UNIVERSITAS GADJAH MADA
DEPARTMENT OF CIVIL AND ENVIRONMENTAL ENGINEERING
BACHELOR IN CIVIL ENGINEERING

Statistics and Probability

Confidence Intervals

Confidence intervals

- Estimate on parameters
  - Probability distribution has a number of parameters, which are mostly unknown.
  - Their values are estimated based on values obtained from samples.
  - Estimates
    - point estimates
    - confidence intervals

Confidence intervals

- Examples of point estimates
  - Mean of sample as an estimate of population mean

    $\bar{X} \rightarrow \mu$

  - Standard deviation of sample as an estimate of standard deviation of the population

    $s_X \rightarrow \sigma$

Confidence intervals

- Estimate on parameter θ
  - $\hat{\theta} \rightarrow \theta$

    estimate ↓   ↓ parameter
  - Find an interval [L,U] such that the probability that this interval contains θ is $(1-\alpha)$
  - $\text{prob}(L < \theta < U) = 1 - \alpha$
  - L and U are both random variables
    - L lower confidence limit
    - L upper confidence limit
    - $1-\alpha$ confidence level

Confidence intervals

- Example
  - Records on discharge at River A from 1981 to 2000 indicate that the average discharge at this river is 77 m^3/s.
    - We may assume that the average discharge at River A is 77 m^3/s.
    - We realize that the above estimate might be wrong; moreover, we know that from probability theory point of view, the probability of discharge equal to 77 m^3/s is nearly impossible.

$$\mathrm{prob}\left(\bar{Q} = 77\ \mathrm{m}^3/s\right) = 0$$

Lower and upper confidence limits

- Method of pivotal quantities: Ostle method
  - Finding a random variable V that is a function of the parameter θ but whose distribution does not involve any other unknown parameters.
  - Determine v_1 and v_2 such that

$$\text{prob}(v_1 < V < v_2) = 1 - \alpha$$

  - The above equation is then changed into the following form

$$\text{prob}(L < \theta < U) = 1 - \alpha$$

  where L and U are random variables and function of V, but they are not function of θ

Confidence interval: mean of a normal distribution

- Looking for an interval of $[L, U]$ that contains μ with the probability of $(1 - \alpha)$

$$\text{prob}(L < \mu < U) = 1 - \alpha$$

- A random variable V is defined as follow

$$V = \frac{\bar{X} - \mu}{s_{\bar{\bar{X}}}}$$

  - V has a *t* distribution with $(n - a)$ degrees of freedom
  - n is the sample size (number of observations) used to estimate the sample mean $\bar{X}$

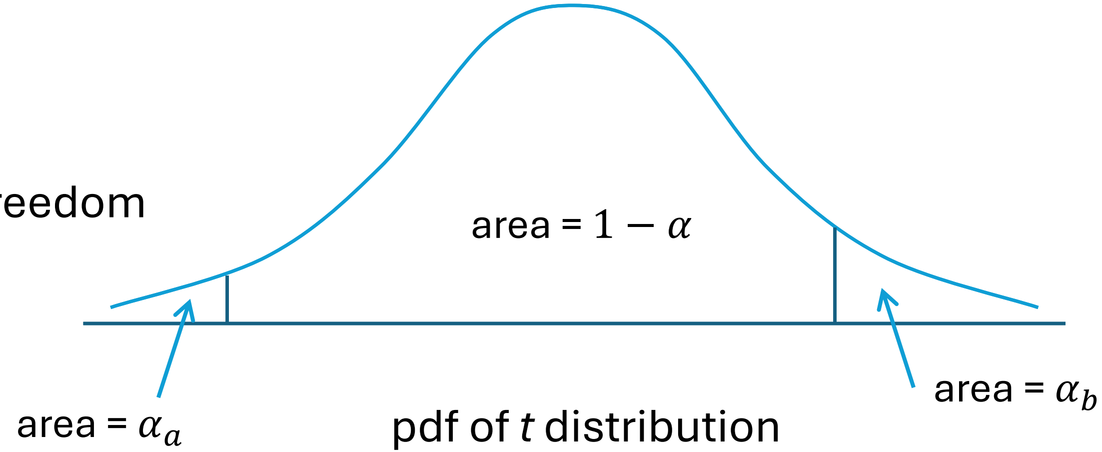
$$\text{prob}(v_1 < V < v_2) = 1 - \alpha \Rightarrow \text{prob}\left(v_1 < \frac{\bar{X} - \mu}{s_{\bar{X}}} < v_2\right) = 1 - \alpha$$

