MA679 Lab3: More on Bootstrap

Why bootstrap?

Bootstrap allows estimation of the sampling distribution of **almost any statistic** using resampling technique.

Applications

- **standard error**
  On each of B bootstrap replicates:
  - Resample $X_i, i = 1, \dots, n$ with replacement from $X_1, \dots, X_n$
  - Compute the statistic of interest

- **confidence interval**
  Implementation in R: boot.ci()
  - t^*: bootstrap estimate
  - t_0: estimate of the statistic using the original full dataset
  - se^*: standard error of bootstrap estimate
  - b: bias, $b = t^* - t_0$
  - α: level of significance
  - z_α: 1- α quantile of the standard normal distribution
  - θ_α: α^{th} percentile of distribution of bootstrap realizations

  Percentile CI
  $$(\theta_{(1-\alpha)/2}, \theta_{1-(1-\alpha)/2})$$
  Normal CI
  $$(t_0 - b - z_{\alpha/2} \cdot se^*, t_0 - b + z_{\alpha/2} \cdot se^*)$$
  Basic CI
  $$(2t_0 - \theta_{1-\frac{1-\alpha}{2}}, 2t_0 - \theta_{\frac{1-\alpha}{2}})$$

- **Regression: resampling residuals**
  1. Fit the model, obtain fitted values $\hat{y}_i$ and residuals $\hat{\epsilon}_i = y_i - \hat{y}_i, i = 1, \dots, n$
  2. On each of B bootstrap replicates:
     a. Resample $\hat{\epsilon}_i^*, i = 1, \dots, n$ with replacement from $\hat{\epsilon}_1, \dots, \hat{\epsilon}_n$
     b. Create synthetic response variable $y_i^* = y_i + \hat{\epsilon}_i^*, i = 1, \dots, n$
     c. Refit the model on $(x_i, y_i^*), i = 1, \dots, n$
     d. Retain the statistic or quantity of interest