BIE-PST – Probability and Statistics

Lecture 10: Interval estimation of parameters

Winter semester 2023/2024

Lecturer:
Francesco Dolce

Department of Applied Mathematics
Faculty of Information Technology
Czech Technical University in Prague

Table of contents

1 Interval estimation

1.1 Confidence intervals

Instead of a point estimator of a parameter θ we can be interested in an *interval*, in which the *true value* of the parameter lies with a *certain* large *probability* $1-\alpha$:

Definition 1.1. Let $X_1, \ldots, X_n$ be a random sample from a distribution with a parameter θ. The interval (L, U) with boundaries given by statistics $L \equiv L(\mathbf{X}) \equiv L(X_1, \ldots, X_n)$ and $U \equiv U(\mathbf{X}) \equiv U(X_1, \ldots, X_n)$ fulfilling

$$\mathrm{P}\left(L < \theta < U\right) = 1 - \alpha$$

is called the $100 \cdot (1-\alpha)\%$ *confidence interval* for θ.

Statistics L and U are called the *lower* and *upper* bound of the confidence interval.

The number $(1-\alpha)$ is called *confidence level*.

- It holds that
$$\mathrm{P}\left(\theta \in (L, U)\right) = 1 - \alpha.$$
- Which means that
$$\mathrm{P}\left(\theta \notin (L, U)\right) = \alpha.$$
- For a *symmetric* or *two-sided* interval we choose L and U such that
$$\mathrm{P}(\theta < L) = \frac{\alpha}{2} \quad \text{and} \quad \mathrm{P}(U < \theta) = \frac{\alpha}{2}.$$
- The most common values are $\alpha = 0.05$ and $\alpha = 0.01$, i.e., the ones that gives a 95% confidence interval or a 99% confidence interval.

If we are interested only in a lower or upper bound, we construct statistics L or U such that

$$\mathrm{P}\left(L < \theta\right) = 1 - \alpha \quad \text{or} \quad \mathrm{P}\left(\theta < U\right) = 1 - \alpha.$$

This means that

$$\mathrm{P}\left(\theta < L\right) = \alpha \quad \text{or} \quad \mathrm{P}\left(U < \theta\right) = \alpha,$$

and intervals $(L, +\infty)$ or $(-\infty, U)$ are called the upper or lower *confidence intervals*, respectively.

In this case we speak about *one-sided confidence intervals*.

There are several possible ways how to construct confidence intervals, depending on the underlying distribution and meaning of estimated parameters. We will use the following approach:

- Find a statistics $H(\theta)$, which:
  - depends on the random sample $X_1, \ldots, X_n$,
  - depends on the estimated parameter θ,

  - has a known distribution.

- Find such bounds h_L and h_U, for which

$$\mathrm{P}\left(h_L < H(\theta) < h_U\right) = 1 - \alpha.$$

- Rearrange the inequalities to separate θ and obtain

$$\mathrm{P}\left(L < \theta < U\right) = 1 - \alpha.$$

The statistics $H(\theta)$ is often chosen using the distribution of a point estimate of the parameter θ, i.e., sample mean for the expectation or sample variance for the theoretical variance.

1.2 Confidence intervals for the expectation

1.2.1 Known variance

Theorem 1.2. *Suppose we have a random sample $X_1, \dots, X_n$ from the normal distribution $\mathrm{N}(\mu, \sigma^2)$ and suppose that we know the value of σ^2. The two-sided symmetric $100 \cdot (1-\alpha)\%$ confidence interval for μ is*

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

where $z_{\alpha/2} = \Phi^{-1}(1-\alpha/2)$ is the critical value *of the standard normal distribution, i.e., such a number for which it holds that $\mathrm{P}(Z > z_{\alpha/2}) = \alpha/2$ for $Z \sim \mathrm{N}(0,1)$.*

The One-sided $100 \cdot (1-\alpha)\%$ confidence intervals for μ are then

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

using the same notation.

Proof. First we show that the sample mean of i.i.d. random variables with a normal distribution has a normal distribution, too, but with different parameters. The proof is obtained using the *moment generating function* $M_X(s) = \mathrm{E}[e^{sX}]$.