$$\alpha_a + \alpha_b = \alpha$$

$$\text{prob}(t < v_1) = \alpha_a$$

$$\text{prob}(t > v_2) = \alpha_b$$

with $(n - 1)$ degrees of freedom



pdf of *t* distribution

- Table of *t* distribution
- Functions in Microsoft Excel

$$\text{prob}\left(v_1 < \frac{\bar{X}-\mu}{s_{\bar{X}}} < v_2\right) = 1-\alpha$$

$$\text{prob}\left(t_{\alpha_a,n-1} < \frac{\bar{X}-\mu}{s_{\bar{X}}} < t_{1-\alpha_b,n-1}\right) = 1-\alpha$$

$$\text{prob}\big(\underbrace{\bar{X} + t_{\alpha_a,n-1}\cdot s_{\bar{X}}}_{\text{lower limit}} < \mu < \underbrace{\bar{X} + t_{1-\alpha_b,n-1}\cdot s_{\bar{X}}}_{\text{upper limit}}\big) = 1-\alpha$$

Confidence limits

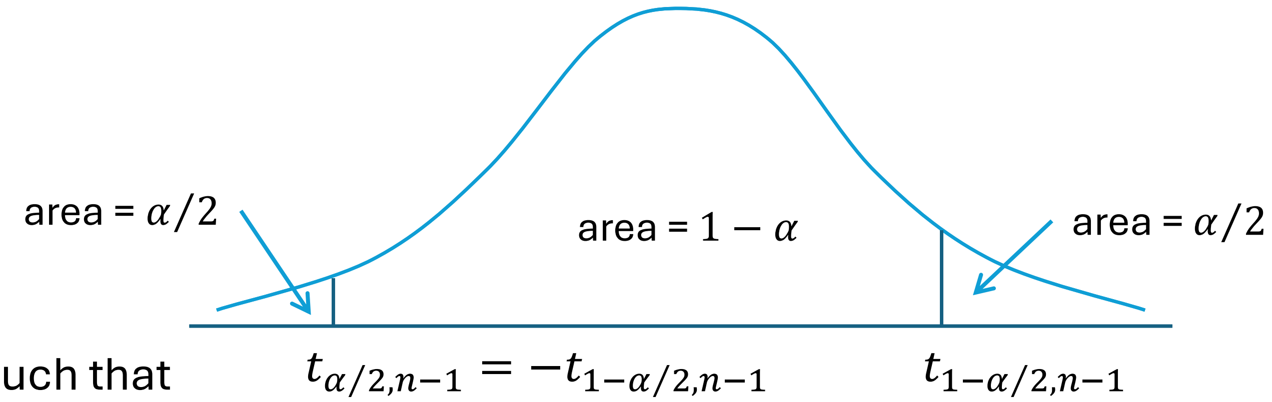
$$\ell = \bar{X} + t_{\alpha_a,n-1}\cdot s_{\bar{X}}$$

$$u = \bar{X} + t_{1-\alpha_b,n-1}\cdot s_{\bar{X}}$$

$s_{\bar{X}} = s_X/\sqrt{n}$

$t_{\alpha_a,n-1}$, $t_{1-\alpha_b,n-1}$: table of t distribution

- If it is desired that the confidence interval be symmetrical in probability
  - v_1 and v_2 can be chosen such that the probability that a random t is less than v_1 equals the probability that a random *t* exceeds v_2
  - Since the probability is symmetrical, thus $\alpha_a = \alpha_b = \alpha/2$
  - We look for $(1-\alpha) = 100(1-\alpha)\%$ confidence interval, thus $\text{prob}(t < v_1) = \alpha/2 = \text{prob}(t > v_2)$

