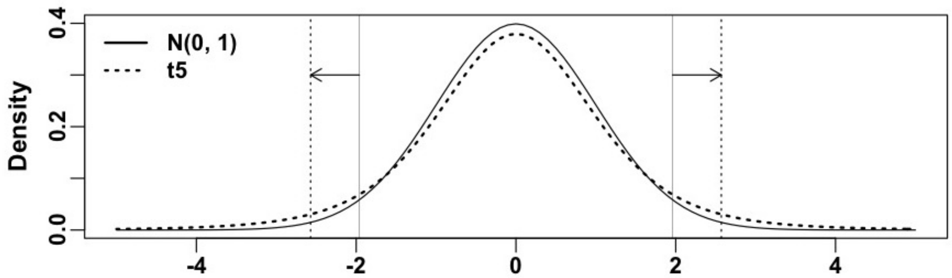
Packet 4: Interval Estimation

Large sample results:

The t_ν-distribution is heavy-tailed:



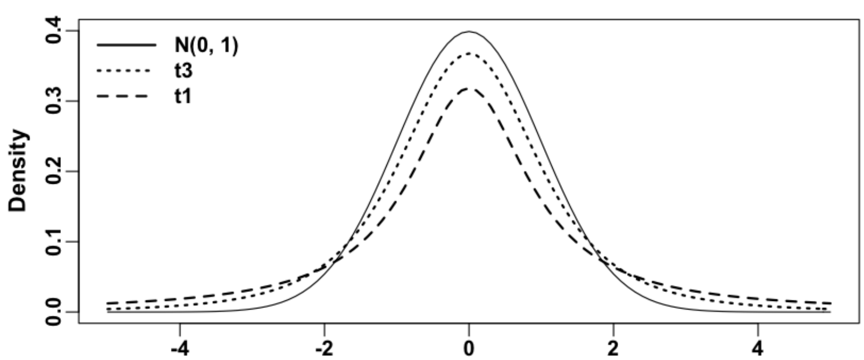
For example, if $T \sim t_5$ ($n = 6$),

$$P(-1.96 < Z < 1.96) = P(-2.57 < T_5 < 2.57) = 0.95.$$

$$P(-2.23 < T_{10} < 2.23) = P(-2.09 < T_{20} < 2.09) = 0.95.$$

$$P(-2.04 < T_{30} < 2.04) = P(-1.98 < T_{100} < 1.98) = 0.95.$$

As the sample size n increases, the t_{n-1} distribution approaches the standard Normal distribution.



Thus, if n is large (typically $n \geq 30$), $t_{n-1} \dot\sim N(0,1)$, and a 95% confidence interval for μ becomes close to

$$\left(\bar{X} - 1.96\frac{S}{\sqrt{n}}, \quad \bar{X} + 1.96\frac{S}{\sqrt{n}}\right).$$

If $X_1, X_2, \ldots, X_n$ are i.i.d. but do NOT follow a $N(\mu, \sigma^2)$, we can still obtain approximated C.I. of μ from C.L.T., which tells us that, with sufficiently many i.i.d. samples collected, the sample mean $\bar{X}$ follows $N(\mu, \sigma^2/n)$ approximately. In addition, $\sigma^2 \approx S^2$ when n is large.

The $100(1-\alpha)\%$ confidence interval of μ is

$$\left(\bar{X} - Z_{\alpha/2}\frac{S}{\sqrt{n}}, \quad \bar{X} + Z_{\alpha/2}\frac{S}{\sqrt{n}}\right).$$

Example: If we knew there were 100 ducks living in the neiborhood. We observed 40 of them when we visited Penn State Duck Pond. We assume that $X_1, X_2, \ldots, X_{100} \overset{i.i.d.}{\sim}$ Bernoulli(θ) where X_i is 1 if the duck i were at Duck Pond when we visited and 0 otherwise.

Find a 95% confidence interval for θ the probability of a duck appeared at Duck Pond.

