

Monte Carlo Simulations
Political Science 597B
Spring 2005

Class Lectures: Monday 8:30-9:45, Room 236 Pond Laboratory
Lab Sessions: Thursday 1-2:15, Room 215 Burrowes Building

Suzanna De Boef
Associate Professor
Office Hours: Monday and Wednesday 10:00-11:30

SDeBoef@psu.edu
320 Pond Laboratory
863-9402

Kyle Joyce
Ph.D. Candidate
Office Hours: Monday and Wednesday 10:00-11.30

kjoyce@psu.edu
215 Pond Laboratory
863-0732

The Purpose

This course is designed as part of the graduate methods sequence for students in Political Science. It introduces students to Monte Carlo Simulation methods, an increasingly popular technique used in Political Science. Monte Carlo Simulation uses computer algorithms and pseudo-random numbers to conduct mathematical experiments on statistical estimators. These experiments can be used to assess the sampling distribution of a new estimator, the sensitivity of an empirical application to variations in population conditions, sensitivity to violations of estimator assumptions, and variations in model specification. It can also help us understand the properties of estimators, such as OLS or MLE, in the presence of small samples. This is not a required course but students interested in doing advanced methods should enroll.

The Plan

This course combines in-class lectures with a lab session which will be used to provide instruction in R, an open-source software package, which will be used to conduct the Monte Carlo simulations. No prior knowledge of R is required for the course. We will assume that no one is familiar with R, and students will be taught how to use R in order to conduct their homework assignments and class project. At the beginning of the course each student will receive a cd-rom with the most recent version of R, LaTeX, WinEdt, and Adobe Reader.

Evaluation

Evaluation for the course will be based on: weekly homework assignments (50%) and a final research project, which will be presented both as a paper (25%) and a poster (25%).

Weekly homework assignments

The homework assignments will be a mix of reading assignments (to be completed before the class period for which they are assigned), written work, and computer exercises. I will grade all assignments exclusive of the of computer exercises, which Kyle will grade. All written assignments should be done in LaTeX using the fullpage package (i.e., `\usepackage{fullpage}`). Also, please use the `\singlespace` option as appropriate to conserve paper. Computer exercises will focus on the tasks central to conducting Monte Carlo analysis and therefore parallel both the lectures and the lab sessions. All exercises must be kept in a lab book. We recommend that you also keep any R resources and handouts there as well. All computer exercises will be programmed in the R language. Kyle will be teaching you R from the ground up. Programs will be evaluated for their efficiency and good coding practices; it is essential that they are well commented.

The final research project

The purpose of the final research project is to conduct an original Monte Carlo analysis that informs your own substantive research. Each student will select a research question that requires Monte Carlo analysis, conduct the analysis and write a research paper presenting the results. In addition to the paper, the project must also be presented as a poster. The poster session will be open to faculty and students in the department and a prize will be given to the best poster.

The first draft of the final paper will be due Friday, April 15. The poster session will be Monday, April 25th over the noon hour in 302, pending final approval from the folks downstairs. The final paper will be due Wednesday of finals

week. These projects may be collaborative, with my approval. You will be able to tailor many of your computer exercises to contribute to your final project; in fact I recommend it.

Poster hints, LaTeX, and R resources will be posted on the course website as well as the Statistics Partnership website.

The Texts

Mooney, Christopher Z. 1997. *Monte Carlo Simulation*. Thousand Oaks, CA: Sage Publications.

Evans, Merran, Nicholas Hastings, and Brian Peacock. 2000. *Statistical Distributions*, 3rd ed. New York: John Wiley & Sons.

We will be providing a set of lecture notes which will be posted on the course web site before class each week, as well as a series of handouts which describe some of the R commands which are necessary to conduct the Monte Carlo simulations.

Articles from *The Political Methodologist* are available at: <http://polmeth.wustl.edu/thepolmeth.php>

Resources for R

Most manuals for R are available online at: <http://www.r-project.org/>. In addition, there are a couple of books that you might find useful.

Maindonald, John and John Braun. 2003. *Data Analysis and Graphics Using R*. Cambridge, UK: Cambridge University Press.

Fox, John. 2002. *An R and S-PLUS Companion to Applied Regression*. Thousand Oaks, CA: Sage Publications, Inc.

Resources for LaTeX

Kopka, Helmut and Patrick W. Daly. 2004. *Guide to LaTeX*, 4th ed. Boston, MA: Addison-Wesley.