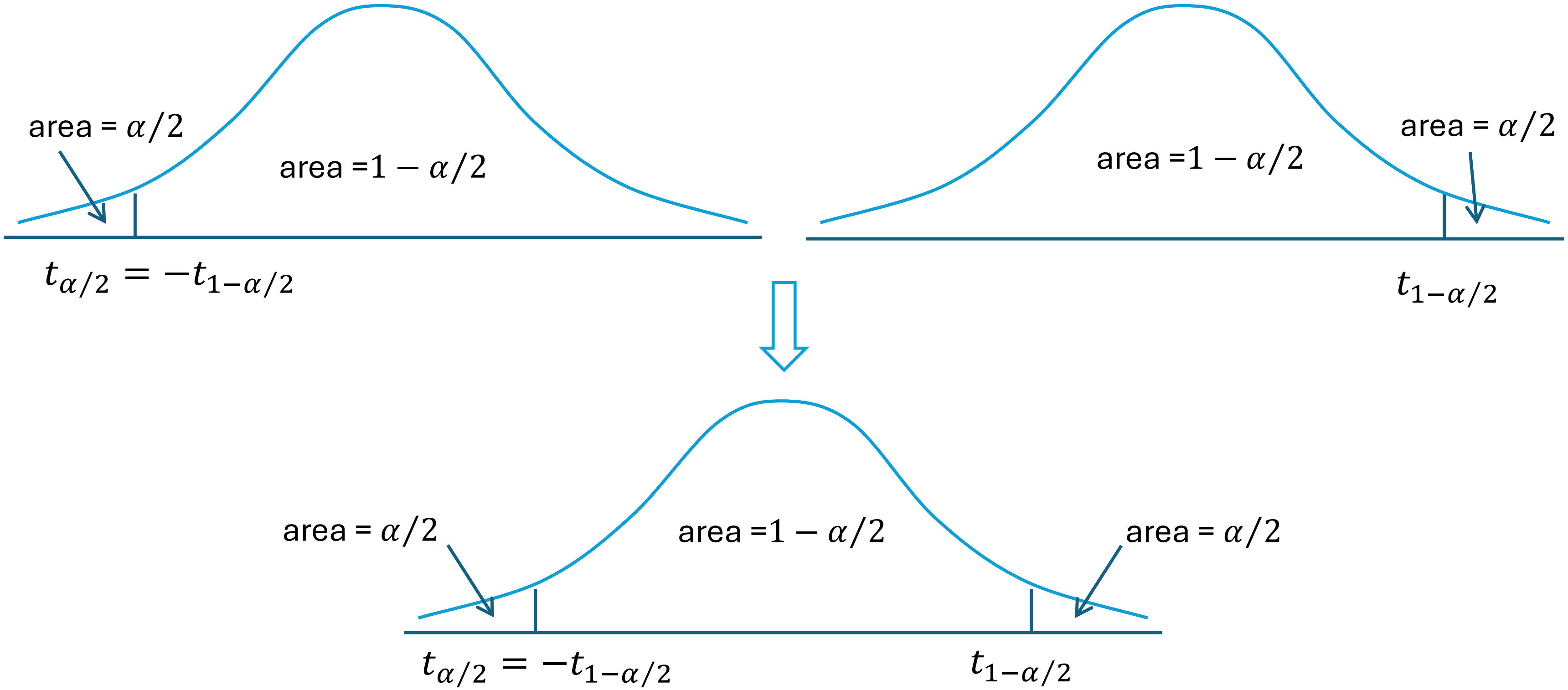


t_α is value of *t* such that $\text{prob}(T < t_\alpha) = \alpha$

pdf of *t* distribution

t distribution

degrees of freedom $n - 1$



- The confidence limits when the probability of the confidence interval is symmetrical are

$$\ell = \bar{X} - t_{1-\alpha/2,n-1} \cdot s_{\bar{X}}$$

$$u = \bar{X} + t_{1-\alpha/2,n-1} \cdot s_{\bar{X}}$$

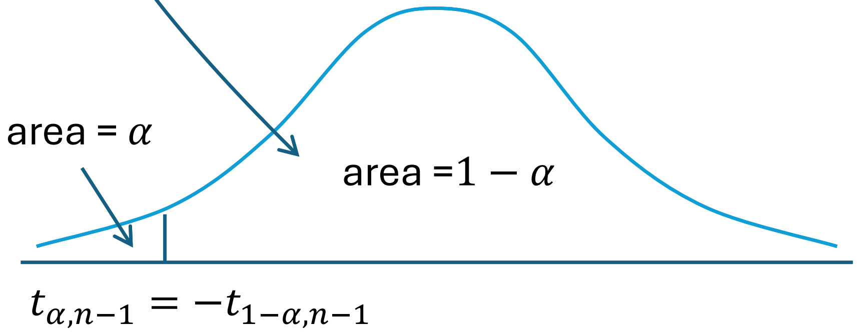
$$\text{prob}\left(\bar{X} + t_{\propto_a,n-1}\frac{s_X}{\sqrt{n}} < \mu < \bar{X} + t_{1-\propto_b,n-1}\frac{s_X}{\sqrt{n}}\right) = 1 - \alpha$$