One-sided confidence interval for μ when $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$.

Sometimes we might want only a lower or upper bound on μ, and it is also possible to form a one-sided confidence interval, for example,

- Rainfall amount for flood hazards.

  The upper one-sided confidence interval is $(-\infty,\ u(X)]$.

  $$P(\theta \leq u(X)) = 1 - \alpha.$$

- Wind speed for power supply.

  The lower one-sided confidence interval is $[l(X),\ \infty)$ such that

  $$P(l(X) \leq \theta) = 1 - \alpha.$$

Example: In vdB (1985), 148 ($= n$) galactic globular clusters are observed, and the average brightness is $\bar{x} = -7.1$ magnitude (an astronomical measure of brightness). Assuming that $\sigma = 1.2$ mag,

1. Find a 95% confidence interval for the population average brightness μ.
2. Find a 95% one sided confidence interval with the lower bound for the population average brightness μ.

Summary of confidence interval for μ

1. If σ^2 is known, a $100(1-\alpha)\%$ confidence interval for μ is

$$\left(\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}},\quad \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right),$$

   where $z_{\alpha/2}$ is a constant satisfying $P(Z > z_{\alpha/2}) = \frac{\alpha}{2}$ if $Z \sim N(0,1)$.

   An one sided $100(1-\alpha)\%$ confidence interval with lower bound for μ is

$$\left(\bar{X} - z_{\alpha}\frac{\sigma}{\sqrt{n}},\quad \infty\right);$$

   An one sided $100(1-\alpha)\%$ confidence interval with upper bound for μ is

$$\left(-\infty,\quad \bar{X} + z_{\alpha}\frac{\sigma}{\sqrt{n}}\right),$$

   where z_{α} is a constant satisfying $P(Z > z_{\alpha}) = \alpha$ if $Z \sim N(0,1)$.

2. If σ^2 is unknown, a $100(1-\alpha)\%$ confidence interval for μ is

$$\left(\bar{X} - t_{\alpha/2}\frac{S}{\sqrt{n}},\quad \bar{X} + t_{\alpha/2}\frac{S}{\sqrt{n}}\right),$$

   where $t_{\alpha/2}$ is a constant such that $P(T > t_{\alpha/2}) = \frac{\alpha}{2}$ if $T \sim t_{n-1}$.

   An one sided $100(1-\alpha)\%$ confidence interval with lower bound for μ is

$$\left(\bar{X} - t_{\alpha}\frac{S}{\sqrt{n}},\quad \infty\right);$$

   An one sided $100(1-\alpha)\%$ confidence interval with upper bound for μ is

$$\left(-\infty,\quad \bar{X} + t_{\alpha}\frac{S}{\sqrt{n}}\right),$$

   where t_{α} is a constant satisfying $P(T > t_{\alpha}) = \alpha$ if $T \sim t_{n-1}$.

3. If σ^2 is unknown and n is large ($n \geq 30$), a $100(1-\alpha)\%$ confidence interval for μ is

$$\left(\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \quad \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right).$$

An one sided $100(1-\alpha)\%$ confidence interval with lower bound for μ is

$$\left(\bar{X} - z_{\alpha}\frac{\sigma}{\sqrt{n}}, \quad \infty\right);$$

An one sided $100(1-\alpha)\%$ confidence interval with upper bound for μ is

$$\left(-\infty, \quad \bar{X} + z_{\alpha}\frac{\sigma}{\sqrt{n}}\right),$$

where the standard deviation σ could be replaced by the sample standard deviation $S = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}}$ if σ is unknown.

Factors that affect the width of the confidence interval

Take two sided cofidence interval as an example:

Chap 7.2 Confidence interval for the difference of two means

We are interested in comparing two populations denoted by X and Y. We independently collect two sets of random samples:

$X_1, \ldots, X_n \sim N(\mu_X, \sigma_X^2)$, $Y_1, \ldots, Y_m \sim N(\mu_Y, \sigma_Y^2)$.