The moment generating function of the normal distribution with parameters μ and σ^2 is:

$$M_X(s) = \mathrm{E}[e^{sX}] = \int_{-\infty}^{+\infty} e^{sx} \cdot \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\,\mathrm{d}x = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2-2x\mu+\mu^2-2\sigma^2 sx}{2\sigma^2}}\,\mathrm{d}x$$

$$= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{\left(x-(\mu+\sigma^2 s)\right)^2 + mu^2 - (\mu+\sigma^2 s)^2}{2\sigma^2}}\,\mathrm{d}x$$

$$= e^{\mu s - \frac{\sigma^2 s^2}{2}} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{\left(x-(\mu+\sigma^2 s)\right)^2}{2\sigma^2}}\,\mathrm{d}x \underbrace{\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{\left(x-(\mu+\sigma^2 s)\right)^2}{2\sigma^2}}\,\mathrm{d}x}_{1} = e^{\mu s - \frac{\sigma^2 s^2}{2}}.$$

The moment generating function of a sum of independent random variables is the product of their generating functions.

The moment generating function of a sum of i.i.d. normal variables is:

$$\begin{aligned} M_{\text{sum}}(s) &= \mathrm{E}[e^{s\sum_{i=1}^n X_i}] = \mathrm{E}[e^{sX_1}\cdot\dots\cdot e^{sX_n}] \overset{\text{independence}}{=} \mathrm{E}[e^{sX_1}]\cdot\dots\cdot\mathrm{E}[e^{sX_n}] \\ &= \prod_{i=1}^n M_i(s) \overset{\text{identical distribution}}{=} (M(s))^n \\ &= \left(e^{\mu s - \frac{\sigma^2 s^2}{2}}\right)^n = e^{n\mu s - \frac{n\sigma^2 s^2}{2}}. \end{aligned}$$

Comparing with the moment generating function of one normal variable we see that the generating function of the sum corresponds with the normal distribution $\mathrm{N}(n\mu, n\sigma^2)$. Thus $\sum_{i=1}^n X_i \sim \mathrm{N}(n\mu, n\sigma^2)$ and therefore $\bar{X}_n \sim \mathrm{N}\left(\mu, \frac{n\sigma^2}{n^2}\right) = \mathrm{N}\left(\mu, \frac{\sigma^2}{n}\right)$.

Thus after *standardization* we have

$$Z = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \sim \mathrm{N}(0,1).$$

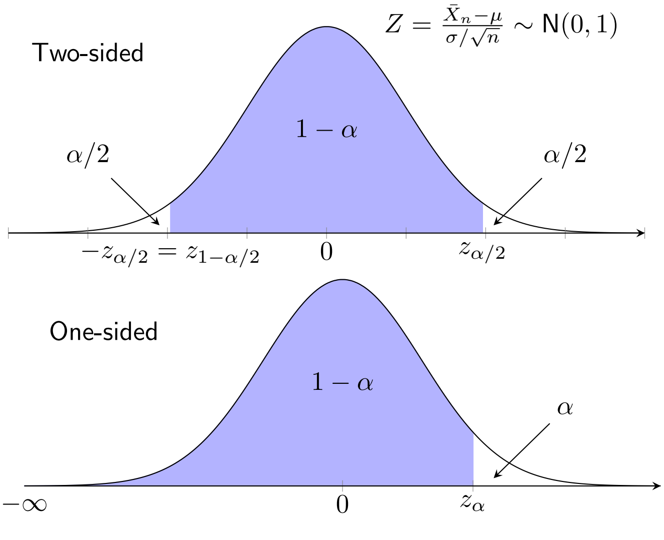
From the definition of the *critical value* $z_{\alpha/2}$: $\mathrm{P}(Z > z_{\alpha/2}) = \alpha/2$ it follows that $\mathrm{P}(Z < z_{\alpha/2}) = 1 - \mathrm{P}(Z > z_{\alpha/2}) = 1 - \alpha/2$. It means that

$$\mathrm{P}(z_{1-\alpha/2} < Z < z_{\alpha/2}) = \mathrm{P}(Z < z_{\alpha/2}) - \mathrm{P}(Z < z_{1-\alpha/2}) = 1 - \alpha/2 - (1 - 1 + \alpha/2) = 1 - \alpha.$$

