Review - Confidence intervals

Interval estimation
Let $X_1, \dots, X_n$ be a random sample from F_θ, depending on an unknown $\theta \in \Theta$.
An **interval estimate** of θ is an interval with random bounds, convering the real value of θ with $1-\alpha$ probability.

- An interval estimate of θ of confidence level $1-\alpha$ is a pair of random variables (L_n, U_n), such that

$$\mathrm{P}_\theta(L_n < \theta < U_n) = 1 - \alpha \quad \text{for all } \theta \in \Theta.$$

- The random variable $L_n = L_n(X_1, \dots, X_n)$ is called the lower $1-\alpha$ confidence bound of the one-sided interval estimate of θ, if $\mathrm{P}_\theta(L_n < \theta) = 1-\alpha$ for all $\theta \in \Theta$.
- The random variable $U_n = U_n(X_1, \dots, X_n)$ is called the upper $1-\alpha$ confidence bound of the one-sided interval estimate of θ, if $\mathrm{P}_\theta(\theta < U_n) = 1-\alpha$ for all $\theta \in \Theta$.

The confidence $1-\alpha \in (0,1)$ is chosen as desired. Most common in practice is to take $\alpha = 0.05$.

General construction of the confidence interval for θ
We find a function h of both $X_1, \dots, X_n$ and θ, with a known distribution which does not depend on θ. Let $h_{\alpha/2}$ and $h_{1-\alpha/2}$ be the quantiles of this distribution. Then

$$P_\theta\left(h_{\alpha/2} < h(X_1, \dots, X_n; \theta) < h_{1-\alpha/2}\right) = 1 - \alpha \quad \text{for all } \theta \in \Theta.$$

Then we rearrange the inequalities to such a form, so that θ is in the middle between two bounds which do not depend on θ.

Confidence intervals for the expectation μ

- For $X_1, \dots, X_n$ i.i.d. from the normal distribution with known variance σ^2:

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

  where $\bar{X}_n$ is the sample mean and $z_{\alpha/2}$ is the $\alpha/2$ critical value of the standard normal distribution.
- For an uknown variance, replace σ^2 with the sample variance s_n^2 and $z_{\alpha/2}$ with the critical value of the Student's t-distribution with $n-1$ degrees of freedom $t_{\alpha/2,n-1}$.
- For one-sided intervals, replace one bound with $\pm\infty$ and in the other one use α instead of $\alpha/2$.
- It works approximatively also for other distributions because of CLT.

Confidence intervals for the variance σ^2

- For $X_1, \dots, X_n$ i.i.d. from the normal distribution:

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

  where $\bar{X}_n$ is the sample variance and $\chi^2_{\bullet,n-1}$ are critical values of the χ^2 distribution with $n-1$ degrees of freedom.
- For one-sided intervals, proceed as above.
- It does not work for other distributions.

Exercises 10 - Confidence intervals

1. We want to estimate the average length of a database server transaction (in milliseconds). Suppose that the transaction lengths $(X_1, \ldots, X_n)$ are independent and identically distributed random variables with a finite expectation $\mathrm{E}\, X_i = \mu$ and finite variance $\mathrm{var}\, X_i = \sigma^2$. We observed 50 transactions and computed:

$$\sum_{i=1}^{50} X_i = 684.2 \ [ms] \quad \text{and} \quad \sum_{i=1}^{50} X_i^2 = 18651.3 \ [ms^2].$$

a) Find point estimates of the mean μ and variance σ^2.

b) Find the two-sided and one-sided 99%-confidence intervals for μ.

c) Why do we need finite variance?

d) Why can't we construct confidence intervals for σ^2?

2. Suppose that a random variable Y is normally distributed. We want to estimate its expected value using a symmetric 95% confidence interval of width equal to 1. From previous measurements we observed that $\sigma^2 = 4$ and take this value as known and fixed.

a) Estimate how large a sample do we need.

b) How does the needed number of observations change, if we need the width of the interval to be 0.1?

3. Suppose we observe a random sample of $n = 16$ observed values from the normal distribution. The sample mean and sample variance are $\bar{X}_{16} = 10.3$ and $s_{16}^2 = 1.2$.

a) Find the two-sided interval estimate of the expected value with confidence level of 90%.

b) Find the two-sided interval estimate of the variance with confidence level of 90%.

4. Suppose we observe a random variable X with normal distribution with variance $\sigma^2 = 4$. We want to estimate the expected value using a 97.5% lower confidence interval with the distance between its right bound and the sample mean being 0.56. How large a random sample do we need?

Additional exercises - Confidence intervals

Interval estimates

5. Let $X_1, \ldots, X_n$ be the IQ scores of 8^{th} grade children. Suppose that $X_1, \ldots, X_n$ form a random sample from the normal distribution with an unknown expectation μ and known variance $\sigma^2 = 9$.

a) Find a point estimate of the expected IQ score. What is the distribution of this estimate?

b) Find a symmetric interval estimate of $(1-\alpha)\%$ confidence for the expected IQ score.

c) How would the interval change if we changed the confidence level?
How would the interval change if we added more children into the study, with the sample mean remaing the same?

d) After doing the measurement, we got

$$111,\ 116,\ 105,\ 111,\ 110,\ 114,\ 108,\ 106,\ 112,\ 108,\ 112,\ 111,\ 105,\ 111,\ 108,\ 110.$$

Use this values to estimate the typical IQ of 8^{th} graders. Find both point and 95% interval estimates.

e) Find the lower bound of the one-sided upper 95% confidence interval.

6. A poll was done by the MA&Č (Mikuláš, Anděl and Čert) Statistical Society to find out the percentage of Czech children who behaved well this year. There were $n = 400$ kids considered, of which 240 deserved to be rewarded with sweets and the others were to be punished with potatoes.

a) Estimate the proportion of Czech kids getting candies. What model do you use? What are the properties and the asymptotic distribution of this estimate?

b) Find the 95% asymptotic confidence interval for the percentage of children who will enjoy the sweets.

c) How many children must we have in the study, for the asymptotic 95% confidence interval to have a width of only 3%? Suppose that the proportion of well-behaved kids does not change.