Group H - Assignment 3 Report

1. Exercise - theta estimator proposition and bootstrap

The bootstrap, technique relying on random sampling with replacement, was used to solve multiple tasks in this assignment. The method allows assigning a measure of the accuracy of the obtained result for mean, standard deviation etc. Bootstrap simply runs over a large number of iterations while resampling every time from the approximating distribution. In this exercise, the bootstrap is used to evaluate precision bias, standard error, and confidence level of a proposed estimator.

Dataset containing the population of 49 US cities in two different years was used in the exercise. The aim was to propose an estimator theta = E(X)/E(U) where X and U are populations in the year 1930 and 1920 retrospectively. Division of the two means of the data was proposed as theta and bootstrap was performed. The mean of the estimate was 1.241 with the average bias of 0.002 and the standard error of 0.036. The boundaries of 90% confidence interval are 1.188 and 1.302.

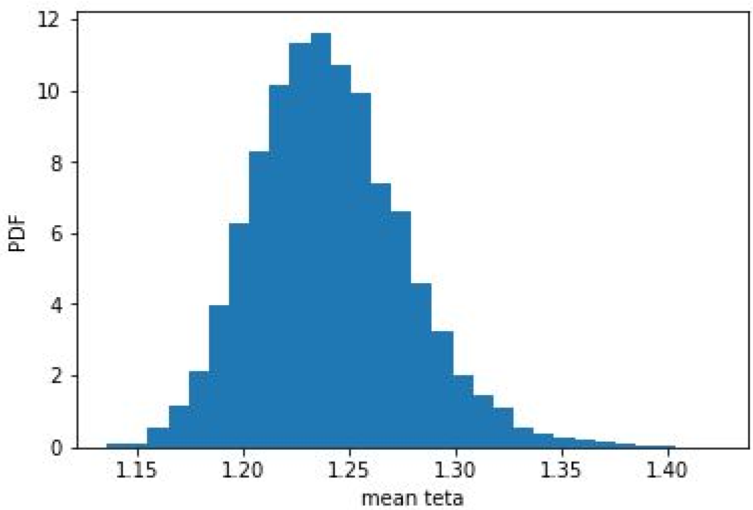

Exhibit 1 - Distribution of bootstrapped mean of proposed theta

2. Exercise - Weibull distribution

The main goal of the second exercise was to implement the Inverse Transform Method (ITM) to generate pseu-random values from a distribution when given its cumulative distribution function. The results are plotted in the Exhibit 2, similar results were achieved, only the right tail on higher values was not observed in the data generated by ITM.

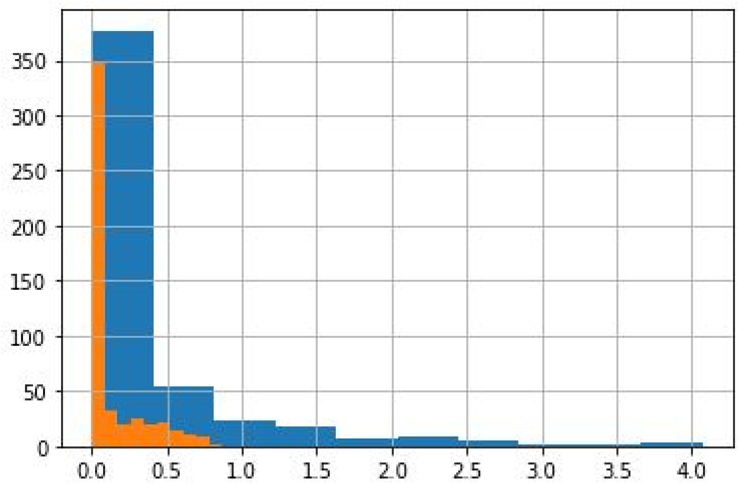

Exhibit 2 - Comparison of the implemented ITM generated values (orange) and the draw from the statistical package (blue)

3. Exercise - Expectation-maximization (EM) algorithm

The EM is an iterative way how to find maximum likelihood estimates of parameters in statistical models. In our case, the algorithm was used to distinguish between points from two different distribution and estimate the mean and deviation of the two distributions. The aim of the exercise was the implement the deviation prediction into the algorithm, which was achieved and Exhibit 3 shows that both of the distributions converged to the right parameters as they exactly fit both clusters of data.

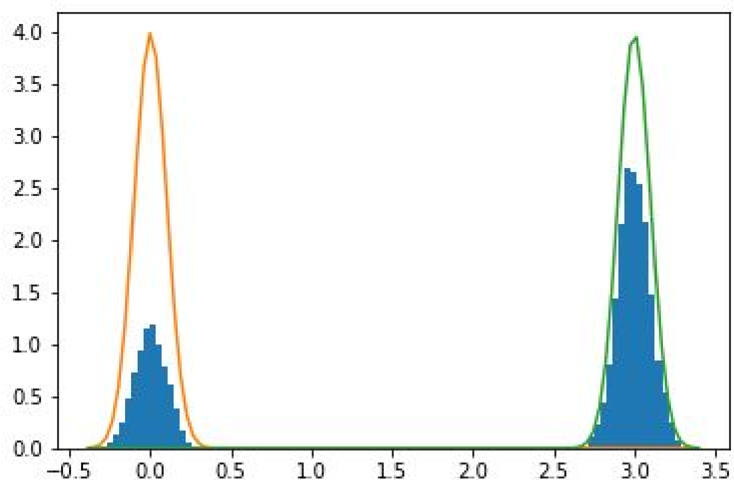

Exhibit 3 - The fit of the final iteration of all the parameter for the two distributions using EM