From the symmetry of $\mathrm{N}(0,1)$ it follows that $z_{1-\alpha/2} = -z_{\alpha/2}$. And we have

$$\begin{aligned} 1-\alpha &= \mathrm{P}(z_{1-\alpha/2} < Z < z_{\alpha/2}) = \mathrm{P}\left(-z_{\alpha/2} < \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} < z_{\alpha/2}\right) \\ &= \mathrm{P}\left(-z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \bar{X}_n - \mu < z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = \mathrm{P}\left(z_{\alpha/2}\frac{\sigma}{\sqrt{n}} > \mu - \bar{X}_n > -z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) \\ &= \mathrm{P}\left(-z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu - \bar{X}_n < z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = \mathrm{P}\left(\bar{X}_n - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{X}_n + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right). \end{aligned}$$

□



To obtain the confidence interval for the expectation, we used the fact that for $X_i \sim \mathrm{N}(\mu, \sigma^2)$ the sample mean has the normal distribution:

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \sim \mathrm{N}(0,1).$$

The *central limit theorem* tells us that for any random sample with expectation μ and finite variance σ^2, the sample mean converges to the normal distribution with increasing sample size:

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{n\to\infty} \mathrm{N}(0,1).$$

This fact can be utilized to form confidence intervals also for other than normal distributions.

As a consequence of the *central limit theorem*, for large n we can use the same confidence intervals even for a random sample from *any distribution* with a finite variance:

Suppose we have a random sample $X_1, \ldots, X_n$ from a distribution with $\mathrm{E}\, X_i = \mu$ and $\operatorname{var} X_i = \sigma^2$, and suppose that we *know* the variance σ^2.

For n large enough, the *two-sided* $100 \cdot (1-\alpha)\%$ *confidence interval* for μ can be taken as

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

where $z_{\alpha/2}$ is the critical value of $\mathrm{N}(0,1)$. The one-sided confidence intervals are constructed analogously.

- The *approximate* confidence level of such intervals $\mathrm{P}\left(\mu \in (\cdots)\right)$ is then $1-\alpha$.
- *Large enough* usually means $n = 30$ or $n = 50$. For some distributions which are further away from the normal distribution (e.g., not unimodal, skewed), n must be even larger.

1.2.2 Unknown variance

Most often in practice we do not know the variance σ^2, but only have the observed data at our disposal.

As seen last time, the variance can be estimated using the sample variance

$$s_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2.$$

We will now show how to utilize the sample variance and adjust the intervals so that the confidence level would be exactly $1-\alpha$.

Chi-square and Student's t-distribution

We use the following new distributions:

Definition 1.3. Suppose we have a random sample $Y_1, \ldots, Y_n$ from the normal distribution $\mathrm{N}(0,1)$. Then we say that the random variable

$$Y = \sum_{i=1}^{n} Y_i^2$$

has the chi-square *(χ^2) distribution with n degrees of freedom.*

Definition 1.4. Suppose we have a random sample $Y_1, \ldots, Y_n$ from $\mathrm{N}(0,1)$, $Y = \sum_{i=1}^{n} Y_i^2$ and an independent variable Z also from $\mathrm{N}(0,1)$. Then we say that the random variable

$$T = \frac{Z}{\sqrt{Y/n}}$$

has the Student's t-*distribution with* n *degrees of freedom.*

The critical values for both distributions can be found in tables.

We estimate the unknown variance σ^2 using the sample variance

$$s_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2.$$

The distribution of the sample variance is connected with the chi-square distribution:

Theorem 1.5. *Suppose we have a random sample* $X_1, \ldots, X_n$ *from the normal distribution* $\mathrm{N}(\mu, \sigma^2)$*. Then*

$$\frac{(n-1)s_n^2}{\sigma^2}$$

has the chi-square *distribution with* $n-1$ *degrees of freedom.*

Proof. See literature. □

The distribution of the sample mean with σ replaced by $s_n = \sqrt{s_n^2}$ is connected with the t-distribution:

Theorem 1.6. *Suppose we have a random sample* $X_1, \ldots, X_n$ *from the normal distribution* $\mathrm{N}(\mu, \sigma^2)$*. Then*

$$T = \frac{\bar{X}_n - \mu}{s_n/\sqrt{n}}$$

has the Student's t-*distribution with* $n-1$ *degrees of freedom.*

Proof. We can rewrite T as:

$$T = \frac{\bar{X}_n - \mu}{\sqrt{s_n^2/n}} = \frac{\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{(n-1)s_n^2}{\sigma^2(n-1)}}}.$$