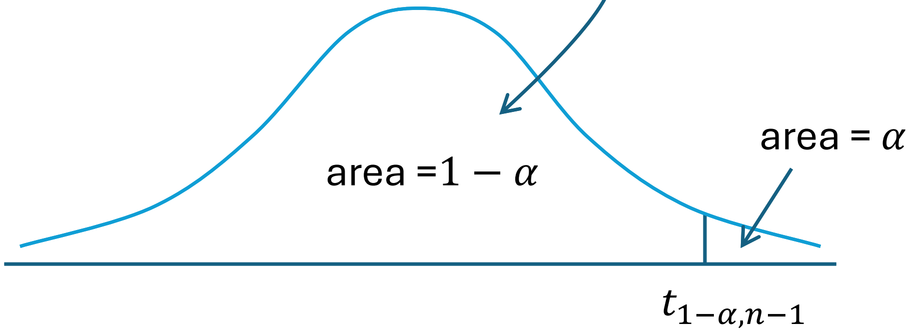
$$\text{prob}\left(\bar{X} - t_{1-\frac{\alpha}{2},n-1}\frac{s_X}{\sqrt{n}} < \mu < \bar{X} + t_{1-\frac{\alpha}{2},n-1}\frac{s_X}{\sqrt{n}}\right) = 1 - \alpha$$

- If it is desired that the confidence interval be one-sided in probability:
  - lower limit → $\text{prob}(t < v_1) = \alpha$
  - upper limit → $\text{prob}(t > v_2) = \alpha$

$$\text{prob}(t > v_1) = 1 - \alpha \Longrightarrow \text{prob}\left(\frac{\bar{X} - \mu}{s_{\bar{X}}} > v_1\right) = 1 - \alpha$$

$$\text{prob}(t < v_2) = 1 - \alpha \Longrightarrow \text{prob}\left(\frac{\bar{X} - \mu}{s_{\bar{X}}} < v_2\right) = 1 - \alpha$$





t distribution

- Functions in Microsoft Excel
  - =T.DIST(...)
  - =T.DIST.2T(...)
  - =T.DIST.2R(...)
  - =T.INV(...)
  - =T.INV.2T(...)

Confidence interval: mean of a normal distribution, known σ²

- If the population variance of the normal distribution is **known** (it is a rare cases)
  - The random variable *V* is defined as follow

$$V = \frac{\bar{X} - \mu}{\sigma_{\bar{X}}}, \qquad \sigma_{\bar{X}} = \sigma_X / \sqrt{n} \qquad V \text{ is normally distributed}$$

  - The confidence interval (symmetrical)

$$\text{prob}\left(\bar{X} - z_{1-\alpha/2}\frac{\sigma_X}{\sqrt{n}} < \mu < \bar{X} + z_{1-\alpha/2}\frac{\sigma_X}{\sqrt{n}}\right) = 1 - \alpha$$

Confidence interval: mean of a normal distribution, known σ^2

- Confidence limits

$$\ell = \bar{X} + z_{\alpha_a} \frac{\sigma_X}{\sqrt{n}}$$

$$u = \bar{X} + z_{1-\alpha_b} \frac{\sigma_X}{\sqrt{n}}$$

- Symmetrical confidence limits

$$\ell = \bar{X} - z_{1-\alpha/2} \frac{\sigma_X}{\sqrt{n}}$$

$$u = \bar{X} + z_{1-\alpha/2} \frac{\sigma_X}{\sqrt{n}}$$

Confidence interval variance of a normal distribution

- Looking for an interval of $[L, U]$ that contains σ^2 with the probability of $(1-\alpha)$