The Weekly Plan

January 10

Introduction to Stochastic Simulation.

Today we will introduce stochastic simulation and compare how we can use simulation to conduct statistical inference to how we usually do statistical inference. We will talk about how stochastic simulation has been used (or could be used) in political science and why you might want to use simulation. We will also cover details of the course and schedule the lab sessions.

Assignment: Read the following before the first class session.

Mooney, Chapter 1.

Metropolis, N. 1987. The Beginning of the Monte Carlo Method. *Los Alamos Science* Special Issue: 125-130.

Franklin, Charles. 1992. A Primer on Simulation of Statistical Models. *The Political Methodologist* 5(1): 14-17.

Mooney, Christopher Z. and George A. Krause. 1997. Review Article: Of Silicon and Political Science-Computationally Intensive Techniques of Statistical Estimation and Inference. *British Journal of Political Science* 27: 83-110. Read pages 83-87, 94-101 and 101-103 on bootstrapping.

Mooney, Christopher Z. 2003. Monte Carlo Simulation. In *The SAGE Encyclopedia of Social Science Research Methods* edited by Michael S. Lewis-Beck, Alan E. Bryman and Tim Futing Liao. Thousand Oaks: Sage Publications, Inc. Volume 2: 662-664.

January 13 **Lab Session** Software installation and basics for R and LaTeX.

January 17 **Probability Review.**

Today we will review probability theory. An understanding of probability theory is central to the conduct of Monte Carlo simulations. We will focus on the most widely used discrete and continuous distributions and their fit with different political phenomenon.

Assignment: Read before class this week.
Evans, Hastings and Peacock Chapters 1 and 2.

Mooney, Section 2.2.

Erikson, Robert S., Michael B. MacKuen and James A. Stimson. 2002. *The Macro Polity*. Cambridge University Press. Chapter 10, including its Appendix.

Writing Assignments: Pick three statistical distributions and describe which political processes match the fundamental principles of those distributions. Due January 24.

Summarize the purpose of the simulations in *The Macro Polity* and what they tell us that we cannot learn otherwise. Discuss the effectiveness of the simulations for teaching us about the macro polity: What problems/difficulties does the analysis present? What strengths does the method have? How does the analysis compare to the early Monte Carlo work done at the Los Alamos labs? Due January 17.

January 20 **Lab Session:** Introduction to R, conducted by the Statistics Partnership, 12-1pm 302 Pond Laboratory.

January 24 **Generating Random Variables.**

Random variables are the key to Monte Carlo analysis! Today we will talk about how pseudo-random variables are generated, methods for testing their randomness, and methods for testing how well they approximate the distribution.

Readings: Read before class this week.
Mooney, Chapter 2.

RAND Corporation. 2001. *A Million Random Digits with 100,000 Normal Deviates*. Santa Monica, CA: RAND. (Forward, Introduction, and skim tables of random digits and random deviates)

Tomz, Michael and Robert P. Van Houweling. 2003. How Does Voting Equipment Affect the Racial Gap in Voided Ballots? *American Journal of Political Science* 47(1): 46-60.

Writing Assignment:

Summarize the purpose of the simulations in Tomz and Van Houweling and discuss what information they provide. Identify the random variables in the simulation, their distributions, and the assumptions used to justify the distributions. Discuss the appropriateness of the assumptions. Due January 24.

January 27 **Lab Session:** Generating random variables in R.

Computer Exercise: Generate a series of random variables in R, present a histogram (or density) of the generated variables, and test to see how random they really are. Due February 3.

January 31

The Data Generating Process.

Now that you know how to generate the random portion of the model, we need to specify the data generating process (the mathematical model) that describes the representation of our theory. How is Y generated as a function of X and the random process? What parameters link X to Y? How strong is the relationship? What constants are included? How simple or complex should the relationship be? What types of variation are involved? We'll also discuss the practical issues such as how do we pick parameter values, what should the variance of the error term be, the signal to noise ratio in the data, etc.?

Readings: Read before class this week.

Mooney, Chapter 3-3.1; 2-2.1.

Bartels, Larry M. 1991. Instrumental and "Quasi-Instrumental" Variables. *American Journal of Political Science* 35(3): 777-800.

