Computational Statistics
Chapter 6: Bootstrap

1. Let $X_1, \ldots, X_n$ 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 $i = 1, \ldots, n$. 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 $x_1, \ldots, x_n$ 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 estimate its standard error.

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

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

   (f) Estimate the coverage probabilities of the confidence intervals computed in the previous two questions.

2. We consider 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 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 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).

3. We consider the wage-productivity data `wages.txt` from the book "Basic Econometrics" by D. N. Gujarati, McGraw-Hill, 4th edition, 2003. These data consist in indexes of real compensation per hour (Y) and output per hour (x) in the business sector of the U.S. economy for the period 1959 to 1998. The base of the indexes is 1992=100. We consider the following model:

$$Y_t = \beta_0 + \beta_1 x_t + \beta_2 x_t^2 + \epsilon_t. \tag{1}$$

(a) Plot the data.

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

(c) Plot the residuals. Using the function `dwtest` in the package `lmtest`, apply the Durbin-Watson test. What do you conclude?

(d) Using the function `tsboot` in the package `boot`, resample the residuals using the moving-block bootstrap method. Compute 95% bootstrap confidence intervals on the coefficients. Compare them to the conventional ones.