1 Confidence intervals (interval estimates)

Suppose that we have a point estimate θ^*_{obs} of the parameter θ.If the distribution of the sample variable θ^* is continuous the probability that we have the correct parameter value is zero. What should we do about this?

Definition 1 *Let $x_1, \dots, x_n$ be observations of independent random variables $X_1, \dots, X_n$. A* **confidence interval** *$I_\theta(x_1, \dots, x_n)$ is an interval such that*

$$P(\theta \in I_\theta(X_1, \dots, X_n)) = 1 - \alpha$$

Here θ is the true parameter value and the probability $1 - \alpha$ is the **confidence level** *of the interval.*

Remark 1 You can thus view $I_\theta(x_1, \dots, x_n)$ as an observation of $I_\theta(X_1, \dots, X_n)$.

In order to construct a confidence interval for a parameter θ one can proceed in the following steps:

1. Determine a point estimate θ^*_{obs} of θ (possibly using the ML or LS method, see the compiled formulae Sections 9.1 and 9.2 respectively).

2. Determine the distribution of the sample variable (or statistic) θ^* (here you might find Sections 11.1-11.3 in the compiled formulae helpful for normally distributed observations, and the approximations in Sections 5 and 6 for non-normal observations).

3. Derive a confidence interval (Section 12.1-12.4 in the compiled formulae).

1.1 Application to the normal distribution

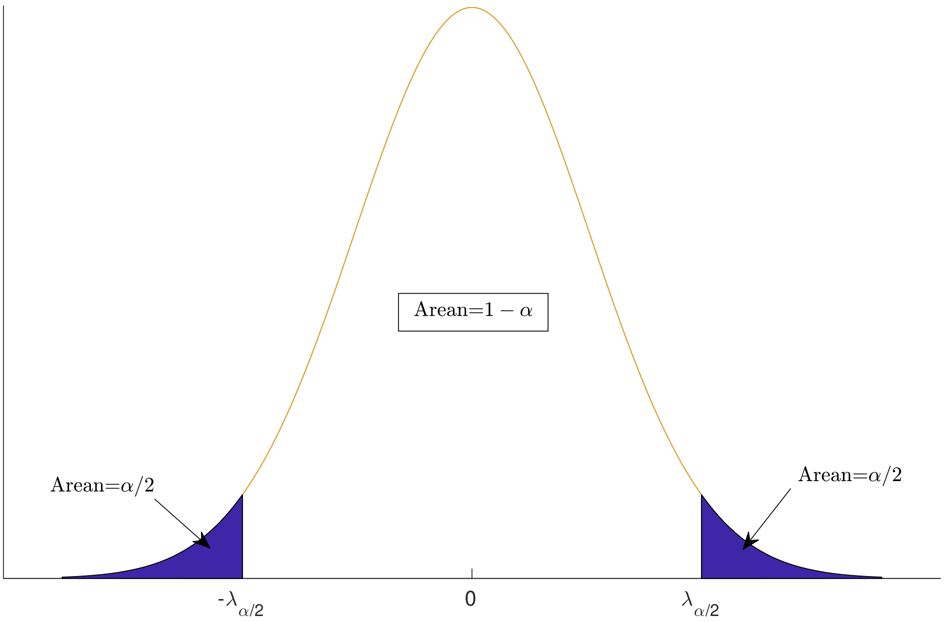
Suppose that we have observations $x_1, \dots, x_n$ of independent random variables $X_1, \dots, X_n$ where $X_i \in N(\mu, \sigma)$.

1.1.1 Confidence interval for μ when σ is known

1. We estimate μ by
$$\mu^*_{obs} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

2. For the random sample μ^* we have that
$$\mu^* = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \in N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$
since it is a linear combination of independent normally distributed random variables. For the calculations of expectation and variance, see Example 3 in Lecture 8.