Writing Assignment:

Write a short paper summarizing the research question and the role of simulations in Bartels' article. What is the DGP and how does he select the parameter values, variance of the error term, sample size, number of repetitions, etc? Explain how Bartels makes his choices and why it matters. Due January 31.

February 3

Lab Session: Specifying a DGP.

Computer Exercise: Write a DGP and explain it in behavioral terms and as much statistical detail as possible. Specify values for parameters (and justify them), specify specific forms that the X process(es) would take, the distribution that the stochastic portion of the DGP would take, etc. Due next February 10.

February 7

Quantities of Interest.

This week we'll devote more time to the practical issues associated with simulation: how many trials, how do we program efficiently, which are the best ways to loop? What information do we want from the experiment and what do we do with it? There are many pieces of information that may be of interest to us. We'll discuss several potential candidates: How biased are estimates of the parameters or their standard errors from a particular estimator and given the specific DGP? Which of two (or more) alternative estimators are more efficient under a specific set of conditions?

Assignment: Read before class this week.

Mooney, Chapter 3.2-3.6.1.

De Boef, Suzanna and Jim Granato. 1997. Near-integrated Data and the Analysis of Political Relationships. *American Journal of Political Science* 41(2): 619-640.

Writing Assignment:

Write a short paper summarizing the research question and the role of simulations in De Boef and Granato's article. Address the following questions: What is the DGP and how do De Boef and Granato select the parameter values, variance of the error term, sample size, number of repetitions, etc. for the simulations? What are the quantities of interest and why? How/why do they select these quantities? And how do they use these quantities to answer their research question? Due February 7.

February 10

Lab Session: Setting up loops and extracting quantities of interesting.

Computer Exercise: Run a small number of trials and collect basic quantities of interest from these trials. February 17.

February 14 **Replication Working Session.** We will discuss Boehmke and begin replication/extension of his 2003 *Political Analysis* paper as a group. The lab sessions for the next 3 periods will continue the replication.

Assignment: Read before class this week.

Boehmke, Frederick J. 2003. Using Auxiliary Data to Estimate Selection Bias Models, with an Application to Interest Group Use of the Direct Initiative Process. *Political Analysis* 11(3): 234-254.

Writing Assignment:

Write a short paper summarizing the research question and the role of simulations in the Boehmke article. Address the following questions: What is the basic DGP used in the paper. Describe in words how you would approach replicating these simulations in R. Focus in particular on how you would generate the data and retrieve the quantities of interest. Due February 14.

February 17 **Lab Session** Replication/extension.

Computer Exercise: Begin the simulation replication/extension.

February 21 **Conducting Multiple Experiments—What to Vary.**

This week discuss the importance of conducting experiments under a variety of conditions. It's important to vary sample size, the relationships among independent variables, the signal to noise ratio in the data, etc. The inter-relationships among these features of the DGP can affect your results so it's important to conduct multiple experiments to test the sensitivity to these variations.

Assignment: Read before class this week.

Mooney, Chapter 3.6.2.

Beck, Nathaniel and Jonathan N. Katz. 2004. Time-Series Cross-Section Issues: Dynamics, 2004. Paper presented at the Annual Meeting of the Society for Political Methodology. Palo Alto, CA.

Writing Assignment:

Write a short paper summarizing the research question and the role of simulations in the Beck and Katz paper. Address the following questions: What is the basic DGP and how do Beck and Katz select the parameter values, variance of the error term, sample size, number of repetitions, etc. for the simulations? What do they vary across the experiments and why is this important or necessary? Due February 21.

February 24 **Lab Session** Replication/extension continuation.

Computer Exercise: Continue from last week.

February 28 **Pulling it together**

This week we'll answer questions and catch up if need be. If we're on target, we'll talk about variance reduction and importance sampling.

Assignment: Read before class this week.

Mooney, Chapter 3.6.2.

An approximately 3 page description of your final project is due before you leave for spring break. You must present your research question, tell me why Monte Carlo analysis is appropriate, provide a detailed description of the DGP, including parameter values, sample sizes, etc., and the variations you will make across different versions of your simulations.

March 3 **Lab Session** Boehmke replication/extension due today.

