

# Experimental analysis of algorithms.

## Lab 12. Confidence intervals

May, 15 2026

Most likely, for an estimator  $\hat{\theta}$  of a population parameter  $\theta$ , we have  $\hat{\theta} \neq \theta$  due to a sampling error. How much can we trust the reported estimator?

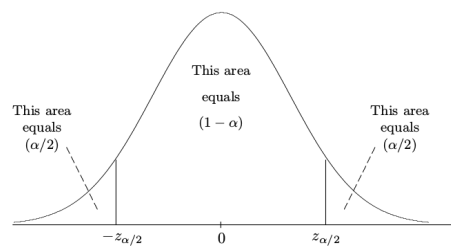
**Definition 1.** An interval  $[a, b]$  is a  $(1 - \alpha)100\%$  **confidence interval** for the parameter  $\theta$  if it contains the parameter with probability  $(1 - \alpha)$ ,  $P\{a \leq \theta \leq b\} = 1 - \alpha$ .

The coverage probability  $(1 - \alpha)$  is called a **confidence level**.

Given a sample of data and a desired confidence level  $(1 - \alpha)$ , how can we construct a confidence interval  $[a, b]$  that will satisfy the coverage condition  $P\{a \leq \theta \leq b\} = 1 - \alpha$ ?

Estimate the parameter  $\theta$ . Assume there is an unbiased estimator  $\hat{\theta}$  that has a Normal distribution. When we standardize it, we get a Standard Normal variable

$$Z = \frac{\hat{\theta} - E(\hat{\theta})}{\sigma(\hat{\theta})} = \frac{\hat{\theta} - \theta}{\sigma(\hat{\theta})}$$



This variable falls between the Standard Normal quantiles  $q_{\alpha/2}$  and  $q_{1-\alpha/2}$ , denoted by  $z_{-\alpha/2}$  and  $z_{\alpha/2}$ , with probability  $(1 - \alpha)$ .

If parameter  $\theta$  has an unbiased, Normally distributed estimator  $\hat{\theta}$ , then  $\hat{\theta} \pm z_{\alpha/2} \cdot \sigma(\theta) = [\hat{\theta} - z_{\alpha/2} \cdot \sigma(\theta), \hat{\theta} + z_{\alpha/2} \cdot \sigma(\theta)]$  is a  $(1 - \alpha)100\%$  confidence interval for  $\theta$ .

Deriving a **confidence interval for the mean  $\mu$**

1. When sampling from a normally distributed population, with **known  $\sigma$** :  
The  $(1 - \alpha)100\%$  confidence interval for  $\mu$  is given by:

$$\bar{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

2. If the population standard deviation  $\sigma$  **is not known**  
The confidence intervals for the population mean  $\mu$  using the  $t$  procedure:

$$\bar{X} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}$$

where  $s$  is the sample standard deviation,  $SE(\bar{X}) = \frac{s}{\sqrt{n}}$  the standard error of  $\bar{X}$  (the estimated standard deviation of the sampling distribution of  $\bar{X}$ ). The  $t$  value is found from the  $t$  distribution with  $n - 1$  degrees of freedom.

**Example:** Consider a sample of 100 individuals participating in a protest. The mean and the stdev of the age of the individuals from the sample are 25, respectively 5. In what range 95% of observations would fall?  
Obs: Critical values for confidence intervals: for 95%,  $z = \pm 1.96$ ; for 99%,  $z = \pm 2.58$ . CI:  $\mu \pm 1.96\sigma$ .

### Homework

1. Follow examples from the following sections: 9.2.2 Confidence interval for the population mean, 9.3.1 Large samples, 9.3.4 Small samples: Student's  $t$  distribution. Solve exercises 9.7a), 9.8a), 9.9a).
2. Compute the confidence interval for the mean of a dataset

**Example** (in python):

```
from scipy.stats import sem, t
from scipy import mean

confidence = 0.95
data = [10, 19, 11, 12, 15, 19, 9, 17, 1, 22, 9, 8]
n = len(data)
m = mean(data)
print("mean: ", m)

#calculates the standard error of the mean
std_err = sem(data)
```

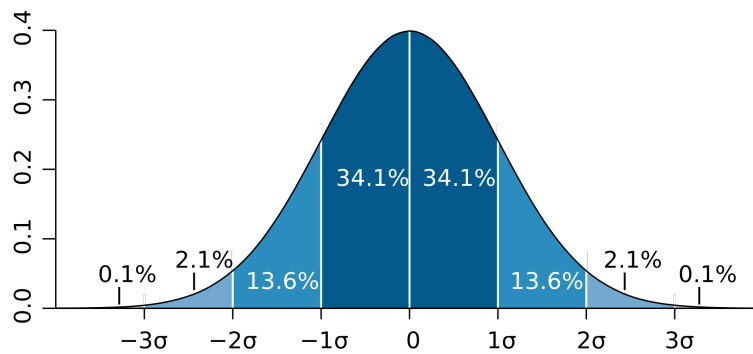
```

#t.ppf = a Student's T continuous random variable
h = std_err * t.ppf((1 + confidence) / 2, n - 1)
start = m - h
end = m + h
print("start: ", start, " end: ", end)

```

## Appendix

### Normal distribution



The **68-95-99.7 empirical rule**: 68% of values drawn from a normal distribution are within 1 stdev  $\sigma$  away from the mean.

**Standard normal distribution**  $\mu = 0, \sigma = 1$

### Bibliography

- *M. Baron. Probability and Statistics for Computer Scientists. Ch. 9. Statistical inference I. (9.2 Confidence intervals), 2013*