TMA4300 Computer Intensive Statistical Methods
Interactive lecture 5, Spring 2018

Problem A: The plug-in principle and bootstrap estimates for the standard error

In this problem we will consider two data sets:

- Data 1: `https://www.math.ntnu.no/emner/TMA4300/2018v/interactive/boot1.txt`
- Data 2: `https://www.math.ntnu.no/emner/TMA4300/2018v/interactive/boot2.txt`

Both the sets are simulated data sets. The values in data set 1 are sampled independently from a normal distribution, whereas the values in data set 2 are sampled independently from a skewed distribution.
For these data sets we will consider the following parameters:

- $\mu = t(F) = \mathrm{E}[X] = \int_{-\infty}^{\infty} x f(x) \mathrm{d}x$
- $\sigma = t(F) = \mathrm{SD}[X] = \sqrt{\mathrm{E}[X^2] - \mathrm{E}[X]^2} = \sqrt{\int_{-\infty}^{\infty} x^2 f(x) \mathrm{d}x - \left(\int_{-\infty}^{\infty} x f(x) \mathrm{d}x \right)^2}$
- $\theta = t(F) = \mathrm{E}\left[\left(\frac{X-\mu}{\sigma} \right)^3 \right] = \frac{\mathrm{E}[X^3] - 3\mu\sigma^2 - \mu^3}{\sigma^3}$ (skewness)

Note that to sample from the elements in a vector x you can in R use the function "sample". The default in this function is to put equal probabilities to each element and to sample without replacement. To sample with replacement you need to specify "replace=TRUE" in the function call.

1. Read in the two data sets and study the two data sets by making histograms with a reasonable number of bins.

2. Using the plug-in-principle, find expressions for the estimate for μ, $\widehat{\mu} = s(x)$. Compute the estimate for data set 1.

3. To get some understanding of the uncertainty of this estimate generate $B = 1000$ bootstrap samples $x^\star$, compute the corresponding bootstrap replications $\widehat{\mu}^\star = s(x^\star)$ and make a histogram of the simulated bootstrap replications.

4. Do items 2 and 3 also for the parameters σ and θ for data set 1. Is the skewness significantly different from zero? Do items 2 and 3 also for each of the three parameters for data set 2. Is the skewness of this data set significantly different from zero?

5. Write down a formula for the ideal bootstrap estimate of the standard error of $\widehat{\mu}$,

$$\mathrm{SD}_{\widehat{F}}\left[\widehat{\mu}^\star\right]$$

Hint: Start by writing a formula for $E_{\widehat{F}}\left[\widehat{\mu}^\star\right]$. Simplify the expression as much as possible. Can you compute the value analytically?

6. Use simulation to estimate the ideal bootstrap estimate of the standard error of $\widehat{\mu}$ for data set 1. Use $B = 200$ bootstrap samples and compare with the analytical value you found in 5.

7. Use simulation also to estimate the ideal bootstrap estimates of the standard error of the other two estimators for data set 1 (can you find the ideal bootstrap estimates also analytically?), and for all three estimators for data set 2.