Take-home MATH 282A Fall 2017

- Take-home exam–open book but please work alone. Prove *all* your statements; all sub-problems have equal weight.
- Part I amounts to 70 points; Part II amounts to 35 points. [100 points is considered full credit!] MONDAY DEC 11 by 11am
- The exam is due ~~Wedneday Dec 6 by 1pm~~ (sharp!). Please leave it in a sealed envelope in the mailbox of Prof. Politis (APM 7th floor). `Keep a copy of your exam for your records in case it is misplaced!`

Part I: Work out the following problems from the book.
Ch. 4: Set 4b: ex. 2, 3; set 4c: ex. 1, 3; misc. set: ex. 4. [Note a mistake in the book: in set 4b, ex. 2 (a), the n in the denominator should be inside the square root.] Ch. 8: Set ~~8b~~ 8a: ex. 2, 3. [Hint for ~~8b~~ 8a, ex. 2: it is the ANOVA identity!]

Part II: Let $\epsilon_i \sim$i.i.d.$N(0, \sigma^2)$, and consider LS estimates under the two models:

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad \text{for} \quad i = 1, ..., n \tag{1}$$

$$Y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \epsilon_i, \quad \text{for} \quad i = 1, ..., n \tag{2}$$

(a) Focus on estimating β_1. Is $\hat{\beta}_1$ unbiased when you fit model (1) while model (2) is true? Vice versa, what is the problem of estimating β_1 by fitting model (2) using LS when model (1) is actually true?

(b) Assume the design points are such that $\bar{x} = n^{-1} \sum_{i=1}^{n} x_i = 0$ so that the design of model (1) is orthogonal. Re-parametrize model (2) to an equivalent quadratic regression with orthogonal design, i.e., let

$$Y_i = \beta_0 + \beta_1 x_i + \tilde{\beta}_2 \, p(x_i) + \epsilon_i, \quad \text{for} \quad i = 1, ..., n \tag{3}$$

where $p(x) = x^2 + c_1 x + c_0$ is a quadratic function with coefficients chosen such that the regression of model (3) has orthogonal design. Is there still a problem when estimating β_1 by fitting model (3) while model (1) is true?

(c) Assume the orthogonal design of part (b) but now consider the goal of prediction of a future response Y^* associated with a regressor value x^*. Is there a penalty to pay if one uses the (estimated) model (3) for prediction purposes while model (1) is true? [Hint: calculate the MSE of prediction.]

(d) Assume straight line model (1), and suppose that you are actually *designing the experiment*, i.e., picking the x_i for $i = 1, ..., n$, where the responses Y_i will be measured. Suppose that the x_i must lie in the interval $[-1, 1]$; assume $\bar{x} = 0$ and that n is an even number. How should you pick $x_1, \ldots, x_n$ in order to minimize the variance of $\hat{\beta}_1$?

(e) Is the optimal design found in part (d) still a good design if the true model is instead the quadratic regression (2)? [Hint: calculate $Var(\hat{\beta}_2)$.]