3. For an estimate θ^*_{obs} such that $\theta^* \in N(\theta, D)$ where D is known we have that
$$\frac{\theta^* - \theta}{D} \in N(0, 1)$$



Figur 1: Illustration of the $\alpha/2$-quantile of the standard normal $N(0,1)$ (Arean=the area)

and therefore it holds with probability $1-\alpha$ that

$$-\lambda_{\alpha/2} \leq \frac{\theta^* - \theta}{D} \leq \lambda_{\alpha/2}$$

or after some rewriting

$$-\lambda_{\alpha/2} \cdot D \leq \theta^* - \theta \leq \lambda_{\alpha/2} \cdot D$$

which finally results in

$$\theta^* - \lambda_{\alpha/2} \cdot D \leq \theta \leq \theta^* + \lambda_{\alpha/2} \cdot D.$$

A confidence interval for θ with confidence level$1-\alpha$ is, given the assumptions above, given by

$$I_\theta = (\theta^*_{obs} - \lambda_{\alpha/2} \cdot D, \theta^*_{obs} + \lambda_{\alpha/2} \cdot D) \qquad (1-\alpha)$$

A confidence interval for μ with confidence level $1-\alpha$ is, given the above assumptions, given by (use $\theta^*_{obs} = \mu^*_{obs} = \bar{x}$ and $D = \sigma/\sqrt{n}$)

$$I_\mu = \left(\bar{x} - \lambda_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}, \bar{x} + \lambda_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right) \qquad (1-\alpha)$$

1.1.2 Confidence interval for μ when σ is unknown

1. We estimate μ by

$$\mu^*_{obs} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

just as before.

2. For the sample variable μ^* it still holds that

$$\mu^* = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \in N(\mu, \frac{\sigma}{\sqrt{n}})$$

or rewritten that

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

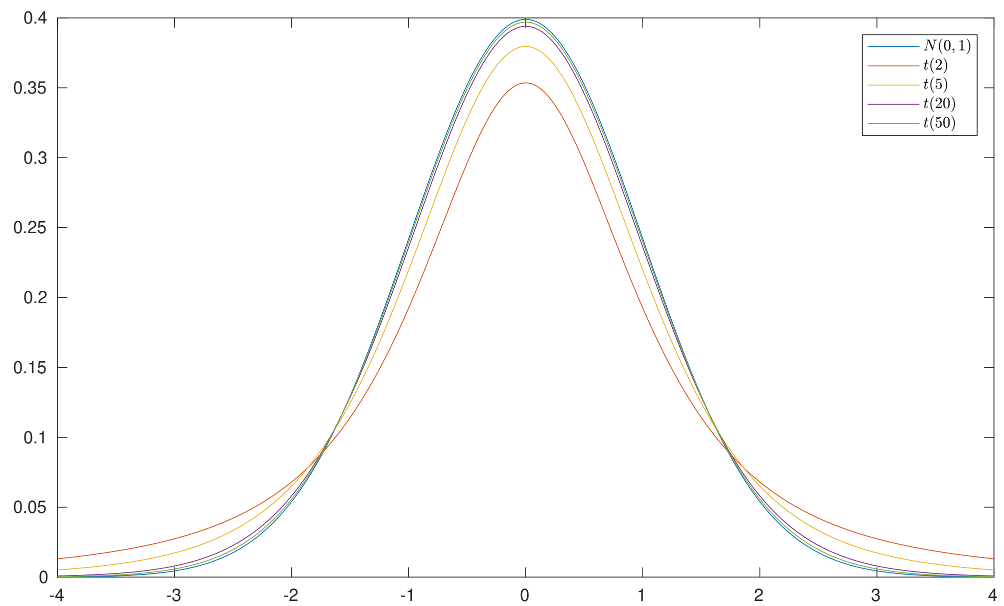
but since σ is unknown this does not help us. Instead we estimate σ by the sample standard deviation

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}.$$

It then holds that

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} \in t(n-1)$$

where $t(n-1)$ is short for the t-distribution with $n-1$ degrees of freedom.

