ST 495 Advanced computing for statistical methods
Homework problem set 10

April 16, 2024

No R packages are permitted for use in this assignment.

1. Using one or more of the various bootstrapping approaches presented in lecture for linear regression, construct and implement a simulation study to investigate the performance of bootstrapping the sampling distribution of an estimator of σ^2. In particular, generate synthetic data according to the model $Y_i = x_i'\beta + U_i$ for $i \in \{1, \ldots, n\}$, with $\beta \in \mathbb{R}^3$ and $U_1, \ldots, U_n \overset{\text{iid}}{\sim} \text{N}(0, \sigma^2)$. For your final analysis, include a histogram of the bootstrapped sampling distribution of $\widehat{\sigma^2}$ with the theoretical sampling density of $\widehat{\sigma^2}$ overlayed as a line. Recall that we performed such an analysis in lecture for the least squares coefficient estimates in the case of synthetic simple linear regression data.

2. Assume that $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{N}(\mu, \sigma^2)$, with σ^2 unknown. Construct a test statistic and a rejection region to test the hypothesis

$$H_0 : \mu = 0 \quad \text{versus} \quad H_1 : \mu < 0,$$

and investigate the power of the test by verifying that the empirical power of the test matches that from the theoretical distribution of the test statistic under various scenarios consistent with H_1 true. In particular, under H_1 true, for $\mu \in \{-2, -1.99, \ldots, -.01\}$ construct and implement a simulation study, and plot the empirical power of the test associated with each $\mu \in \{-2, -1.99, \ldots, -.01\}$ with the theoretical power overlayed as a line. Recall that we constructed and implemented such a simulation study during lecture, but for the case with σ^2 known.

3. Let $Y_1, \ldots, Y_n \overset{\text{iid}}{\sim} \exp(\lambda)$. If the value of y_n is missing, then construct and implement an EM algorithm to find the MLE of λ. Implement a simulation study to repeat this algorithm for a large number of data sets, and plot a histogram of the MLE of λ for each data set.