4. Exercise - X distribution

The aim of this exercise was to implement a given function and then later use it in Acceptance Rejection (AR) sampling algorithm in order to obtain randomized data. At the end, bootstrap statistics were required as the measure of the accuracy of the output once more. For the purpose of the AR, an exponential distribution was chosen as the envelope function. In the first step, a random value from the envelope function is selected along with a random number from 0 to 1. After multiplying the two values, we either accept the resulting number if it lies in the region under the function from which we want to sample or reject it if not.

The implemented function was used for AR with the observed acceptance ratio of 52.44% while using an exponential function with lambda equal to one as a parameter. Sample from the function was generated with a standard error of 0.302 as an estimate of the deviation with bounds of 90% bootstrapped confidence interval of 0.304 and 0.307.

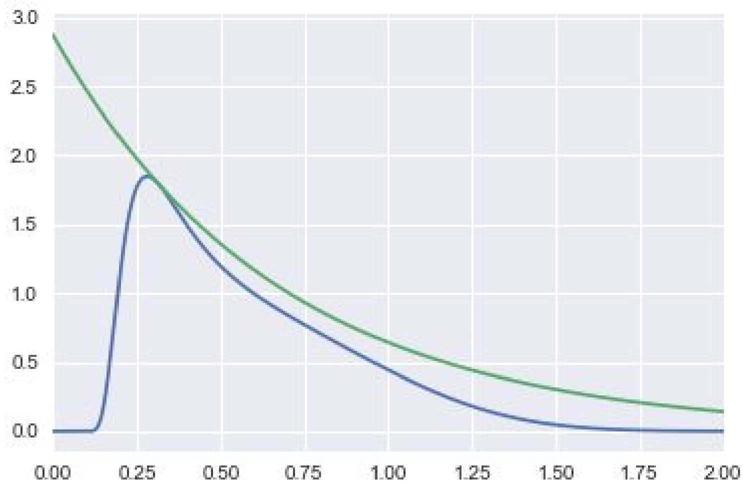

Exhibit 4 - Plotted envelope function (green) and the original function (blue) used for AR

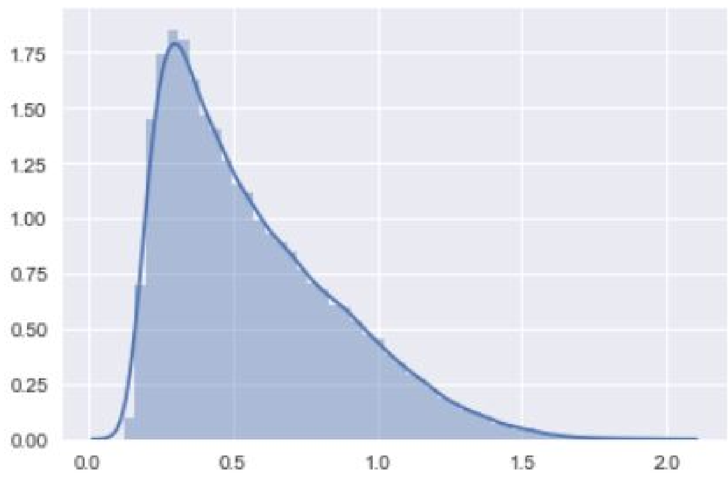

Exhibit 5 - The distribution of accepted draws from the AR algorithm for the given function

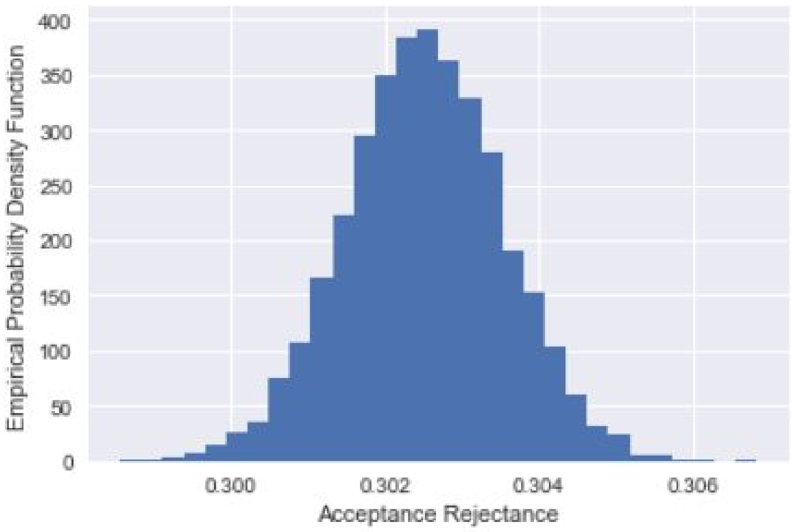

Exhibit 6 - Distribution of bootstrapped values of the standard error for implementation of AR