Find $100(1-\alpha)\%$ confidence interval of $\mu_X - \mu_Y$.

Case I: If σ_X^2 and σ_Y^2 are both known,

C.I. for $\mu_X - \mu_Y$:

1. Two sided $100(1-\alpha)\%$ confidence interval:

$$\left(\bar{X} - \bar{Y} - z_{\alpha/2}\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}},\ \ \bar{X} - \bar{Y} + z_{\alpha/2}\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}\right),$$

2. An one sided $100(1-\alpha)\%$ confidence interval with lower bound for μ is

$$\left(\bar{X} - \bar{Y} - z_{\alpha}\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}},\ \ \infty\right);$$

3. An one sided $100(1-\alpha)\%$ confidence interval with upper bound for μ is

$$\left(-\infty,\ \ \bar{X} - \bar{Y} + z_{\alpha}\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}\right),$$

Case II: If σ_X^2 and σ_Y^2 are both unknown, but we may assume $\sigma_X^2 = \sigma_Y^2$

C.I. for $\mu_X - \mu_Y$:

1. Two sided $100(1-\alpha)\%$ confidence interval:

$$\left(\bar{X} - \bar{Y} - t_{\alpha/2,(n+m-2)}\sqrt{\frac{S_p^2}{n} + \frac{S_p^2}{m}},\ \ \bar{X} - \bar{Y} - t_{\alpha/2,(n+m-2)}\sqrt{\frac{S_p^2}{n} + \frac{S_p^2}{m}}\right),$$

2. An one sided $100(1-\alpha)\%$ confidence interval with lower bound for μ is

$$\left(\bar{X} - \bar{Y} - t_{\alpha,(n+m-2)}\sqrt{\frac{S_p^2}{n} + \frac{S_p^2}{m}},\ \ \infty\right);$$

3. An one sided $100(1-\alpha)\%$ confidence interval with upper bound for μ is

$$\left(-\infty,\ \ \bar{X} - \bar{Y} + t_{\alpha,(n+m-2)}\sqrt{\frac{S_p^2}{n} + \frac{S_p^2}{m}}\right),$$

Case III: If σ_X^2 and σ_Y^2 are both unknown and $\sigma_X^2 \neq \sigma_Y^2$

When n and m are both large, then

Case IV: paired samples

Suppose X and Y are always collected in pairs (X_1, Y_1), $(X_2, Y_)$,..., (X_n, Y_n).

$X_i \sim N(\mu_X, \sigma_X^2)$ – blood pressure before treatment for patient i.

$Y_i \sim N(\mu_Y, \sigma_Y^2)$ – blood pressure after treatment for patient i.

To know whether treatment is effective, we estimate $\mu_X - \mu_Y$.

Note that X_i and Y_i are not independent! Let $D_i = X_i - Y_i \sim N(\mu_X - \mu_Y, \sigma_X^2 + \sigma_Y^2)$. It is reduce to a one-population problem with $D = X - Y$, and the parameter of interest being $\mu_D = \mu_X - \mu_Y$.

Two sided $100(1-\alpha)\%$ confidence interval:

1. If σ_D^2 is known:

2. If σ_D^2 is unknown:

3. If sample size is large ($n \geq 30$):

Example 7.2-1: $n = 15$, $m = 8$, $\bar{x} = 70.1$, $\bar{y} = 75.3$, $\sigma_X^2 = 60$, $\sigma_Y^2 = 40$. X and Y are independent and both follow normal distributions. Find a 95% confidence interval.

Exercise 7.2-9: Students attend a health-fitness program. $X_1, \ldots, X_{10}$ are the % of body fat before the program, and $Y_1, \ldots, Y_{10}$ are the % of body fat after the program. Assume the % of body fat follows a normal distribution. Find a 90% confidence interval for the mean reduction of the % of body fat before the program.

More Examples: 7.2-2, 7.2-4