Y}$ = estimator.

An estimator gives us the rule of what to calculate from our sample.

SAMPLING DISTRIBUTION

The sampling distribution of an estimator $\tilde{\theta}$ is the distribution that $\tilde{\theta}$ follows.

Example $Y_1, \dots Y_n \sim G(\mu, \sigma)$
σ known

$$\boxed{\bar{Y} \sim G(\mu, \sigma/\sqrt{n})}$$

SAMPLING DISTRIBUTION OF $\bar{Y}$

PIVOTAL QUANTITY

Definition: A pivotal quantity Q is a function of $Y_1, \dots Y_n$ and θ such that $P(Q \geq a)$, $P(Q \leq b)$ can be computed without knowing what the value of θ is

$$\bar{Y} \sim G(\mu, \sigma/\sqrt{n})$$

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

Q = pivotal quantity

We use the pivotal quantity to construct the COVERAGE INTERVAL.

$\boxed{\text{COVERAGE INTERVAL}}$

A coverage interval is two r.v.s $[L, U]$ such that

$$P(L \leq \theta \leq U) = 0.95$$

= or some pre-specified probability

CONFIDENCE INTERVAL

A confidence interval is $[l, u]$ which are estimates of L, U.

where l, u are calculated from our sample

THE CHI-SQUARED (χ^2) DISTRIBUTION

Properties

- $W \sim \chi^2_n$ if $W = Z_1^2 + \dots + Z_n^2$

  $Z_i \sim N(0,1)$

  Z_i's independent.

- W takes values $(0, \infty)$

- $E(W) = n$ = degrees of freedom

  $V(W) = 2n = 2 \times df$.

- If $W_1 \sim \chi^2_{n_1}$ and $W_2 \sim \chi^2_{n_2}$

then $W_1 + W_2 \sim \chi^2_{n_1+n_2}$

if W_1 and W_2 are independent.

Question 1

$W_1 \sim \chi^2_1$
$W_2 \sim \chi^2_2$
$\vdots$

(1) $W_i \sim \chi^2(i) \quad i = 1, \ldots 5$

W_i's are independent.

Let $Y = \sum_{i=1}^{5} W_i$

What distribution does Y follow?

(a) $\chi^2(5)$

$\boxed{\text{(b) } \chi^2(15)}$

(c) $N(0, 15)$

d) can't say

$Y = \chi^2 \quad 1+2+3+4+5 = 15$

Question 2

Suppose $Y \sim G(3, 9)$

Y and Z indep. $Z \sim G(0, 1)$

Let ~~Y~~ $W = \left(\frac{Y-3}{9}\right)^2 + Z^2$.

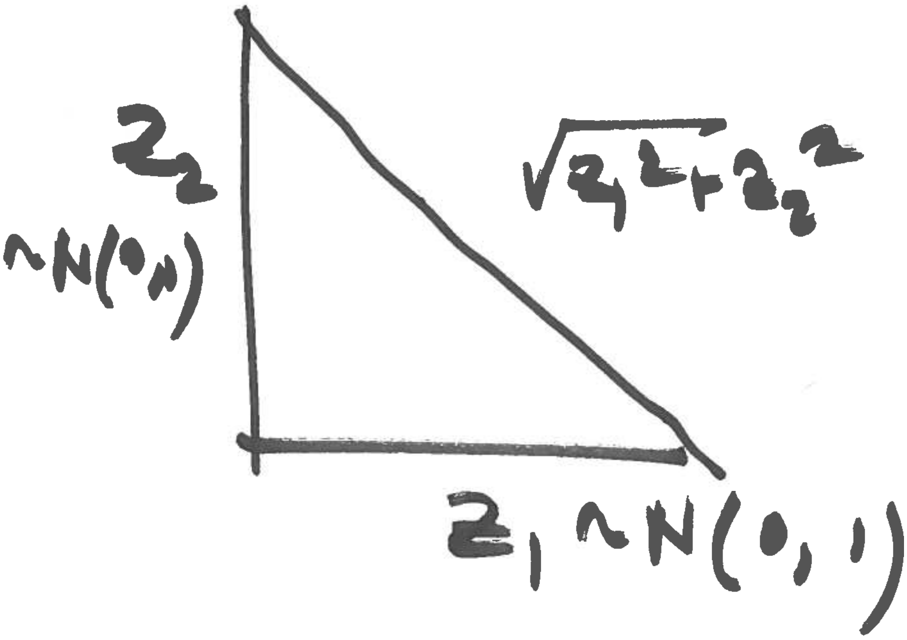
What distribution does W follow?

(a) $\chi^2(1)$ $\boxed{\text{(b) } \chi^2(2)}$ 77%

(c) $N(0,2)$ (d) none of the above.

Calculation purposes $k = df$

- $k = 1 \longrightarrow W = Z^2$
- $k = 2 \longrightarrow W \sim Exp(2)$
- $k > 50 \longrightarrow W \sim G(k, \sqrt{2k})$
- $k \in (2, 50) \longrightarrow$ Consult tables to get probabilities



$$z_1^2 + z_2^2 \sim \chi^2(2)$$

CHI-SQUARED TABLE

Z-table / Chi-Square.

