

Distribution of regression parameters

Use `lm` in R for least squares estimation.

Question 1.1. Take $N = 1000$ and $p = 100$. Simulate a single dataset where \mathbf{y} consists of i.i.d. $N(0, 1)$ variables and each \mathbf{x}_i is a p -dimensional vector of i.i.d. $N(0, 1)$ variables. Then

$$Y = \beta_0 + \sum_{i=1}^p \beta_i X_i + \varepsilon$$

with all $\beta_i = 0$. Estimate β and plot the absolute values in decreasing order.

Question 1.2. Replicate the above simulation $M = 500$ (or perhaps just $M = 100$) times. Store for each replication the three numbers

$$\max_i |\beta_i|, \quad \beta^T \beta = \|\beta\|^2, \quad \sum_{i=1}^p |\beta_i|$$

and make a histogram of their distribution.

Question 1.3. Repeat the previous question with $p = 500$. What happens? Can you find the general picture of the behaviour of the three distributions when p changes?