- March 7 **Spring Break – No Class.**
- March 14 **Variance Reduction, Importance Sampling, Bootstrapping, and other fancy terms relevant to Monte Carlo analysis.** We will talk today about how Monte Carlo analysis relates to each of these terms.
- Assignment:** Read before class this week.
Mooney, Christopher Z. 1996. Bootstrap Statistical Inference: Examples and Evaluations for Political Science. *American Journal of Political Science* 40(2): 570-602.
- More readings TBA.
- March 17 **Lab Session** The lab session will focus on organizing the logic of your program; how do we program efficiently; which are the best ways to loop?
- Computer Exercise:** TBA.
- March 21 **Working Session or Problems and Progress**
- March 24 **Lab Session** The lab session will discuss bootstrapping and miscellaneous topics of interest.
- Computer Exercise:** TBA.
- March 28 **Presenting results.**
Presenting Monte Carlo results in a way that readers can easily grasp the findings is a challenge. There is often too much information to present in tabular form, and when you can it's just not effective! We'll talk about alternative options: three-dimensional graphs and other ways of presenting results.
- Assignment:** Read before class this week.
Readings: Franklin, Charles. 1989. Graphical Displays in Political Science. *The Political Methodologist* 2(2): 7-9.
- Adolph, Christopher. 2003. Visual Interpretation and Presentation of Monte Carlo Results. *The Political Methodologist* 11(2): 30-35.
- March 31 **Lab Session** The lab session will cover graphing in R.
- Computer Exercise:** Turn in some nontabular presentation of some portion of their results. Due April 7.
- April 4 **Working Session.**
- April 7 **Lab Session** Working session.
- Computer Exercise:** TBA.
- April 11 **In Class Presentations.**
Lab Session Working session, if necessary.
- April 14 **Computer Exercise:** No computer exercises the rest of the semester.
- April 18 **In Class Presentations.**

- April 21 **Lab Session** Working session, if necessary.
- April 25 **Poster Presentations (we won't have this during class, but in lieu of class)**
- April 28 **Lab Session** No lab session.

Academic Dishonesty¹

The Department of Political Science, along with the College of the Liberal Arts and the University, takes violations of academic dishonesty seriously. Observing basic honesty in one's work, words, ideas, and actions is a principle to which all members of the community are required to subscribe.

All course work by students is to be done on an individual basis unless an instructor clearly states that an alternative is acceptable. Any reference materials used in the preparation of any assignment must be explicitly cited. In an examination setting, unless the instructor gives explicit prior instructions to the contrary, whether the examination is in-class or take-home, violations of academic integrity shall consist of any attempt to receive assistance from written or printed aids, or from any person or papers or electronic devices, or of any attempt to give assistance, whether the one so doing has completed his or her own work or not.

Other violations include, but are not limited to, any attempt to gain an unfair advantage in regard to an examination, such as tampering with a graded exam or claiming another's work to be one's own. Violations shall also consist of obtaining or attempting to obtain, previous to any examinations, copies of the examination papers or the questions to appear thereon, or to obtain any illegal knowledge of these questions. Lying to the instructor or purposely misleading any Penn State administrator shall also constitute a violation of academic integrity.

In cases of a violation of academic integrity it is the policy of the Department of Political Science to impose appropriate penalties that are consistent with University guidelines.

¹Much of the text above has been directly obtained from the sections of the Princeton University website (<http://www.princeton.edu/pr/pub/rrr/99/pages/OI.htm>) concerning academic integrity (Rights, Rules, Responsibilities introductory text as well as pages 55-69) as well as from the website of the Department of Economics at The Pennsylvania State University.

Disabilities

The Pennsylvania State University encourages qualified people with disabilities to participate in its programs and activities and is committed to the policy that all people shall have equal access to programs, facilities, and admissions without regard to personal characteristics not related to ability, performance, or qualifications as determined by University policy or by state or federal authorities. If you anticipate needing any type of accommodation in this course or have questions about physical access, please tell the instructor as soon as possible. Reasonable accommodations will be made for all students with disabilities, but it is the student's responsibility to inform the instructor early in the term. Do not wait until just before an exam to decide you want to inform the instructor of a learning disability; any accommodations for disabilities must be arranged well in advance.

Visit our web site

The Political Science Department is in the process of upgrading its web site and will continue to do so during this and future academic years. In the undergraduate section you will find a wealth of information including course schedules, faculty office hours, faculty home pages describing their areas of teaching and research activities, answers to questions about advising, internship opportunities, announcements,

and much, much, more. Check back often: we will continuously update our information about internships and career opportunities: <http://polisci.la.psu.edu/>