Stats 201B: Statistical Modeling and Learning
Problem Set 1

January 9, 2017

Due January 16 by 11:55pm, submitted on course website.

Unlike other problem sets, **you must complete this entirely on your own without discussing with others**. It is purely review material. It will be graded. Responses should be typeset in LaTeX, Rmarkdown, KnitR, or similar.

Notation

In problem sets we will maintain the notation used in class, in which the density function for random variables Z is given by $p(Z)$. Note that we use p instead of f, which may be more familiar to some of you. Furthermore, we use $p(\cdot)$ whether it is a probability density function (for continuous random variables) or a probability mass function (for discrete random variables). Finally, except where it is needed, we drop the subscript and assume that the density is for the random variable referenced in the parentheses – i.e. we write $f_X(X)$ as simply $p(X)$.

Random Variables

1. (2pt) Consider continuous random variable X with probability distribution $p(X)$. How is $\mathbb{E}[X]$ defined? (Give the definition, not the estimator you'd use given a sample).

2. (2pt) How is $\text{Var}(X)$ defined?

3. (2pt) Further suppose Y is a continuous variable as well, and you have joint density $p(X, Y)$. How is $\mathbb{E}[Y|X]$ defined? (Write it out in terms of an integral and density function).

4. (2pt) If X and Y were independent, what is the relationship between $p(X, Y)$, $p(X)$, and $p(Y)$? What does $\mathbb{E}[X|Y]$ reduced to when X and Y are independent (show why this is, writing out the definition of $\mathbb{E}[Y|X]$ first).

For the following questions, draw random variables X_1, X_2,..., X_N, all independently from common density $p(X)$.

5. (2pt) Suppose you have scalars, a, b, c. What is $\mathbb{E}[aX_1 + bX_2 + cX_3]$ equal to (in terms of $\mathbb{E}[X]$)? What is $\text{Var}[aX_1 + bX_2 + cX_3]$?

6. (3pt) Let $\overline{X} = \frac{1}{N}\sum_i^N X_i$. Is $\overline{X}$ unbiased for $\mathbb{E}[X]$? Prove it. (Do not just cite a theorem!)

7. (3pt) Derive the variance of $\overline{X}$. What happens to it as $N \to \infty$?

Matrix Algebra,OLS, and R Practice

Consider random variables $Y \in \mathbb{R}$ and $X \in \mathbb{R}^{\mathbb{P}}$, drawn from joint density $p(X, Y)$. You collect a sample of draws from this distribution, $\{(Y_1, X_1), ..., (Y_N, X_N)\}$.

Let $\mathbf{X}$ be a $N \times (1 + P)$ matrix, with row i equal to $[1\ X_i^\top]$ (i.e., there is an intercept and then a column for each "covariate"). Consider an OLS model, $Y = \mathbf{X}\beta + \epsilon$, where $E[\epsilon|X] = 0$ by assumption.

8. (5pt) Using matrix notation at each step, derive the ordinary least squares estimator for β:

$$\beta_{OLS} = \underset{\beta \in \mathbb{R}^{P+1}}{argmin} (\mathbf{Y} - \mathbf{X}\beta)^\top (\mathbf{Y} - \mathbf{X}\beta)$$

9. (4pt) Show R code that would achieve the following (there is no need to submit this code in a separate file; just include it in your problem set write-up using an environment such as `verbatim`):

   a. Construct a matrix X to represent $\mathbf{X}$ in the above, with $N = 100$, one column of ones, and two columns of randomly drawn numbers (from any distribution you like).

   b. Using $\beta = [1\ 2\ 3]^\top$, compute vector Y equal to $X\beta + \epsilon$, where ϵ is drawn from a standard normal distribution.

   c. Compute $\hat{\beta}_{OLS} = (X^\top X)^{-1}(X^\top Y)$.

   d. Compare the result to the coefficients obtained using `lm` with the data you have constructed.

10. (4pt) Show (analytically) the unbiasedness of $\hat{\beta}_{OLS}$ for β. (Hint: compute $\hat{\beta}_{OLS}$, but replacing Y with $\mathbf{X}\beta + \epsilon$).

11. (4pt) Compute the variance, $\mathbb{E}[(\hat{\beta}_{OLS} - \beta)(\hat{\beta}_{OLS} - \beta)^\top]$, again sticking with matrix notation. You may assume $\mathbb{E}[\epsilon\epsilon^\top|X] = \sigma^2 I_N$, where I_N is the $N \times N$ identity matrix.

12. (2pt) What meaning would you give to the matrix $\mathbb{E}[\epsilon\epsilon^\top|X]$? Give an intuitive explanation of what the assumption that this matrix equals $\sigma^2 I$ implies.