Final Exam

EC 303: Empirical Economic Analysis

December 20, 2019

This exam is worth **100 points**, and is timed to take 90 minutes (you have **120 minutes** to complete it). You may use a two-sided, hand-written formula sheet (letter size) and a simple calculator. You **may not** use your cell phone or any other electronic device. If necessary, round your final answer (not the numbers in intermediate steps) to the third decimal place.

Please keep answers neat and organized. If you must separate parts of an answer, indicate this in the blue book so that a grader can clearly find your work. Any suspected academic misconduct will be reported to the Dean's office.

1 True/False Questions

(25 points total; 20 minutes). Answer each question true (T) or false (F) with *no more than* 3 sentences justifying your answers. Justifications may take the form of counterexamples, proof sketches, or intuition.

- Note that you get one point for free. Happy holidays!

1. (3 pts) In probability theory, we had to assume that all probabilities summed to 100% (rather than being able to derive it).
2. (3 pts) The Law of Large Numbers justifies both probability and estimation by asserting that the probability of recovering the true population parameter increases with sample size.
3. (3 pts) To find the marginal distribution of X from a joint probability distribution $f(x, y)$, we compute $\int_{-\infty}^{\infty} f(x, y)dx$ over the support of X.
4. (3 pts) The Method of Moments asserts that a random sample should approximate important population moments, particularly for large n.
5. (3 pts) The p-value is the probability of obtaining your data if the null hypothesis is true.
6. (3 pts) Conjugate priors are useful only to simplify mathematical calculations.
7. (3 pts) A high R^2 is strong indication of a linear relationship in a regression model.
8. (3 pts) To obtain unbiased estimators of β_0 and β_1 in a regression $y = \beta_0 + \beta_1 x + \epsilon$, we must assume that ϵ is normally distributed with mean 0 and variance σ^2.

2 Reviewing the Foundations

(30 points total; 30 minutes). This section contains 4 problems from Chapters 1 through 8. You may **choose two of them** to complete. Please clearly indicate which problems you have chosen in your blue book.

Problem 1: Probability. (15 points; 15 minutes). Prove that if A and B are independent events, then $\overline{A}$ and $\overline{B}$ are independent.

- *Hint:* A good way to start is by computing $P(\overline{A} \cap \overline{B})$ as $1 - P(A \cup B)$.

Problem 2: Moments. (15 minutes). Let X be a continuous random variable with a pdf given by

$$f(x) = \begin{cases} 0.15e^{-0.15x} & x \geq 0 \\ 0 & \text{Otherwise.} \end{cases}$$

a. (10 points). Find the moment generating function $M_X(t)$.

b. (5 points). Use your answer to (a) to calculate $\mathbb{E}[X]$ and $\mathbb{V}[X]$.

Problem 3: Sampling Distributions. (15 minutes). Let Y be a random variable indicating the number of job offers a person receives after graduating college. Suppose that no one receives more than 2 offers, so that Y has a probability mass function given by

y	0	1	2
$p(y)$	0.2	0.45	0.35

a. (10 points). Suppose you randomly sample $n = 2$ college graduates. What is the sampling distribution for $\overline{Y}$, the average number of job offers for your sample? Assume that job offers are independent across graduates.

b. (5 points). Based on your answer to (a), what is the probability that the average graduate receives more than one job offer?

Problem 4: Maximum Likelihood Estimation. (15 points; 15 minutes). Suppose that $\{w_1, w_2, ..., w_n\}$ is an i.i.d. random sample, where w_i follows a normal distribution, with an unknown mean μ, but a known standard deviation of one. Find the maximum likelihood estimator of μ.

- Recall that the normal pdf is given by

$$f(w; \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(w-\mu)^2}{2\sigma^2}}$$

3 Testing, Bayesian Inference, & Regression

(45 points total; 40 minutes). This section contains 3 problems from our current unit. You may **choose two of them** to complete. Please clearly indicate which problems you have chosen in your blue book.

- Again, you get one point for free!

