Math475: Statistical Computation

Fall 2021

Instructor:	Professor Nan LIN
	Office: Room 205, Cupples I
	Email: nlin@wustl.edu
Time and location:	12 -12:50pm MWF at Wrighton 250
Office hours:	1-2pm Monday and 11-11:50am Wednesday

General information

Textbook: *Computational Statistics*, 2nd Edition, by Geof H. Givens and Jennifer A. Hoeting. Wiley, 2012. ISBN: 9780470533314

Reference: *Statistical Computing with R*, 2nd edition, by Maria L. Rizzo. Chapman and Hall/CRC, 2019. ISBN: 1466553324 (Available electronically at the WUSTL library)

Class webpage: All homework assignments, handouts, and other information will be available on Canvas (http://mycanvas.wustl.edu/). Students are responsible for checking the class webpage frequently for updates.

Course Description

The aim of this course is to introduce modern computational statistics methods rather than being a programming course. Prior knowledge of basic R programming is assumed. Topics to be covered in this class include: numerical methods in R, Pseudo-random number generation, Monte Carlo approximation, bootstrap procedures, Markov chain Monte Carlo methods, permutation tests, etc..

Learning Objectives

Students are expected to gain theoretical understanding to the statistical methods covered in class, and are able to implement the methods in R in real applications and integrate them into future research.

Prerequisite

- **Math and statistics**: Calculus; linear algebra; probability and statistics at the level of Math493/494 is required. To benefit more from this class, students should have also taken at least one course on statistical models such as math322, math420, math434, math439, or equivalent.

- **Programming**: Programming skills are essential for this class. Familiarity with R is required. To review basic R programming, students may refer to the online book "An Introduction to R" at `http://cran.r-project.org/doc/contrib/Lam-IntroductionToR_LHL.pdf`. Another useful online book on R is "simpleR - Using R for Introductory Statistics" at `http://cran.r-project.org/doc/contrib/Verzani-SimpleR.pdf`.

Course schedule

The following course schedule is tentative and subject to changes.

Week	Topic
1	Introduction, Vectorization
2	Numerical optimization, **Labor Day**
3	Numerical optimization, **HW1 due**
4	EM Algorithm
5	Simulation: Pseudo-random number generators
6	Monte Carlo methods, **HW2 due**
7	Monte Carlo methods, **Fall Break**
8	MCMC
9	MCMC, **HW3 due**
10	Bootstrap and Jackknife
11	Bootstrap and Jackknife
12	Permutation test, **HW4 due**
13	Density estimation, **Thanksgiving**
14	Student Presentation,
15	Student Presentation

Computing

Students are **required** to use R to complete all assignments. R can be downloaded freely from http://cran.r-project.org/, and it works under major operating systems, including Windows, Linux and Mac OS. The instruction will be mainly through the Mac OS platform.

Homework

There will be four homework assignments for the entire semester. For each homework, students must submit the homework electronically on Canvas, including well-organized written responses and R programs. The R program must be well structured and contain sufficient comments to ensure it is easy to read the codes. Students must work on these assignments independently. Late homework will not be graded and receives zero point.

Final Presentation

In the last one or two weeks of the semester (depending on the size of the class), every student needs to give a brief presentation in class about a computational statistical method. The instructor will provide candidate topics and relevant reading materials. Students may choose their own topics with the instructor's approval. Every student must attend all presentations. Grading of the presentation is based on a weighted average of peer evaluation and the instructor's assessment.

Grading

Grades will be based on the homework sets (80%) and the final presentation (20%). Cr means C or better if you elect "Credit/No Credit." The final letter grade is given according to the following scale.

$[95, 100]$	$A+$	$[83, 85)$	$B+$	$[65, 75)$	C
$[87, 95)$	A	$[77, 83)$	B	$[60, 65)$	D
$[85, 87)$	$A-$	$[75, 77)$	$B-$	< 60	Fail

Learning Tips

1. Try to show up in all the lectures. Make good notes.

2. Ask questions in class. Your questions may be others' as well. No questions are too elementary, and all deserve to be answered.

3. Discuss with your classmates about your questions.

4. Finish homework in time.

Class Policies

1. No auditing is allowed. Students must formally register for this class.

2. Late homework will not be graded and receives zero point.

3. Collaboration: I encourage discussion among students in broad. But students must complete the homework independently. For example, each student must write his/her own programs in entirety.

4. Academic Integrity: All students are expected to adhere to the university's academic integrity policy at `https://artsci.wustl.edu/academic-integrity`. Any student who is found to have cheated on an assignment or exam will receive a zero score for that work, regardless of the extent of the offense.

5. About COVID-19: Please follow the university policy at `https://covid19.wustl.edu/`.