Computational Statistics
Chapter 6: Bootstrap

1. Let $X_1, \dots, X_n$ be an iid sample from an exponential distribution with rate θ, i.e., the pdf of X_i is

$$f_\theta(x_i) = \theta \exp(-\theta x_i) \mathbb{1}_{[0,+\infty)}(x_i)$$

for $i = 1, \dots, n$. The purpose of this exercise is to estimate the median $m = F_\theta^{-1}(0.5)$ of this distribution using the bootstrap.

   (a) Generate a sample $x_1, \dots, x_n$ of size $n = 50$, with $\theta = 0.5$. What is the true value of m?

   (b) Write a function `bootstrap` that generates B bootstrap replicates of the observed data.

   (c) Compute the plug-in estimate $\widehat{m}$ of m and its standard error.

   (d) Compute a 95% confidence on m using the percentile method.

   (e) Compute a 95% confidence on m using the bootstrap-t method. (You will need to use a double bootstrap loop).

2. We consider again the data in the file `investment.txt`, which contains 15 yearly observations of U.S. investment data for the period 1968-1982. The variables are

   - `Year` = Date,
   - `GNP` = Nominal GNP,
   - `Invest` = Nominal Investment,
   - `CPI` = Consumer price index,
   - `Interest` = Interest rate,
   - `Inflation` = rate of inflation computed as the percentage change in the CPI.

   We consider the linear regression model $\mathbf{y} \sim \mathcal{N}_n(X\boldsymbol{\beta}, \sigma^2 I_n)$ with the dependent variable `Invest/(10*CPI)` and, as covariates, the time trend (a vector of integers from 1 to 15), `GNP/(10*CPI)`, `Interest` and `Inflation`.

(a) Compute the least-squares estimates of the regression coefficients as well as 95% confidence intervals based on the normal theory (use the function `confint`).

(b) Install the package `boot`. Using the functions `boot` and `boot.ci`, compute 95% bootstrap confidence intervals on the regression coefficients by case-based resampling. (Use the bootstrap t, percentile and BCa methods to construct the confidence intervals).

(c) Compute 95% bootstrap confidence intervals on the regression coefficients using the model-based approach (bootstrapping the residuals).

(d) Compare the different confidence intervals obtained and draw some conclusions.