The numerator has standard normal distribution $\mathrm{N}(0,1)$, under the square root in the denominator we have χ^2_{n-1} divided by $(n-1)$. The distributions of $\bar{X}_n$ and s_n^2 are independent (see literature), thus the whole fraction has indeed the t_{n-1} distribution. □

Confidence intervals for the expectation

If the variance σ^2 is unknown we estimate the σ by taking the square root of the sample variance $s_n = \sqrt{s_n^2}$. Standardization of $\bar{X}_n$ with s_n leads to the *Student's t-distribution*:

Theorem 1.7. *Suppose we have a random sample $X_1, \dots, X_n$ from the normal distribution $\mathrm{N}(\mu, \sigma^2)$ with unknown variance. The two-sided symmetric $100 \cdot (1-\alpha)\%$ confidence interval for μ is*

$$\left(\bar{X}_n - t_{\alpha/2,n-1}\frac{s_n}{\sqrt{n}}\ ,\ \bar{X}_n + t_{\alpha/2,n-1}\frac{s_n}{\sqrt{n}}\right),$$

where $t_{\alpha/2,n-1}$ is the critical value *of the Student's t-distribution with $n-1$ degrees of freedom. The one-sided $100 \cdot (1-\alpha)\%$ confidence intervals for μ are*

$$\left(\bar{X}_n - t_{\alpha,n-1}\frac{s_n}{\sqrt{n}}\ ,\ +\infty\right) \quad \textit{and} \quad \left(-\infty\ ,\ \bar{X}_n + t_{\alpha,n-1}\frac{s_n}{\sqrt{n}}\right)$$

using the same notation.

As a consequence of the *central limit theorem*, for large n we can use the same confidence interval even for a random sample from *any distribution.*

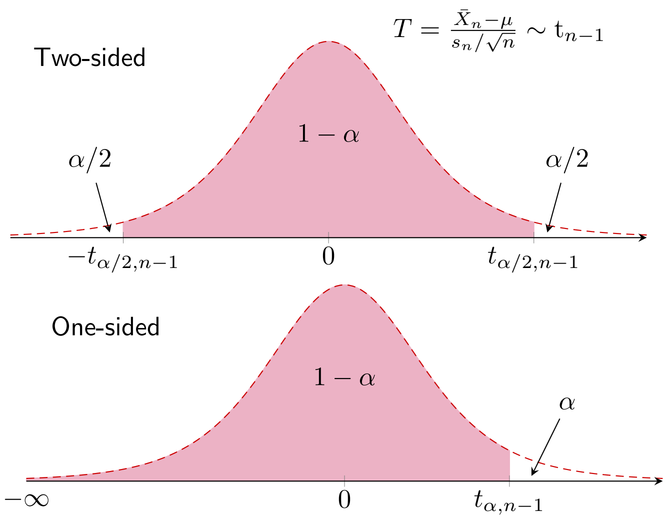
Suppose we observe a random sample $X_1, \dots, X_n$ from any distribution with $\mathrm{E}\, X_i = \mu$ and $\mathrm{var}\, X_i = \sigma^2$ and suppose that we *do not know* the variance σ^2.

For n large enough, the *two-sided* symmetric $100 \cdot (1-\alpha)\%$ *confidence interval* for μ can be taken as

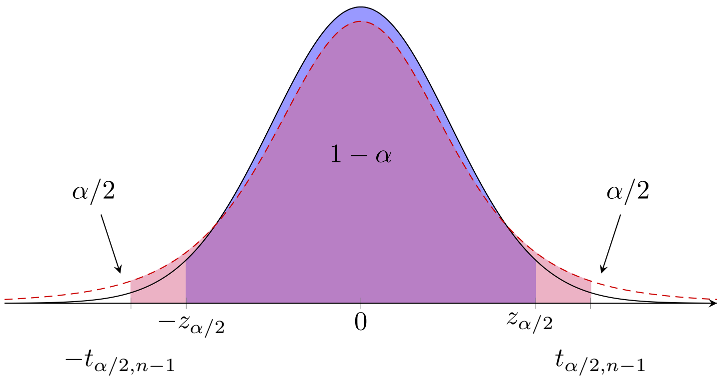
$$\left(\bar{X}_n - t_{\alpha/2,n-1}\frac{s_n}{\sqrt{n}}\ ,\ \bar{X}_n + t_{\alpha/2,n-1}\frac{s_n}{\sqrt{n}}\right),$$

where $t_{\alpha/2}$ is the critical value of the *Student's t-distribution with $n-1$ degrees of freedom* t_{n-1}. The one-sided confidence intervals are constructed analogously.

