Economics 241A Professor Startz

Fall 2017 Class Assignments

A student will be chosen in class to present each of the assignments below. That means you need to come to class prepared to teach the answer. Bring a thumb drive with a PowerPoint or Beamer presentation. Many of the assignments call for a Matlab program. You need to explain why your program is written the way it is, not just show the answer. The focus is on *teaching* the answer.

Lecture 1:

Assignment before lecture Introduction

Consider the problem of a farmer who plants seeds using labor L and grows a crop of size X. Suppose the production function is

$$X = \alpha L$$

and that the farmer's utility is

$$U(X, L) = X^{\beta} + (\bar{L} - L)$$

where $\bar{L} - L$ is leisure and $\alpha > 0, 0 < \beta < 1$. How many hours should the farmer put in and how does the marginal productivity of labor effect the decision? Write a Matlab function that plots the optimal labor supply against the marginal product of labor.

Lecture 2:

1. Suppose that there are four types of potential research assistants. There are research assistants who are productive and who make good comments in class; there are research assistants who are productive but sit quietly; there are unproductive research assistants who do make good comments and finally there are unproductive research assistants who don't talk in class. The probabilities for the four kinds are

Type	probability
Productive, talks	.4
Productive, quiet	.2
Useless, talks	.1
Useless, quiet	.3

Now suppose that the value of a productive RA is 1 and the value of an unproductive RA is zero and that the cost of an RA is 0.7 and that you want to maximize

$$V = P(\text{Prod}) \times (1 - .7) + P(\text{Unprod}) \times (0 - .7)$$

A. Should you hire an RA if all you know is the unconditional distribution?
B. Suppose you now observe whether a student talks in class

2. Consider a health insurance plan. If a person is sick, the plan will have to spend C. A healthy person costs nothing. The probability that a young person will be sick is p_y; for an old person p_O, where $p_o > p_y$. A fraction f of the population is young. People voluntarily buy insurance if the premium is equal to or below their expected cost of illness. The insurance company prices the plan so as to break even.
   (a) Suppose participation is mandatory. What will be the premium?
   (b) Suppose participation is voluntary. What will be the premium?

3. Consider a test statistic that rejects a true null hypothesis a fraction $p = 0.05$ of the time. Suppose we have $n = 20$ independent applications of this test. Draw a plot of the pdf of the number of rejections (an application of the binomial distribution). What is the probability of more than 1 rejection?

Lecture 3:

(1) Write a Matlab program that plots two figures (one figure for small n and one for a large n), each comparing two cdfs. For both figures, one cdf is the number of heads out of n coin flips with a probability of heads equals to $1/2$. The other cdf is for a normal distribution with mean $1/2$ and standard deviation $\sqrt{n/4}$. The Matlab function binocdf will give you the former and normcdf will give you the latter.

(2) Suppose we draw n independent random variables $x_i \sim N(0,1)$ and generate the average $\bar{x}$. We know that $\Pr\left(\left|\bar{x}/\sqrt{1/n}\right| > 1.96\right) = 0.05$. Suppose we were to generate m such samples and count up the number of times that $\left|\bar{x}/\sqrt{1/n}\right| > 1.96$. Call this count divided by m, α.

   a. What is the standard deviation of α?
   b. Suppose we wanted there to be a 90 percent chance that the estimated α is between 0.04 and 0.06. How large must m be? You may approximate the distribution of α by the normal distribution.
   c. Write a simulation program that sets $n = 10$ and uses a variety of values of m, some smaller than your value given in part (b) and some larger. Then run this for a number of times saving α. Make a plot that shows the theoretical standard deviation of α (from part (a))as a function of m as well as the empirical standard deviation of α from the simulations. Effectively, do a Monte Carlo of doing a Monte Carlo.

(3) Write a Matlab program that draws a simulated $\chi^2(4)$ density and an analytic density on the same figure. To simulate the χ^2 generate 4 standard normals (randn(4,1)), square them and add them up. Do this a lot of times and then draw the histogram using the Matlab function histogram. Then draw the analytic expression using the chi2pdf function with 4 degrees of freedom.

(4) The pdf of the extreme value distribution can be written $f(x) = \frac{1}{\beta} e^{-\left(\left(\frac{x-\mu}{\beta}\right)+e^{-\left(\frac{x-\mu}{\beta}\right)}\right)}$, where $mean(x) = \mu + \beta\gamma, \gamma \approx 0.5772$ and $median(x) = \mu - \log(\log(2))$. Generate a bunch of samples of standard normals each with 120 observations. Find the maximum value for each sample. Find the mean and median of the distribution of the maxima. Now solve for the values of μ and β implied by the mean and median. Plot the empirical density (use the histogram function) of the maxima along with the theoretical pdf implied by the values of μ and β.

