MECO 7312: Advanced Statistics and Probability

Dr. Khai Chiong

Fall, 2024

E-mail: khai.chiong@utdallas.edu
Web: www.khaichiong.com/teaching
Office Hours: by appointment
Office: JSOM 13.324
Class Hours: Wednesday 1-3:45pm
Class Room: JSOM 13.501
Course modality: in-person

Course Description

The goal of this PhD-level course is to develop a rigorous theoretical foundation necessary for research in applied econometrics and statistics. It covers the core topics of probability theory and statistical inference, including properties of random variables and probability distributions, frequentist and bayesian estimation, asymptotic theory, and various methods of hypothesis testing.

Texbooks

The main required textbook is *Statistical Inference* by Casella and Berger.

- Casella, George, and Roger L. Berger. *Statistical inference*. 2nd Edition
- Davidson, Russell, and James G. MacKinnon. *Econometric theory and methods*.
- Cameron, A. Colin, and Pravin K. Trivedi. *Microeconometrics: methods and applications.*

Grades

Course grade is based on the weighted average of four problem sets and one final exam. Please work together in a team of 4 to 5 for the problem sets.

Programming and Statistical Computing

At the end of this course, students will be expected to be familiar with the following:

LaTeX

LaTeX is a program for typesetting mathematical notations. There are many online resources for LaTeX. To install LaTeX:, see: https://guides.nyu.edu/LaTeX/installation. You will also need a front-end text editor, I recommend TeXstudio.

You will be expected to typeset all your assignments using LaTeX. Exams are exempted and can be handwritten. There is an initial start-up cost and a learning curve, but it will pay off in the longer run.

Another common way to use LaTeX is Overleaf https://www.overleaf.com/. It provides an online, web-based editor for LaTeX and it compiles LaTeX on the cloud. It is helpful in a group setting where several people are working on the same document. It tracks changes and handles versioning automatically. Please see: https://www.overleaf.com/learn/latex/Learn_LaTeX_in_30_minutes

Python

I will be using Python for statistical programming and Monte Carlo simulations. Students are expected to learn and be familiar with Python.

Python is a general purpose programming language. How to get started:

- Python in Microsoft Visual Studio Code. https://code.visualstudio.com/docs/python/python-tutorial
- Anaconda. https://www.anaconda.com/products/individual
- Google Colab.

For this course, we will be writing Jupyter Notebooks to run and execute Python codes. You can either use Microsoft VS Code or Google Colab to write Jupyter Notebooks.

Google Colab is a free Jupyter Notebook environment that runs in the cloud and stores its notebooks on Google Drive. Since it online web-based, you do not have to install anything on your computer.

R

Another important programming language to learn is R. There are several ways to use R (these are all free and open-source):

- Install R locally (https://rstudio-education.github.io/hopr/starting.html), and use the RStudio application to write and run R (https://posit.co/downloads/).
- Change your runtime to R in Google Colab.
- Run R within Python via rpy2.

Mathematica

Occasionally, we will use Mathematica. It is useful for quickly visualizing functions and working through excessively tedious algebra. It can be installed for free as a UTD student: https://oit.utdallas.edu/howto/mathematica/

Schedule

The schedule is tentative and subject to change.

Week 1: Basic probability theory (Casella-Berger, chapter 1): sample spaces, event spaces, probability spaces, random variables, probability density functions, cumulative density functions.

Week 2: Properties of random variables: (Casella-Berger, chapters 2): transformations of random variables, probability integral transformation, expectations and higher moments.

Week 3: Multivariate random variables (random vectors) (Casella-Berger, chapter 4): joint and marginal distributions, conditional distributions, independence of random variables; covariance and correlation, bivariate transformations

Week 4: Common families of distributions (Casella-Berger, chapters 3): Multivariate Normal distribution, Gamma distribution, truncated random variables, concentration inequalities.

Week 5: Properties of a random sample (Casella-Berger, chapter 5): sampling distributions, order statistics, unbiasedness and consistency, convergence concepts (convergence in probability, convergence almost surely, convergence in distribution).

Week 6: Point estimation (Casella-Berger, chapter 5): asymptotics, central limit theorem, delta method, continuous mapping theorem, asymptotic variance.

Week 7: Point estimation (Casella-Berger, chapter 7): Method of Moments estimator, Generalized Method of Moments.

Week 8: (Point estimation (Casella-Berger, chapter 7): Maximum Likelihood estimator, loss functions, mean square error, Fisher's information, Probit models.

Week 9: (Casella-Berger, chapter 7. Cameron-Trivedi, Chapter 13): Bayesian methods versus Frequentist. Conjugate prior. Bayes Theorem. Bayesian estimation.

Week 10: (Casella-Berger, chapter 8): Hypothesis testing, Likelihood Ratio test, Wald's t-test, Type-1 and Type-2 errors.

Week 11: (Casella-Berger, chapter 8): Lagrange-multiplier test. Size and power of a test. p-values. Neyman-Pearson lemma. Asymptotic distribution of test statistics.

Week 12: (Casella-Berger, chapter 9): Confidence interval as inversion of test statistics. Coverage probabilities. Bayesian intervals.

Week 13: (Casella-Berger, appendix): Data-resampling and simulation techniques. Bootstrapping. Importance sampling. Monte Carlo sampling and integration. Non-parametric methods. Kernel estimators. k-nearest neighbors.

Week 14: (Davidson-Mackinnon, chapters 2-3): Statistical properties of linear regressions. Matrix notation. Bias and consistency. Frisch-Waugh-Lovell Theorem. The geometry of OLS estimation. Multicollinearity.

Week 15: (Davidson-Mackinnon, chapters 2-5): OLS covariance matrix. Hypothesis testing and inference involving OLS estimators. Heteroskedasticity consistent covariance matrix estimator. Serial correlation.

Optional topics, when time permits, may include brief overviews and introductions to the following: Causal inference methods. Average treatment effect. Propensity score matching. Difference-in-difference methods. Discrete-choice models. Probit and Logit models. Binary and multinomial models. Time-series and panel data. Testing for serial correlations. Random and fixed effects.

Related courses

The first part of this course (the *probability* part of this course) has some overlap with OPRE 7310 Probability and Stochastic Processes. The treatment of probability in OPRE 7310 is more rigorous and measure-theoretic.