$$\text{prob}\left(v_1 < \sigma^2 < v_2\right) = 1 - \alpha$$

- A random variable V is defined as follow

$$V = \frac{(n-1){s_X}^2}{{\sigma_X}^2}$$

  - V has a chi-square distribution with $(n-1)$ degrees of freedom

Confidence interval variance of a normal distribution

$$\text{prob}(v_1 < V < v_2) = 1 - \alpha$$

$$\text{prob}\left(v_1 < \frac{(n-1){s_X}^2}{{\sigma_X}^2} < v_2\right) = 1 - \alpha$$

select $v_1 = \chi^2_{\alpha/2,n-1}$

$v_2 = \chi^2_{1-\alpha/2,n-1}$

such that $\text{prob}\left(\chi^2_{\alpha/2,n-1} < \frac{(n-1){s_X}^2}{{\sigma_X}^2} < \chi^2_{1-\alpha/2,n-1}\right) = 1 - \alpha$

or $\text{prob}\left(\frac{(n-1){s_X}^2}{\chi^2_{1-\alpha/2,n-1}} < {\sigma_X}^2 < \frac{(n-1){s_X}^2}{\chi^2_{\alpha/2,n-1}}\right) = 1 - \alpha$

Confidence interval variance of a normal distribution

- The lower and upper limits of the confidence interval containing the variance of normal distribution is
  - Lower limit $\ell = \dfrac{(n-1){s_X}^2}{\chi^2_{1-\alpha/2,n-1}}$
  - Upper limit $u = \dfrac{(n-1){s_X}^2}{\chi^2_{\alpha/2,n-1}}$

$$\text{prob}\left(\frac{(n-1){s_X}^2}{\chi^2_{1-\alpha/2,n-1}} < {\sigma_X}^2 < \frac{(n-1){s_X}^2}{\chi^2_{\alpha/2,n-1}}\right) = 1-\alpha$$

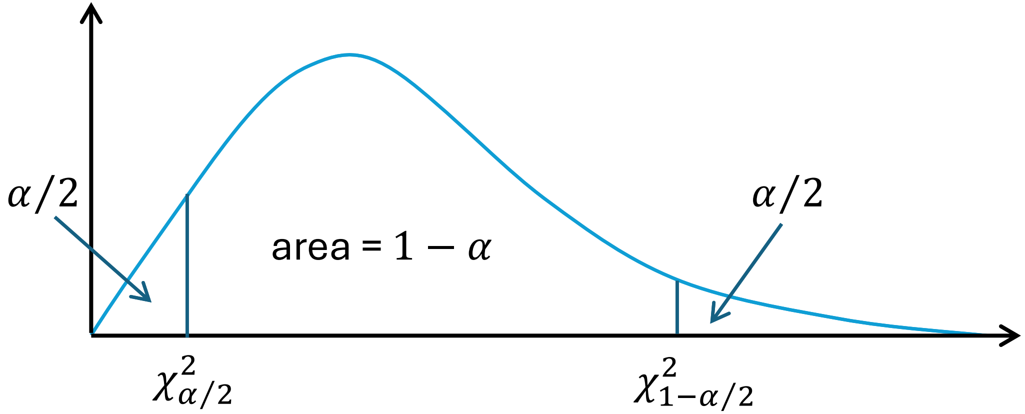
  - X is normally distributed and χ^2 is chi-squarely distributed

Confidence interval variance of a normal distribution

- The chi-square distribution is not symmetrical, thus

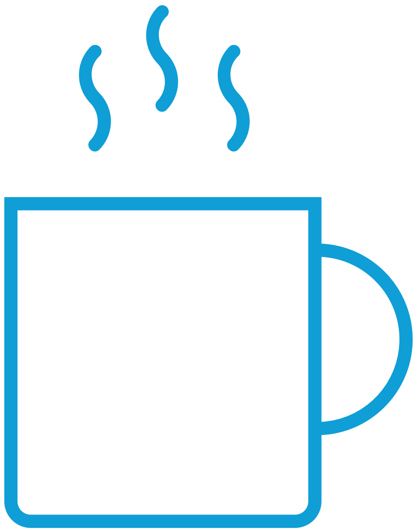
$$s_X{}^2 - \ell \neq u - s_X{}^2$$

- For large sample size, the chi-square distribution approaches a symmetrical distribution



Chi-square distribution

- Functions in Microsoft Excel
  - =CHISQ.DIST(...)
  - =CHISQ.DIST.RT(...)
  - =CHISQ.INV(...)
  - =CHISQ.INV.RT(...)



Statistics and Probability

Confidence Intervals