Pol Sci 200C - Problem Set 0

Chad Hazlett

Due April 8, 11:59pm

This problem set is designed to (a) help remind you of some essential material which you are expected to know; (b) help you to clarify whether you have the necessary background for the course; and (c) help me see what you know and don't know. Do what you can and get help as needed, but don't be worried about your grade; we will score this on evidence of effort. If you are entirely unclear on something even after seeking help or having it explained, I'd rather know that than see a copied answer... so try to learn what you need but if you just don't have the requisite background, please say so in your write-up.

Instructions:

- Responses should be typeset in something nice, like TeX or Rmarkdown.
- Submit your completed problem set as a single PDF via the course website. If you are not using RMarkdown or similar, please include a copy of your code in your write-up (e.g. using the `verbatim` environment). If in doubt about formatting issues, please check with the TA.

1. Probability Theory I: Events

The 1982 (original) movie *Blade Runner* (starring Harrison Ford) is set in a future world where there are robots that are designed to look and behave exactly like human beings. The only way to tell if a randomly selected individual (who appears to be human) is in fact a human being or a robot is to administer a test. In our version of the test, suppose that if an individual taking the test is in fact a robot, the test will report that the individual is a robot 95% of the time. If an individual taking the test *is not* a robot, the test will report that the individual *is* a robot 3% of the time. Based on the number of robots manufactured, we can estimate that about 2% of all individuals we would test are actually robots.

a) (4pts) What is the probability that a randomly selected individual given the test will be classified as a robot?

b) (4pts) Given that an individual is classified as a robot by the test, what is the probability that individual is actually a robot?

c) (4pts) Given that an individual is classified as a human by the test, what is the probability that individual is actually a robot?

2. Probability Theory II: Random Variables

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

[1] In problem sets we will maintain the notation used in class, in which $p(Z)$ is a density function for random variable Z. Note that we use p instead of f, and that we use this for both probability density functions and probability mass functions. Further, we drop the subscript and assume that the density is for the random variable referenced in the parentheses (i.e. $f_X(X)$ is simply $p(X)$)

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

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

d) (2pt) If X and Y were independent, what does $\mathbb{E}[X|Y]$ reduce to? (Show why this is, writing out the definition of $\mathbb{E}[X|Y]$ first).

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

e) (2pt) Suppose you have scalars, a, b, c. What is $\mathbb{E}[aX_1 + bX_2 + cX_3]$? What is $\text{Var}[aX_1 + bX_2 + cX_3]$?

f) (4pt) 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!)

g) (4pt) Derive the variance of $\overline{X}$. What happens to it as $N \to \infty$?

3. Matrix algebra and OLS

Some of this question involves linear/matrix algebra. You should work through this question, getting however much help you need. But we note that linear algebra is now classified as "very useful'' rather than required for the class. Still it is a good chance to learn/practice these things and that is worthwhile in its own right. That said, if something makes no sense to you at after getting help, I'd rather you let me know that so we have a sense of where people are.

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, $\{(Y_1, X_i), ..., (Y_N, X_N)\}$.

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 a model, $Y = \mathbf{X}\beta + \epsilon$, where we assume $\mathbb{E}[\epsilon|X] = 0$.

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

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

b) (5pt) 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` or a non-evaluated code chunk in Rmd, etc.):

   i. 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).

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

   iii. Compute $(X^\top X)^{-1}(X^\top Y)$. Use the `solve` function in `R`.

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

c) (5pt) Show the unbiasedness of $\hat{\beta}_{OLS}$ for β. (Hint: compute $\hat{\beta}_{OLS}$, but replacing Y with $\mathbf{X}\beta + \epsilon$).

d) (5pt) Compute the variance, with X taken as given, i.e. $\mathbb{V}[\hat{\beta}_{OLS}|\mathbf{X}]$, again sticking with matrix notation. You may assume $\mathbb{E}[\epsilon\epsilon^\top|\mathbf{X}] = \sigma^2 I_N$, where I_N is the $N \times N$ identity matrix.

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

4. Some Statistics Review

True or false? For credit, explain your choice **briefly**.

a) (3pt) If there is perfect collinearity, the OLS estimator will give biased and inconsistent estimates.

b) (3pt) A very large p-value for the estimated coefficient for an explanatory variable provides strong evidence that the variable has zero effect on the outcome.

c) (3pt) If an estimator is unbiased it is also consistent.

d) (3pt) You have a model $Y_i = X_i^\top \beta + \epsilon$, and you fit it by OLS. If the OLS residuals are uncorrelated with X, our estimate of β are unbiased.

5. Simulations of Key Properties of the Mean

Suppose $X_1 \sim N(5, 2)$ and $X_2 \sim exp(\lambda = 1)$ (where exp indicates the exponential distribution). In `R`, construct two vectors, `X1` with 10000 draws from the same distribution as X_1, and `X2` with 10000 draws from the same distribution as X_2.

We will take sub-samples from these two variables to evaluate the coverage probability of 95% confidence intervals using different types of data and different sample sizes.

(a)(3pt) Describe the distribution of `X1` and `X2` using your favorite graphical approach for looking at continuous distributions. Mark the true expectation on your plot using a line.

(b)(4pts) Consider for a moment the random variables $\overline{X}_1$ and $\overline{X}_2$, representing sample means you could get from taking the average of 8 randomly sampled values of X_1 or X_2 respectively. Using math (not `R`), give solutions for:

- $\mathbb{E}[\overline{X}_1]$?
- $\mathbb{E}[\overline{X}_2]$?
- $Var(\overline{X}_1)$?
- $Var(\overline{X}_2)$?

(c)(5pts). Now, for X_1, draw 5000 samples of size N=6. Get the mean and compute the 95% confidence interval each time. What portion of the confidence intervals you computed include the true expectation? Repeat this for X_2.

(d)(5pt) Repeat (c) for samples of size 6, 20, 50, and 500. Report the coverage probability for each of your eight simulations in a table. How do your results change? What differences do you see between `X1` and `X2`?

(e)(2pt) Explain your findings in parts (c) and (d).