Lecture 4:

(1) Consider a Bernoulli random variable with possible outcomes $x_1 = 10, x_2 = 80, p(X = x_1) = .4$. Consider the function $g(x) = \log(x)$. Find $g(\mathrm{E}(x))$ and $\mathrm{E}\big(g(x)\big)$.

(2) One version of Chebychev's inequality states that if a distribution has $\mathrm{E}(x) = \mu$ and $\mathrm{var}(x) = \sigma^2$ and $d \geq 0$, then

$$\mathrm{P}(|\mathrm{x} - \mu| \geq \mathrm{d}) \leq \sigma^2/\mathrm{d}^2$$

Consider $x \sim N(0,1)$.

(a) Prove that the inequality holds for $d = 1$, giving a proof that takes no more than five seconds.

(b) Draw a graph in Matlab that plots both the bound from Chebychev's inequality and the actual value from the standard normal for interesting values of d.

Lecture 5:

(1) Generate 1,000 samples of 100 $\mathrm{iid\ exponential}(\lambda)$ variables with $\lambda = 2$, $\mathrm{f(x)} = \lambda \mathrm{e}^{-\lambda \mathrm{x}}, 0 < \mathrm{x} < \infty$.

Plot the distribution of the 25th, 50th, and 75th percentiles of your samples.

(2) Suppose you have a pair of honest dice. What is the probability of a natural (7 or 11)? What is the probability of craps (2, 3, or 12)?

Lecture 6:

Prove the following theorem

Theorem: If you want to minimize expected square error given the variable x, conditional expected value is right way to forecast. (Note that $\mathrm{E}(y|x)$ is a function of x.)

$$\mathrm{y^e} = \underset{\mathrm{y^e}}{\mathrm{argmin}}\, \mathrm{E}((\mathrm{y} - \mathrm{y^e})^2)$$

$$y^e = \mathrm{E}(y|x)$$

Lecture 7:

Suppose that employers observe a signal of productivity, mp_i, with error.

$$s_i = mp_i + \varepsilon_i$$

where marginal product and the error are joint normal, and uncorrelated, and the mean error is zero. So the error is pure noise. If employers observe the signal for individuals and pay expected marginal product, what does the wage schedule look like, and in particular what happens to an individual's wage if they increase their productivity by 1.0?

Lecture 8:

1. Try the following problem. Draw m iid standard normals, $x_i, i = 1, \dots, m$. Compute the statistic

$$T = \frac{1}{m}\sum_{i=1}^{m} I\big(x_i < \Phi^{-1}(.025) \cup x_i > \Phi^{-1}(.975)\big)$$

Now repeat this experiment n times and show the distribution of T. Do this for $m = 100$ and $n = 1000$. Show what you get empirically as well as what the theoretical answer should be. (Remember that the test statistic is essentially an average of Bernoulli trials.)

2. My daughter is going to run a survey in Nigeria. Enumerators are expensive, so each marginal response costs $c = 2500$ naira. A field experiment is planned. A baseline survey was run before any treatment, which indicated that the standard deviation of the variable of interest is $\sigma = 4$ and that the responses are approximately normally distributed. My daughter believes the true effect size is $\mu = 2$. Finding a positive estimated effect size will result in a publication which will have a NPV for her career equal to $50,000. (A reminder that sample means are roughly normally distributed.) How many surveys should my risk-neutral daughter buy?

Lecture 9:

(1) Generate $n = 4$ standard normals, compute the sample mean $\bar{x}$, sample variance s^2, and the t-statistic given by $t = \bar{x}/\sqrt{s^2/n}$. Do this a lot of times, saving the result. (While you're at it, use tic/toc to compute how long your simulation takes.) Now make three plots showing the empirical and theoretical pdfs, one for the sample mean, one for $(n-1)$ times the sample variance, and one for the t-statistic.

(2) Draw a sample of size $n = 2$ of uniform(0,1) independent random variables and compute the sample mean. Do this a lot of times and save the sample means. Plot the empirical distribution. Now do it again with $n = 1{,}000$. Does either distribution look familiar?

Lecture 10:

(1) The mean, $\bar{x}$, of n Bernoulli trials with $p = 1/2$ is distributed approximately $\bar{x} \overset{A}{\sim} N\left(p, \frac{p(1-p)}{n}\right)$. That suggests that 10 percent of the time the mean should turn out to be greater than $p + 1.2816 \times \sqrt{p(1-p)/n}$. Run a series of Monte Carlo simulations to find out how good an approximation this is for a variety of values of n.

(2) Suppose $x \sim U(0, u)$. Consider the function $g(\theta) = \log(\theta)$. It turns out that the first-order Taylor series approximation to the variance of a function of a random variable is

$$\left[\frac{dg(\theta)}{d\theta}_{|\mathrm{E}(\theta))}\right]^2 \mathrm{var}(\theta)$$

Use this expression to calculate and approximate variance of $\log(\theta)$ for $u = 1$ and for $u = 0.01$. Now redo the calculation by generating simulated $\theta's$ and finding the variance of $\log(\theta)$.