Problem 5: Hypothesis Testing. (20 minutes). A government program subsidizes electricity installation costs in a developing country. Before the program was started, 15% of households had electricity installed. After the first year of the program, the government surveyed 26 households, and estimated that 27% had electricity installed.

a. (12 points). Perform a **one-sided** hypothesis test at the $\alpha = 0.05$ level of the null hypothesis that the new fraction of households with electricity is statistically different than the original estimate. Be sure to include all of the steps of a formal hypothesis test, including an interpretation of your results.

b. (5 points). How does your conclusion change if, instead, you were to perform a **two-sided** test? Which test would you want to perform in this scenario, and why?

c. (5 points). Do you think your results are economically meaningful? Provide a short discussion (of about 3 sentences).

Problem 6: Bayesian Inference. (20 minutes). We are interested in the average number of months a worker will be unemployed in a typical economy. That is, we let θ be the true average unemployment spell (in months) for our country.

a. (7 points). Suppose that our prior distribution is exponential: $f(\theta) = e^{-\theta}$, for θ defined over the interval $[0, \infty)$. Sketch a graph of the prior distribution and interpret it in context.

b. (9 points). We now sample n workers once they've been re-hired and ask each of them to report their unemployment spell x_i for $i = 1, 2, ..., n$. Suppose that each worker is independent, and the true data generating process for each x_i is given by a Poisson distribution with parameter $\lambda = 4$:

$$f(x_i; \lambda = 4) = \frac{4^{x_i} e^{-4}}{4!}$$

If $n = 1$, what should the posterior be? Which of the distributions does your posterior look closer to: the exponential or the Poisson?

- Don't worry about re-normalizing your posterior.

c. (6 points). What happens to the posterior as we sample many workers, all with the same data generating process? As $n \to \infty$, what should $\mathbb{E}[\theta]$ be?

- It may help you to plot the likelihood function for a few values of x (try integers less than 5) to recall its shape. Alternatively, you can calculate its expectation.

Problem 7: Multiple Regression. (20 minutes). This problem is loosely based on a recent working paper published by Ole Agersnap, Amalie Sofie Jensen, & Henrik Kleven. These researchers were interested in examining a sociological theory that suggests immigrants are drawn to countries with generous welfare policies (called "welfare magnet" countries). Specifically, these researchers estimated how decreasing the amount of welfare available to immigrant families affected the flow of immigration.

First, the researchers estimated the following simple regression, where y is the immigrant fraction of population and x is the monthly welfare benefits (measured in 100 USD) available to immigrants. Standard errors are shown in parentheses below:

$$\underset{}{y} = \underset{(0.002)}{0.05} + \underset{(0.0007)}{0.0044}x \tag{1}$$

a. (5 points). What does $\beta_0 = 0.05$ tell you in this regression? Do you think it is meaningful in this context?

b. (6 points). How much would welfare benefits have to be (according to this model) for immigrants to make up 10% of the population? Does this seem plausible?

- Make sure to watch your units.

The researchers then estimated a regression to examine what *type* of immigrants are attracted by these policies. They estimated this by including regressors for the amount of monthly benefits available to each type of immigrant visa (still 100s of USD). The y variable has not changed in this regression:

$$\begin{aligned} y = \underset{(0.002)}{0.05} &+ \underset{(0.0005)}{0.0028}(\text{Benefits for Asylum Seekers}) + \underset{(0.0005)}{0.0022}(\text{Benefits for Family Visas}) \\ &+ \underset{(0.0003)}{0.0004}(\text{Benefits for Work \& Study Visas}) \end{aligned} \tag{2}$$

c. (6 points). Which of these 4 regression coefficients is significant with 95% confidence?

d. (5 points). Which type of immigrant is most attracted by welfare benefits? Provide a short discussion of how you would use Equations (1) and (2) to respond to the statement: "Providing access to government welfare universally attracts more immigration."

4 References

- Agersnap, Jensen, & Kleven (2019). The welfare magnet hypothesis: Evidence from an immigrant welfare scheme in Denmark. *NBER Working Paper 26454* https://www.nber.org/papers/w26454.pdf.