MATH 392 Problem Set 8

Warm-up

Using matrix notation and the hat matrix, find $Var(\hat{Y})$.

MLR Simulator

The core of this assignment is the creation of a MLR simulator that you will use to investigate the properties of the method. The model under investigation is the following.

$$Y = X\beta + \epsilon$$

Where Y and ϵ are vectors of length n, X is an $n \times 3$ design matrix (the first column is just ones) and β is a vector of length 3.

You're welcome to select any values that you like for the parameters and any distribution that you like for the x's (you may be want to check out the `rmvnorm()` function in the `mvtnorm` package). The core bit of code should look something like:

```
# set params
B0 <- ___
B1 <- ___
B2 <- ___
B <- c(B0, B1, B2)
sigma <- ___

# complete specification
n <- ___
X <- ___

# simulate ys (this part inside a for loop)
epsilon <- r____()
y <- ___
```

Part I. Sampling distributions

Use your simulator to create an MC approximation of the true sampling distribution of the estimates of β_1, $E(Y_s)$, and Y_s corresponding to a fixed new observation x_s. How do these empirical distributions compare to their analytical form in terms of center, shape, and spread?

Part II. A different model

Consider two variations on the model:

1. An alternate distribution of the X's of your choosing.
2. An alternate distribution for ϵ (though it still should be centered at 0).

Does the inference change for the statistics investigated in the previous exercise? For this part, you should be able to recycle much of your code from the previous part.

Part III. Variance/Covariance

Generate a scatterplot matrix involving all pairwise relationships between the three simulated regression coefficients in the original model. To be clear, this involves generating a good number fitted betas from your simulator and plotting them. Based on a visual assessment of these plots, please characterize the joint distribution of these statistics in terms of center, shape, spread, and correlation. Compare the empirical covariance matrix to the analytical form that we derived in class.