Syllabus: Understanding & Modeling Social Processes – An Introduction to Simulation

Christoph Zangger, Unibe & LMU

WiSe, 2022/2023

E-mail: christoph.zangger@unibe.ch
Web: www.christophzangger.com

Time: Mo 14:15–18:00
(≈ every other week)

Place: Konradstrasse 6
Room 109 / online

---

1 Organization, requirements and grading

1.1 Organization

This course emphasizes practical aspects of statistics and simulation studies. To this end, students are asked to bring their own laptops to every lecture, preferably with `R` or/and `Stata` already installed. Theoretical inputs will usually be complemented by a direct application of concepts to solve problems, which students and researchers frequently encounter in applied social research. Likewise, there will be time to work on the assignments during class such that the workload should be evenly distributed throughout the semester.

There is no required reading for this course, although I suggest you have a look at the book recommended in the last section of this syllabus. All the material, including the assignments, peer assessments, code, slides, screencasts, interactive refreshers, and literature for further reading are provided on Moodle.

1.2 What you should already know...

Among other things, we will use simulation approaches to understand different stochastic and statistical concepts in a more intuitive way. Nevertheless, students should already be familiar with basic univariate and preferably multivariate statistics (as taught in introductory statistics lectures, for example, Statistik I, Angewandte Datenanalyse or Political Science Methods II: Multivariate Data Analysis). Given the applied focus of the course and the pursued approach, students with little knowledge in the rather abstract and dry domain of statistical theory should nevertheless be able to follow the course. Although not a prerequisite, some experience in programming will be of help (e.g., in `R`, `Python`, `Stata`).

1.3 ...what you will (hopefully) learn in this course

Students will use simulation approaches on artificial as well as 'real' data to answer substantial questions regarding both theoretical and methodological concepts. Doing so, they will

get a more intuitive understanding of statistical concepts and they will be able to present meaningful answers to complex phenomena in an accessible way, for example by means of graphical representations of predicted outcomes. Moreover, combining simulation techniques with the counterfactual approach to causal inference, students acquire skills to evaluate policy interventions and assess their potential impact on people's lives.

1.4 ...and how you will get your credit points

Given the applied nature of the course, you are asked to hand in your solutions to *four assignment*. Additionally, you have to complete *three peer assessments*, that is, provide feedback to fellow students. The solutions should be handed in *no later than 3 weeks after the assignment*. The tasks will be based on the topics covered in the course and will usually apply a certain simulation procedure to a commonly encountered problem or to present evidence in a more accessible way to a broader audience. The solutions to these tasks (and the peer assessments) can be handed in either *English* or *German*.

Since we will often collectively apply theoretical concepts to problems during the course, there will be time to work on parts of the assignments during the lectures. The tasks can either be completed individually or in small teams of no more than 3 students. A sample solution will be provided for all tasks and students will get individual feedback on their own solutions (at the same time as you will receive the feedback from your peers). As a guidance, the rubric I use to assess your solutions can be found on Moodle. Moreover, I will provide you with a structured form and corresponding guidelines to conduct the peer review / assessment.

2 Course structure

This course uses different formats. Inputs are usually followed by quizzes, life polls, and other interactive formats to clarify concepts, collect open questions, and to find collective solutions. This is usually followed by an applied part, in which you either work individually or in teams on small problems (e.g., adjusting some code to generate a meaningful answer to a substantial question). Finally, there will also be asynchronous elements that you can complete at any time before or after the actual meeting, for example self-assessments to check whether some of the main ideas stuck.

If the course needs to be held online (and also if not), I would like to plan one or two voluntary "social gatherings" in which we can meet more informally (get some drinks) such that you have an opportunity to get to know each other.

3 Overall learning outcomes

By the end of this course, students ...

- ... are able to construct simple stochastic processes (random variables) using software.
- ... can simulate the asymptotic assumptions of the frequentist approach to statistics and can explain them in an accessible way to peers and people less familiar with quantitative methods.
- ... can test the performance of different estimators using Monte Carlo methods.
- ... can develop simulation-based approaches to improve the robustness of their analyses (e.g., bootstrapping).

- ... can present the results of complex statistical procedures to a non-scientific audience in a meaningful way (simulations based on real microdata).

These skills will not only be of use for your future studies in a variety of disciplines (e.g., when working on your Bachelor's or Master's thesis), but the techniques learned in this course are directly applicable in future jobs: Being able to grasp and model the inherent uncertainty of social processes gives you a more adequate understanding of our social world and enables you to make better predictions. Moreover, the ability to explain and present the results of complex methods and models in an accessible, meaningful way is crucial to pass on knowledge, for example to policy makers.

4 Program

Week 1, 17.10.2022 *Organization & introduction; An applied review of basic probability theory*

Short outline of ...

