Concepts of parameter estimation - confidence interval estimation
In confidence interval parameter estimation based on the sample $\mathbf{V} = (X_1, X_2, ..., X_n)$ a **confidence interval** $[l, u]$ is determined for which it is trusted with a **confidence coefficient** $(1 - \alpha)$ or a **risk coefficient** α that it contains the true value of the estimated parameter θ:

$$P(l \leq \theta \leq u) = 1 - \alpha\,. \tag{1}$$

Confidence intervals can be two-sided or left and right one-sided:

$$l \leq \theta \leq u \qquad \text{or} \qquad l \leq \theta \qquad \text{and} \qquad \theta \leq u\,. \tag{2}$$

Error of interval estimation is $|l - \theta|$ or $|u - \theta|$.

When the distribution of X_i in the sample is *normal with known variance* σ^2, the two-sided confidence interval on **mean** m of this distribution is:

$$\langle x\rangle - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < m < \langle x\rangle + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\,. \tag{3}$$

It is taken into account that the random variable

$$Z = \frac{\langle X\rangle - m}{\sigma/\sqrt{n}} \tag{4}$$

has a normal distribution. The value $z_{\alpha/2}$ is determined by $\Phi(z_{\alpha/2}) = (1 - \alpha)/2$.

When the distribution of X_i in the sample is *arbitrary with unknown variance* σ^2 and the *sample is large* ($n > 30$), the two-sided confidence interval on **mean** m of this distribution is:

$$\langle x\rangle - z_{\alpha/2}\frac{s}{\sqrt{n}} < m < \langle x\rangle + z_{\alpha/2}\frac{s}{\sqrt{n}}\,, \tag{5}$$

where the unknown variance σ^2 is replaced by the corrected sample variance S^2 which is a random variable. In the above equation its realisation s^2 is used which is determined from the sample. It is taken into account that the random variable

$$Z = \frac{\langle X\rangle - m}{S/\sqrt{n}} \tag{6}$$

has a normal distribution based on the central limit theorem for large n. The value $z_{\alpha/2}$ is determined as above.

When the distribution of X_i in the sample is *normal with unknown variance* σ^2 and the *sample is small* ($n < 30$), the two-sided confidence interval on **mean** m of this distribution is:

$$\langle x\rangle - t_{n-1;\alpha/2}\frac{s}{\sqrt{n}} < m < \langle x\rangle + t_{n-1;\alpha/2}\frac{s}{\sqrt{n}}\,, \tag{7}$$

where the unknown variance σ^2 is replaced by the corrected sample variance S^2. It is taken into account that the random variable

$$T = \frac{\langle X\rangle - m}{S/\sqrt{n}} \tag{8}$$

has a Student (or "t") distribution with $n - 1$ degrees of freedom. The value $t_{n-1;\alpha/2}$ is found in Table A.2.

When the distribution of X_i in the sample is *normal*, the two-sided confidence interval on **variance** σ^2 of this distribution is:

$$\frac{(n-1)s^2}{\chi^2_{n-1;\alpha/2}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{n-1;1-\alpha/2}}\,. \tag{9}$$

It is taken into account that the random variable

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2} \tag{10}$$

has a χ^2 distribution with $n-1$ degrees of freedom. The values $\chi^2_{n-1;\alpha/2}$ and $\chi^2_{n-1;1-\alpha/2}$ are found in Table A.3.

The approximate two-sided confidence interval on **proportion** p **of the population** having a *large sample* is:

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < p < \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\,. \tag{11}$$

Here it is assumed that the binomial distribution can be approximated by a normal distribution.

The two-sided confidence interval on **sum (difference) of the means** m_1 and m_2 of the *normally distributed* populations with *known variances* is:

$$\langle x_1\rangle \pm \langle x_2\rangle - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < m_1 \pm m_2 < \langle x_1\rangle \pm \langle x_2\rangle + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\,. \tag{12}$$

If the populations have *arbitrary distributions* with *unknown variances* and the *samples are large*, the above confidence interval is used with variances σ_1^2 and σ_2^2 replaced by sample variances s_1^2 and s_2^2.

The two-sided confidence interval on **sum (difference) of the means** m_1 and m_2 of the *normally distributed* populations with *unknown variances* and having *small samples* is:

$$\langle x_1\rangle \pm \langle x_2\rangle - t_{n_1+n_2-2;\alpha/2}s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} < m_1 \pm m_2 < \langle x_1\rangle \pm \langle x_2\rangle + t_{n_1+n_2-2;\alpha/2}s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\,, \tag{13}$$

where s_p is a realisation of the combined sample standard deviation S_p:

$$S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}\,. \tag{14}$$

The approximate two-sided confidence interval on **sum (difference) of the proportions** p_1 **and** p_2 **of the populations** when the *samples are large* is:

$$\hat{p_1} \pm \hat{p_2} - z_{\alpha/2}\sqrt{\frac{\hat{p_1}(1-\hat{p_1})}{n_1} + \frac{\hat{p_2}(1-\hat{p_2})}{n_2}} < p_1 \pm p_2 < \hat{p_1} \pm \hat{p_2} + z_{\alpha/2}\sqrt{\frac{\hat{p_1}(1-\hat{p_1})}{n_1} + \frac{\hat{p_2}(1-\hat{p_2})}{n_2}}\,. \tag{15}$$

Here it is assumed that the binomial distributions can be approximated by normal distributions.