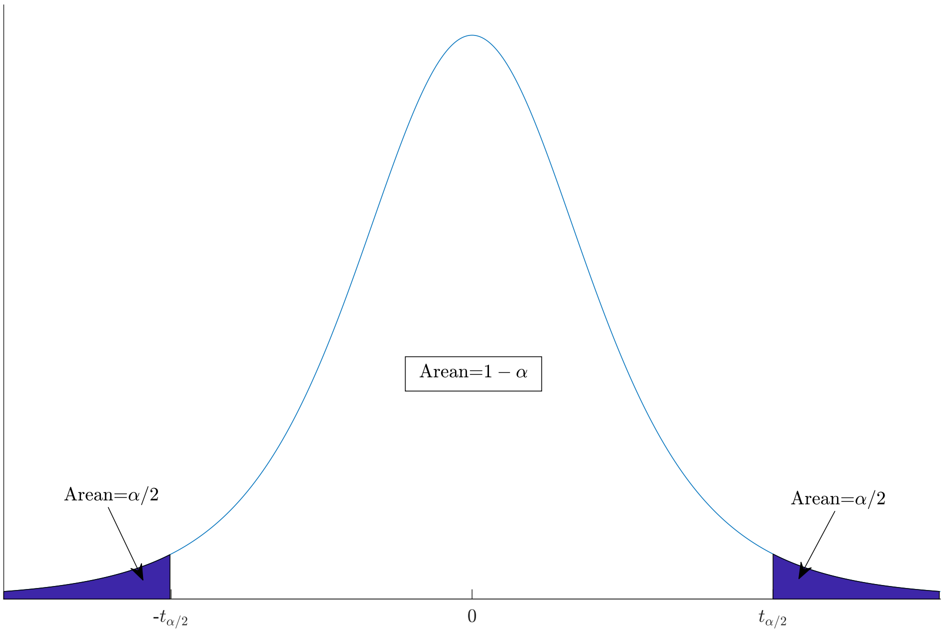


Figur 2: Comparison between the density functions of the standard normal $N(0, 1)$ t-distributions with different degrees of freedom

3. If we now use t-quantiles instead of the quantiles of the normal distribution we get

$$I_\mu = \left(\bar{x} - t_{\alpha/2}(n-1)\cdot\frac{s}{\sqrt{n}}, \bar{x} + t_{\alpha/2}(n-1)\cdot\frac{s}{\sqrt{n}}\right) \qquad (1-\alpha)$$

Example 1 Do problem 1305 and 1306 from [2]. □



Figur 3: Illustration of the $\alpha/2$-quantile of the distribution $t(5)$ (Arean=the area)

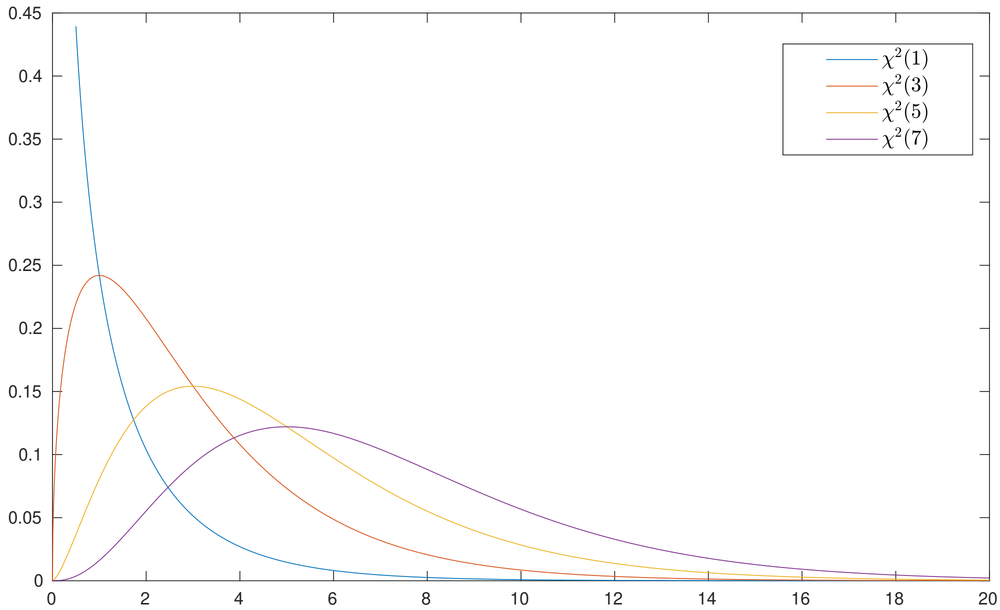
1.1.3 Confidence interval for σ

1. We estimate σ^2 by
$$(\sigma^2)^*_{obs} = s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2.$$

Remark 2 We will make a confidence interval for σ^2 and then we will take the square root of the limits and this will give us a confidence interval for σ.