- ...the topics covered in this course,
- ...theoretical and practical approaches,
- ...and the requirements.

Moreover, a first introduction to some notation and concepts used throughout the course and an assessment of students needs and expectations.

Simulation studies make use of stochastic processes. Probability theory is thus essential to such approaches. Using statistical software, we are going to look at...

- ...how to generate different types of random variables.
- ...how to plot different probability distributions.
- ...how to do this in repeated samples.

Interlude, 24.10.2022 *Voluntary introduction to* `R` (online)

For those less / not familiar with `R`, I'll provide a short voluntary introduction to

- `R` and RStudio
- basics of programming in R
- where to find help when your code doesn't want what you want it to do

Week 2, 31.10.2022 *Another look at statistics using simulations*

It's been a long time...so time for a refresher! Putting asymptotic theory to the test, we will review basic concepts of inferential statistics such as...

- ... standard errors,
- ... confidence intervals,
- ... and how this relates to the frequentist's approach to statistics...

by means of simulating our own data to check these and other concepts.

Simulating single random variables and distributions is nice but doesn't help much in understanding complex social phenomena and statistical methods. To this end, we will learn to...

- ...simulate multivariate relationships,
- ...variance structures,
- ...and how this will be of use to test sociological theories.

Assignment task 1 (due date: 28.11.2021)

Week 3, 21.11.2022 *Understanding linear regression assumptions using simulations*

Is it really that bad if there's multicollinearity among independent variables? What about a non-constant variance? And what do all these things mean anyway?

We will have a look at:

- Some basic of the linear model.
- Simulate such models and evaluate its performance.
- Check how the violations of underlying assumptions affects our estimates.

Continuing with linear models, we will...

- ...further evaluate violations of OLS assumptions,
- ...investigate omitted variable and endogeneity issues,
- ...simulate two-step procedures, namely an instrumental variable approach.

Assignment task 2 (due date: 12.12.2021)

Week 4, 05.12.2022 *Understanding regression assumptions using simulations II*

Since not all outcomes of interest can be modeled using a linear model, we extent the perspective by means of simulating GLMs. We will...

- ...simulate different categorical measures,
- ...evaluate the outcomes of such models,
- ...and—in preparation for the applications that follow—more generally look at predicting outcomes.

Peer assessment 1 (feedback to assignment 2): due date: 02.01.2023

Week 5, 19.12.2022 *Are you sure about that? – Bootstrapping*

So far, we used Monte Carlo simulations. However, they are by no means the only form of simulation studies. In this lecture, we will have a look at other techniques based on real data, namely...

- ...Bootstrapping
- ...and reshuffling techniques.

Assignment task 3 (due date: 09.01.2023)

Week 6, 09.01.2023, online *Evaluating policy I: Labor supply modeling*

As a first illustration of how to use microsimulation studies to answer substantial theoretical questions, we will turn our attention to labor supply models. In this regard, we will...

- ...get to know such models and their purpose,
- ...apply the underlying simulation approach to survey data,
- ...and get some ideas how these models might be of use in other instances.

Peer assessment 2 (feedback to assignment 3): due date: 30.01.2023

Week 7, 16.01.2023 *Evaluating policy II: Causal inference & simulation*

Having learned how to use microsimulation approaches on real data, we are going to investigate how to use such techniques to construct counterfactuals to identify causal effects. Using a study on the effects of educational homogamy on inter-couple income inequality as an example (Wise & Zangger 2017), we learn...

- ...how reshuffling can be used to construct plausible counterfactuals,
- ...how to use them to investigate causal mechanisms,
- ...and how to present the results of such an analysis.

Assignment task 4 (due date: 06.02.2023)

Week 8, 30.01.2023 *Evaluating policy III: Simulations based on experimental data; Further topics*
Based on data from a choice experiment on intra-urban housing decisions, we examine how combining experimental data with microsimulations makes use of the advantages of both methods. Students will...

- estimate the underlying utility function,
- use these estimates as basis for a simulation,
- and make predictions based on this simulation about potential outcomes.

Having worked through different examples and applications of simulation studies, we are going to have a short look into other fields that relate to the approaches treated so far. Depending on students' interest, these could include...

- ...the analysis and treatment of missing data,
- ...Bayesian approaches,
- ...spatial models and spatial microsimulation,
- ...or similar topics.

Likewise, we can also have a deeper look at things already covered in previous lectures.

Peer assessment 3 (feedback to assignment 4): due date: 27.02.2022

5 Literature

You are generally not required to read any textbook or articles during this course. However, feel free to do so. A selection of useful books and articles is provided on Moodle. I especially encourage you to have a look at the book of Carsey & Harden (2014), which provides an accessible introduction to the topic. Moreover, there is a script with additional background material and code that you might also consider looking at (also on Moodle, although it's work in progress).