Lecture 12:

(1) Suppose $x_i \sim iidU[0, b]$. Write the joint pdf for x, being careful about edge conditions. Find the value of b that maximizes the joint probability. (Hint: this is a logic question rather than a calculation question.)

(2) Suppose $\bar{x}|\theta \sim N\left(\theta, \frac{\sigma^2}{n}\right)$—assume σ^2 and n are known constants and $\theta \sim U(-c, c)$. Find an expression for

$$f(\bar{x}) = \int_{-\infty}^{\infty} f(\bar{x}|\theta) f(\theta) d\theta$$

(3) Returning to question (1). Generate $n = \{10{,}100\}$ $iidU(0, b = 1)$ random variables and find the maximum likelihood estimator of b. Do this many times for each sample size and report on the distribution of b_{mle}. In particular, is it unbiased?

Lecture 15:

(1) Write a program that computes the power function for our standard mean of iid normal variables, assuming the null hypothesis is $\mu_0 = 0$ and, that $\sigma^2 = 1$, that the sample is size n, and we do a 5 percent test.(Do not use the Matlab function sampsizepwr; it makes it too easy.) So we need to compute

$$\beta(\mu) = P\left(|z > 1.96|\middle| z \sim N\left(\mu, \frac{\sigma^2}{n}\right)\right)$$

Graph the power function for $n = 20$ and $n = 120$. Suppose you are “accept” the null with 20 observations. Is it convincing that the true $\mu \neq 1$? How about $\mu \neq .25$? What would your answer be for 120 observations?

(2) Suppose $x_i \sim iid(\mu, \sigma^2), i = 1, \dots, n$ with σ^2 known. Show that the Wald, likelihood ratio, and Lagrange multiplier tests of

$$H_0: \mu = \mu_0$$

$$H_a: \mu \neq \mu_0$$

Are all identical, where

$$W = \frac{(\hat{\mu}_{mle} - \mu_0)^2}{I(\hat{\mu}_{mle})^{-1}}$$

$$LR = -2\left(\mathcal{L}^*(\mu_0) - \mathcal{L}(\hat{\mu}_{mle})\right)$$

$$LM = \frac{\partial \mathcal{L}(\mu)}{\partial \mu}\bigg|_{\mu_0}' [I(\mu_0)]^{-1} \frac{\partial \mathcal{L}(\mu)}{\partial \mu}\bigg|_{\mu_0}$$

(3) Income is distributed (very) roughly log-normal, once people with zero income are dropped from the sample. The file cpsMarch2016Income.mat contains wage and salary income from the March 2016 current population survey in the variable wsal_val and gender in the variable fe (fe=1 for women and fe=0 for men.)

You may “remember” (or may have looked it up), that the pdf of a log normal can be written

$$f(y|\mu, \sigma^2) = \frac{1}{y\sigma\sqrt{2\pi}} \exp\left\{-\frac{(\log y - \mu)^2}{2\sigma^2}\right\}$$

And that the mean is $\exp\left[\mu + \frac{\sigma^2}{2}\right]$, while the variance is $[\exp(\sigma^2) - 1] \exp(2\mu + \sigma^2)$.

We are interested in the hypothesis that the variance of the level of income is equal for men and women.

(a) Find the maximum likelihood values of μ and σ^2 and the variance separately for women and for men and then estimated jointly.

(b) Test that the variances are the same for women and men using a likelihood ratio test. You may cheat a little and act as if the hypothesis requires $\mu_F = \mu_M$ and $\sigma_F^2 = \sigma_M^2$, even though that is a little stronger than is required.

(c) Test the hypothesis that the variances are equal with a Wald test, without using the cheat part (b).

Lecture 16:

(1) Generate n values for X and ε. Then pick a value for β and generate $y = \log(\beta x) + \varepsilon$.

a. Plot the sum squared residuals for various values of β.

b. Define $dg = \frac{\partial X\beta}{\partial \beta}$. Pick an initial guess for $\beta, \beta^{(0)}$. Then regress $y - X^{\beta^{(j)}}$ on $dg^{(j)}$ calling the coefficient Δ. Update according to $\beta^{(j+1)} = \beta^{(j)} + \Delta$. Iterate for a while and draw a plot to show what happens.

(2) Compute an estimate of π by the following technique of Monte Carlo integration. Draw random $\{x, y\}$ pairs uniform between 0 and 1. Then decide if the point is inside or outside the unit circle. Since the fraction of points inside the unit circle corresponds to the area of a quarter unit circle, you can estimate π by four times the fraction inside the circle. Do this for a various numbers of random draws and plot the estimates of π against the number of draws.