# Estimation basics

## ECDF intuition

a. Take any univariate continuous distribution that is readily available in R and plot its CDF (F).

b. Draw one sample (n = 1) from the chosen distribution and draw the ECDF (F\_n) of that one sample. Use the definition of the ECDF, not an existing function in R. Implementation hint: ECDF's are always piecewise constant - they only jump at the sampled values and by 1/n.

c. Repeat (b) for n = 5, 10, 100, 1000... Theory says that F\_n should converge to F. Can you observe that?

d. For n = 100 repeat the process m = 20 times and plot every F\_n^{(m)}. Theory says that F\_n will converge to F the slowest where F is close to 0.5 (where the variance is largest). Can you observe that?

## Show that the sample average is, as an estimator of the mean:

a. unbiased,

b. consistent,

c. asymptotically normal.

## Consistent but biased estimator.

a. Show that sample variance (the plug-in estimator of variance) is a biased estimator of variance.

b. Show that sample variance is a consistent estimator of variance.

c. Show that the estimator with (N-1) (Bessel correction) is unbiased.

Tukaj najdeĹˇ dovolj info:

https://stats.stackexchange.com/questions/174137/an-example-of-a-consistent-and-biased-estimator

<https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=2&ved=2ahUKEwjR7f-suMbmAhWMuIsKHSSkBqcQFjABegQIDBAE&url=https%3A%2F%2Fdawenl.github.io%2Ffiles%2Fmle_biased.pdf&usg=AOvVaw07wYsoQWIqXYQaBQKtuVJ4>

## Estimating the median

a. Show that the sample median is an unbiased estimator of the median for N(\mu, \sigma^2)

b. Show that the sample median is an unbiased estimator of variance for any distribution with symmetric density.

https://math.stackexchange.com/questions/119414/prove-that-the-sample-median-is-an-unbiased-estimator

# Bootstrap

## Ideally, a 1-\alpha CI would have 1-\alpha coverage. That is, say a 95% CI should, in the long run, contain the true value of the parameter 95% of the time. In practice, it is impossible to assess the coverage of our CI method, because we rarely know the true parameter. In simulation, however, we can. Let's assess the coverage of bootstrap percentile intervals.

a. Pick a univariate distribution with readly available mean and one that you can easuly sample from.

b. Draw n = 30 random samples from the chosen distribution and use the bootstrap (with large enough m) and percentile CI method to construct 95% CI. Repeat the process many times and count how many times the CI contains the true mean. That is, compute the actual coverage probability (don't forget to include the standard error of the coverage probability!). What can you observe?

c. Try one or two different distributions. What can you observe?

d. Repeat (b) and (c) using BCa intervals (R package boot). How does the coverage compare to percentile intervals?

# Maximum likelihood

##

a. Derive the maximum likelihood estimator of variance for N(\mu, \sigma^2).

b. Compare with results from (Consistent but biased estimator). What does that say about the MLE estimator?

https://www.youtube.com/watch?v=UUgEOSaydyw

## The German tank problem

During WWII the allied intelligence were faced with an important problem of estimating the total production of certain German tanks, such as the Panther. What turned out to be a successful approach was to estimate the maximum from the serial numbers of the small sample of captured or destroyed tanks.

(describe the statistical model used)

a. What assumptions were made by using the above model? Do you think they are reasonable assumptions in practiie?

b. Show that the plug-in estimate for the maximum (i.e. the maximum of the sample) is a biased estimator.

c. Derive the maximum likelihood estimate of the maximum.

d. Check that the following estimator is not biased: (it's at the end of the pdf).

https://engineering.purdue.edu/~ipollak/ece302/FALL09/notes/Classical\_Inference\_1\_Parameter\_Estimation.pdf