- For the interval it holds that $\mathrm{P}\left(\mu \in (\cdots)\right) \approx 1-\alpha$.
- *Large enough* usually means $n = 30$ or $n = 50$. For distributions which are further away from the normal distribution (e.g., not unimodal, skewed), n must be even larger.



Comparison of the critical values of $\mathrm{N}(0,1)$ and t_{n-1}:



- Confidence intervals for μ for unknown variance σ^2 are wider than for σ^2 known.
- For $n \to +\infty$ both distributions (and thus also their critical values) coincide.

Example 1.8 (– fishes' weights)**.** Suppose that the carps' weights in a certain pond in south Bohemia are random with normal distribution $\mathrm{N}(\mu, \sigma^2)$. From 10 previously caught carps we know that:

$$\sum_{i=1}^{10} X_i = 45.65 \text{ kg} \qquad \text{and} \qquad \sum_{i=1}^{10} X_i^2 = 208.70 \text{ kg}^2.$$

Find point estimates and two-sided 90% confidence interval estimates for μ and σ^2.
Point estimates:

- $\bar{X}_{10} = \dfrac{1}{10}\sum_{i=1}^{10} X_i = \dfrac{45.65}{10} = 4.565$ kg.
- $s_n^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2 = \dfrac{1}{n-1}\big(\sum_{i=1}^{n} X_i^2 - n(\bar{X}_n)^2\big).$
- $s_{10}^2 = \dfrac{208.7 - 10\cdot(4.565)^2}{9} = 0.0342$ kg^2.

Find the two-sided 90% confidence interval for μ:

$$\left(\bar{X}_n - t_{\alpha/2,n-1}\frac{s_n}{\sqrt{n}} \;,\; \bar{X}_n + t_{\alpha/2,n-1}\frac{s_n}{\sqrt{n}}\right)$$

$$\left(4.565 - 1.833\frac{\sqrt{0.0342}}{\sqrt{10}} \;,\; 4.565 + 1.833\frac{\sqrt{0.0342}}{\sqrt{10}}\right)$$

$$\bar{X}_{10} = 4.565 \text{ kg}$$
$$s_{10}^2 = 0.0342 \text{ kg}^2$$
$$\alpha = 10\% = 0.1$$
$$t_{0.05,9} = 1.833$$

Two-sided 90% confidence interval for μ is

$$(4.4578 \;,\; 4.6722) \text{ kg}.$$

Find the lower 90% confidence interval for μ:

$$\left(-\infty \;,\; \bar{X}_n + t_{\alpha,n-1}\frac{s_n}{\sqrt{n}}\right)$$

$$\left(-\infty \;,\; 4.565 + 1.383\frac{\sqrt{0.0342}}{\sqrt{10}}\right)$$

$$\bar{X}_{10} = 4.565 \text{ kg} \qquad t_{0.1,9} = 1.383$$
$$s_{10}^2 = 0.0342 \text{ kg}^2$$
$$\alpha = 10\% = 0.1$$

The lower 90% confidence interval for μ is then

$$(-\infty \ , \ 4.646) \ \text{kg}.$$

If the fish seller would tell us that the expected weight is 4.8 kg, we can say with 90% certainty that it is not true.

Such considerations form the basis of *hypothesis testing*, see later.

1.3 Confidence intervals for the variance

Theorem 1.9. *Suppose we observe a random sample $X_1, \ldots, X_n$ from the normal distribution $\mathrm{N}(\mu, \sigma^2)$. The two-sided $100 \cdot (1-\alpha)\%$ confidence interval for σ^2 is*

$$\left(\frac{(n-1)s_n^2}{\chi^2_{\alpha/2,n-1}} \ , \ \frac{(n-1)s_n^2}{\chi^2_{1-\alpha/2,n-1}} \right),$$

where $\chi^2_{\alpha/2,n-1}$ is the critical value *of the χ^2 distribution with $n-1$ degrees of freedom, i.e., $\mathrm{P}(X > \chi^2_{\alpha/2,n-1}) = \alpha/2$ if $X \sim \chi^2_{n-1}$.*