Z-table → #s are probabilities

Chi-Square → Quantiles

Row = degrees of freedom.

Column = Level of quantile

Row = 5   Column = 0.4

$\boxed{3.656}$ → Table

Row = 5 Column = 0.4

Value = 3.656

$P(W \leq 3.656) = 0.4$

where $\chi^2_5 = W$

Example: Suppose $W \sim \chi^2_{20}$

$P(W \leq 13) = ?$ (from the table).

Row = 20 = degrees of freedom

12.443 → 0.1

14.578 → 0.2

The probability of $W \leq 13$ lies between 0.1 and 0.2

STUDENT'S T- distribution

Definition: A random variable T is said to follow a Student's T-distribution with n degrees of freedom. If T is a ratio of two independent r.v.s.

$$T = \frac{Z}{Y} \quad \text{where } Z \sim N(0,$$

$$\cancel{Y = \sqrt{\chi^2_n} \,/\, n}$$

$$Y = \sqrt{\frac{W}{n}} \quad \text{where } W \sim \chi^2_n .$$

Properties of the T- distribution

- The T- distribution is symmetric around zero
- It looks similar to Z but with fatter tails (KURTOSIS > 3)

T will have more extreme obs. compared to Z.

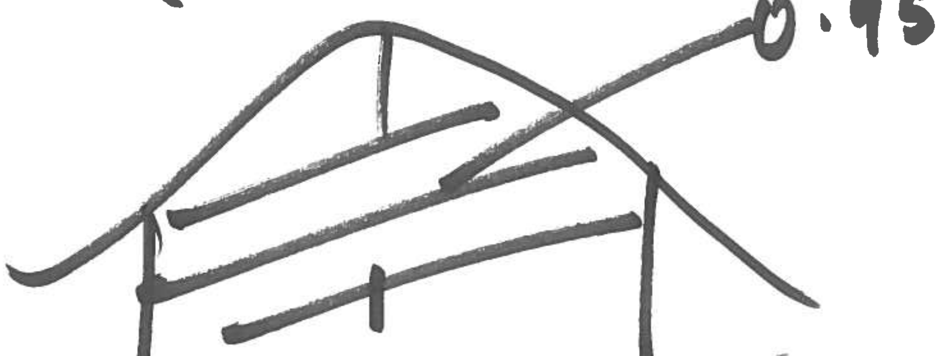
: As the degrees of freedom $n \longrightarrow \infty$, the T distribution approaches the Z- distribution

---

<u>Example</u> : Suppose $T \sim T_{15}$

Find a such that

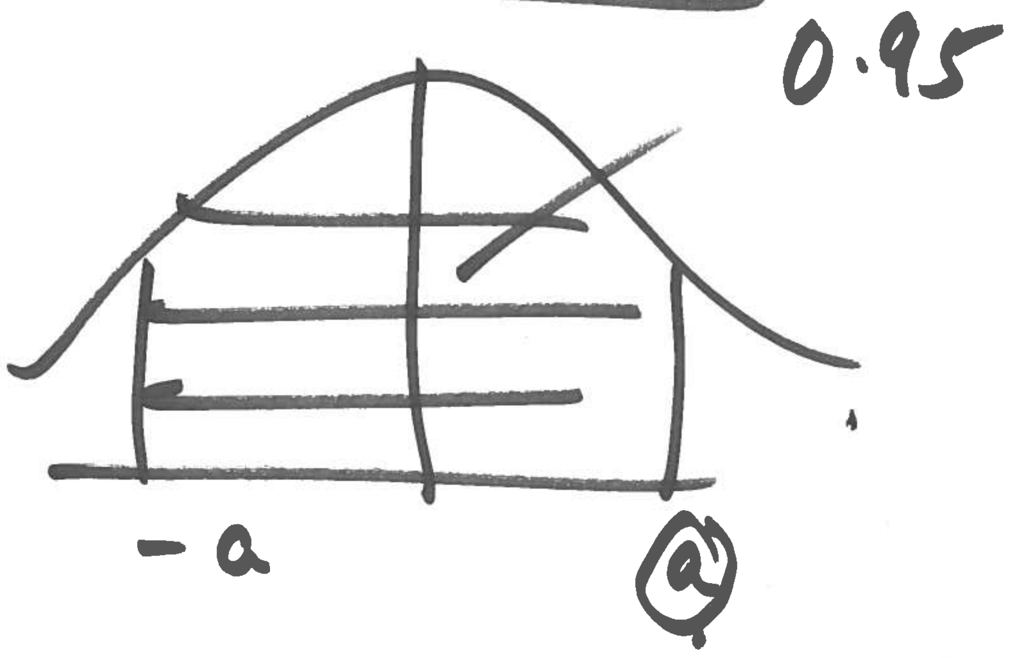
$$P(|T| \leq a) = 0.95$$



T- table

Row = df

Entries: Quantiles



Row : 15 Column = 0.975

a = 2.1314