SYLLABUS

This first half of the first-year PhD sequence on econometrics develops the basic ideas of data analysis, modeling, probability and inference, without assuming background on these topics. However, it assumes firm control of the mathematical tools of linear algebra and calculus and will require computational work with data. It will bring out connections of econometrics to recent developments in machine learning and Bayesian inference. Though the course will not follow a single textbook, the following books could be useful or interesting for those taking the course.

McElreath (2020): *Statistical Rethinking*. An introductory book that works hard to make Bayesian inference intuitive. The order in which it takes up topics and methods does not match what we will do in this course, but its overall perspective does match.

Gelman et al. (2014): *Bayesian Data Analysis*, 3rd edition. Assumes more background of the reader than McElreath, though it does have an introductory section. The strongest coverage of MCMC methods and their applications, though not with a primary emphasis on economic applications.

MacKay (2003): Links information theory, statistical inference, and machine learning, with interesting examples. Has some overlap with what we will do in the course. This book is available online as a legal downloadable pdf.

Casella and Berger (2002): *Statistical Inference*. Thorough and careful explanation of standard topics of a graduate-level statistics and probability course. Little attention to Bayesian approaches, none to machine learning.

Murphy (2013): *Machine Learning: A probabilistic perspective*. The first part of this book has substantial overlap with material in this course.

Hastie et al. (200): *The Elements of Statistical Learning*. Focused on machine learning, but with a probabilistic point of view. More advanced than what we will do in this course.

Imbens and Rubin (2015): *Causal Inference* and Angrist and Pischke (2010): *The Credibility Revolution. . . .* References for the "Causal" approach mentioned below.

1. Approaches to data analysis

Machine learning: Algorithms first. Finding the model, not being given a model.

Bayesian inference: Producing distributions for unknowns, rather than for estimators. Harnesses computer power to make some models easier.

Econometrics: Traditionally took models as given, produced by "theorists", to be "tested". But now being influenced by machine learning approaches.

"Causal" inference: Though the need for assumptions in mapping probability models to predictions about the consequences of actions was well understood at the Cowles

Date: August 30, 2020.

Foundation in the 1950's, it has received new emphasis, especially in the context of estimating binary "treatment effects" in heterogeneous populations.

Exercise: k-means

2. Probability: Definition, interpretations

The mathematical definition. Random variables. Expectation, conditional expectation, conditional probability, exchangeability, independence, Bayes' rule.

Interpretations: Describing a data set. Limiting frequencies in independent trials. Characterizing uncertainty for decision-making. "Market measure".

Reporting a distribution: Mode, mean, variance, quantiles, higher moments, HPD set plots, one and 2-D plots of marginals, conditionals.

Exercises: Conditioning in AK count data.

3. Decision Theory

Probability as a separating hyperplane. Analogy to prices.

Complete class theorem.

Exercise: Admissible umbrellas.

4. Continuous distributions

Integrals vs. sums. Probability density functions. Conditional probability paradoxes.

pdf's for functions of random variables — Jacobian terms.

Evaluating posterior expectations by Monte Carlo.

The Dirichlet distribution. Inference for the multinomial.

The Gamma and Poisson distributions.

Exercise: Error bands in the AK count data. Fitting Gamma to log wage distributions.

5. The multivariate normal distribution

Inference for normal mean and covariance matrix.

Student t, multivariate t, Wishart.

Where t vs. normal matters: Weitzman (2007)

The central limit theorem. Convergence for distributions, random variables.

Gaussian approximations to posterior pdf's. Bernstein-von Mises theorem.

Exercise: Limits of the CLT: city population data.

6. Inference based on pre-sample probability

Unbiased estimates. Confidence intervals and sets. Pivotal statistics. Tests.

Paradoxes from interpreting pre-sample probabilities as post-sample. Müller and Norets (2016): making confidence sets credible. The stopping rule paradox.

Exercise: Coverage probabilities for the AK count posterior credible sets. Confidence intervals for AK data based on CLT.

7. The standard normal linear model

Posterior means, credibility sets.

Unbiased estimates, confidence sets.

Model choice: variable list, non-linearities, interactions.

Heteroskedasticity.

Non-regression linear fits.

Exercise: Returns to schooling in the AK data.

References

Angrist, J. D. and J.-S. Pischke (2010): "The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking the Con out of Econometrics," *Journal of Economic Perspectives*, 24, 3–30.

Casella, G. and R. L. Berger (2002): *Statistical Inference*, Wadworth Group, second ed.

Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin (2014): *Bayesian Data Analysis*, CRC Press, third ed.

Hastie, T., R. Tsibshirani, and J. Friedman (200): *The Elements of Statistical Learning*, Springer, second ed.

Imbens, G. W. and D. B. Rubin (2015): *Causal Inference*, Cambridge University Press.

MacKay, D. J. C. (2003): *Information Theory, Inference, and Learning Algorithms*, Cambridge University Press.

McElreath, R. (2020): *Statistical Rethinking: A Bayesian Coures with Examples in R and Stan*, CRC Press, second ed.

Müller, U. K. and A. Norets (2016): "Credibility of Confidence Sets in Nonstandard Econometric Problems," *Econometrica*, 84, 2183–2213.

Murphy, K. P. (2013): *Machine Learning, a Probabilistic Perspective*, MIT Press.

Weitzman, M. L. (2007): "Subjective Expectations and Asset-Return Puzzles," *American Economic Review*, 97, 1102–1130.