The one-sided $100 \cdot (1-\alpha)\%$ confidence intervals for σ^2 are then

$$\left(\frac{(n-1)s_n^2}{\chi^2_{\alpha,n-1}} \ , \ +\infty \right) \quad \text{and} \quad \left(0 \ , \ \frac{(n-1)s_n^2}{\chi^2_{1-\alpha,n-1}} \right).$$

✓ *The statement holds only for the normal distribution!*

Proof. We know that

$$\frac{(n-1)s_n^2}{\sigma^2}$$

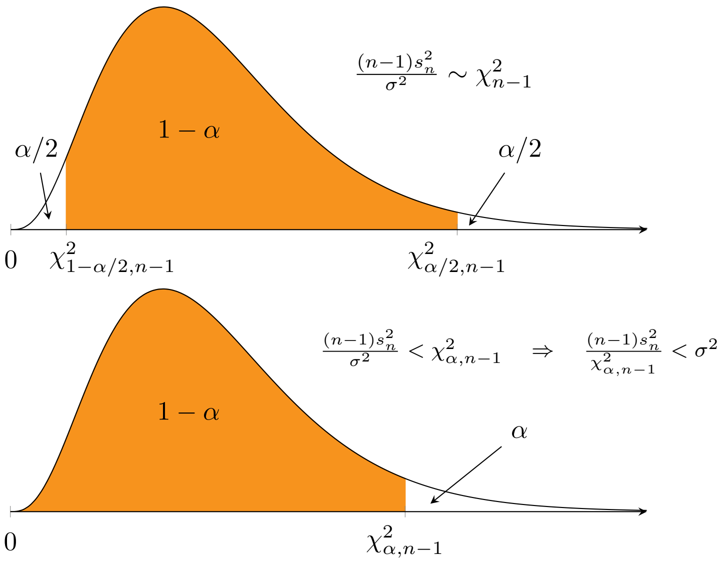
has the chi-square distribution χ^2_{n-1}. Then the confidence interval can be established using the critical values:

$$\mathrm{P}\left(\chi^2_{1-\alpha/2,n-1} < \frac{(n-1)s_n^2}{\sigma^2} < \chi^2_{\alpha/2,n-1} \right) = 1 - \alpha.$$

By multiplying all parts by σ^2 and dividing with the critical values we get that indeed:

$$\mathrm{P}\left(\frac{(n-1)s_n^2}{\chi^2_{\alpha/2,n-1}} < \sigma^2 < \frac{(n-1)s_n^2}{\chi^2_{1-\alpha/2,n-1}} \right) = 1 - \alpha.$$

□



Example 1.10 (– fishes' weights – continuation)**.** *Find the two-sided 90% confidence interval for the variance σ^2 of the carps' weights:*

$$\left(\frac{(n-1)s_n^2}{\chi^2_{\alpha/2,n-1}}\ ,\ \frac{(n-1)s_n^2}{\chi^2_{1-\alpha/2,n-1}}\right) \qquad s_{10}^2 = 0.0342 \text{ kg}^2$$

$$\alpha = 10\% = 0.1$$

$$\left(\frac{9 \cdot 0.0342}{16.919}\ ,\ \frac{9 \cdot 0.0342}{3.325}\right) \qquad \chi^2_{0.05,9} = 16.919$$

$$\chi^2_{0.95,9} = 3.325$$

The two-sided 90% confidence interval for σ^2 is

$$(0.0182\ ,\ 0.0926)\ \text{kg}^2.$$

Find the upper one-sided 90% confidence interval for the variance σ^2 of the carps' weights:

$$\left(\frac{(n-1)s_n^2}{\chi^2_{\alpha,n-1}}\ ,\ +\infty\right) \qquad s_{10}^2 = 0.0342 \text{ kg}^2$$

$$\alpha = 10\% = 0.1$$

$$\left(\frac{9 \cdot 0.0342}{14.684}\ ,\ +\infty\right) \qquad \chi^2_{0.1,9} = 14.684$$

The upper one-sided 90% confidence interval for σ^2 is then

$$(0.0210\ ,\ +\infty)\ \text{kg}^2.$$

If the fish seller would tell us that the variance of the weights is 0.01 kg^2, meaning that the standard deviation is 100 grams, we could say with 90% certainty that it is not true.