2. The distribution of
$$\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2$$
where $(X_i-\mu)/\sigma \in N(0,1)$ and the X_i's are independent, is known as the χ^2-distribution with n degrees of freedom, in short $\chi^2(n)$.

Proposition 1 *The random variable*
$$\frac{n-1}{n-1}\cdot\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\bar{X})^2 = \frac{(n-1)S^2}{\sigma^2}$$
is χ^2-distributed with $n-1$ degrees of freedom and independent of $\bar{X}$.



Figur 4: The density function of the χ^2-distribution with different degrees of freedom.

3. With probability $1-\alpha$ it holds that

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

   Please note that we now have to give both quantiles explicitly, since the χ^2-distribution is not symmetric. Rewriting a bit we obtain that

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

   A confidence interval for σ^2 with confidence level1 $-\alpha$ is given by

$$I_{\sigma^2} = \left(\frac{(n-1)s^2}{\chi^2_{\alpha/2}}, \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}} \right) \qquad (1-\alpha)$$

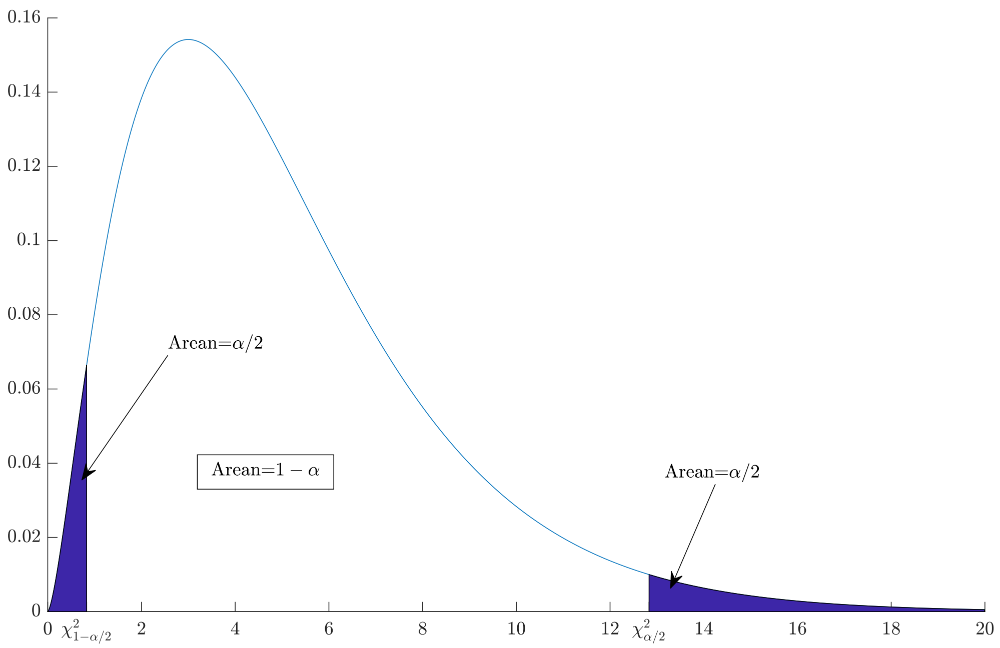
   and for σ by

$$I_{\sigma} = \left(\sqrt{\frac{(n-1)}{\chi^2_{\alpha/2}}} \cdot s, \sqrt{\frac{(n-1)}{\chi^2_{1-\alpha/2}}} \cdot s \right) \qquad (1-\alpha)$$

Example 2 Do problem 1308 from [2]. □

Referenser

[1] Blom, G., Enger, J., Englund, G., Grandell, J., och Holst, L., (2005). Sannolikhetsteori och statistikteori med tillämpningar.



Figur 5: Illustration of the quantiles of the distribution$\chi^2(5)$ (Arean=the area).

[2] Blom, Gunnar, (1989). Probability and Statistics. Theory and Applications.