Assignment 1 — 140.778 (Adv Stat Comp)
Due: 29 Nov — 2nd term, 2000–2001

The bootstrap in linear regression

Part A

Consider the model $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, $i = 1, \ldots, n$, where

- $x_i \sim$ iid U$(0, 10)$ or
  $x_i \sim$ iid N$(\mu = 5, \sigma^2 = 3)$
- $\varepsilon_i \sim$ iid N$(0, \sigma^2)$ or
  $\varepsilon_i \sim$ indep N$(0, \sigma^2 x_i)$ or
  $\varepsilon_i / \gamma \sim$ iid t$(\text{df} = 3)$ or
  $\varepsilon_i + 1/\lambda \sim$ iid exponential(λ)

Write a function (in R, S-plus, or your favorite such package) to simulate such data, where you allow the user to specify n, the models for the x_i and the ε_i and the unknown parameters β_0, β_1 and σ^2, γ or λ.

Part B

1. Download the comma-delimited file `data1.csv` from the course web page.
2. Fit the linear model $y = \beta_0 + \beta_1 x + \varepsilon$ where the ε are assumed to be iid N$(0, \sigma^2)$. Obtain, using the normal theory, estimates and estimated standard errors (SEs) of $\hat{\beta}_0$ and $\hat{\beta}_1$ and confidence intervals (CIs) for β_0 and β_1. Also get an estimate of the correlation between $\hat{\beta}_0$ and $\hat{\beta}_1$.
3. Use a parametric bootstrap (ie, Monte Carlo simulation), using the normal model, to get new SEs and CIs and estimated correlation.
4. Use a nonparametric bootstrap, resampling the residuals, to get another set of SEs, etc.
5. Use a nonparametric bootstrap, resampling the pairs (x, y), to get a final set of SEs, etc.
6. In at least one of 3–5, try to estimate the Monte Carlo error in the estimated SEs and CIs.
7. Discuss the results.

Part C

Use your simulation program from A to investigate the performance of the parametric bootstrap and the two nonparametric bootstraps as a function of the sample size n, the model for ε_i and perhaps the model for the x_i.

Notes:

1. You do not need to complete all aspects of this assignment in great detail. Use your judgement in deciding how much time you wish to devote to it (considering your other responsibilities, your interest, and your need for sleep).
2. In reporting your results, *do not* simply provide raw computer output. I expect to see the sort of report one might give to a consulting client, with any computer code that you feel you should include placed in an appendix.
3. Feel free to ask me for advice, and feel free to work together. The report, however, should be your own.

Useful R functions:
`summary(a <- lm(y~x))`, `sample`, `rnorm`, `rexp`, `rt`, `mean`, `sd`, `cor`, `for`