Title: Algorithmic and Computational Statistics

Instructor: Robert S. Maier

Textbook: G. H. Givens and J. Hoeting, "Computational Statistics", second edition, Wiley, 2013.

Outline:

This Topics course was last successfully taught in Fall 2017 and is proposed for Fall 2019. It will be a higher-level counterpart to the existing course STAT 675 ("Statistical Computing"). Unlike STAT 675, it will not provide training in the use of statistical software, such as R. Instead, it will focus on theory: the mathematics of advanced computational statistics. It should appeal to students in the Applied Mathematics graduate program who would like to see statistical applications of analysis and numerical analysis, and to students in the Statistics graduate program.

Prerequisites: Any of the core courses MATH 523, 527, 575, or 564/566.

Topics:

Aspects of maximum likelihood and Bayesian inference. (Review.)

Numerical optimization and the numerical solution of nonlinear equations (Newton-Raphson and more advanced multivariate schemes; statistical applications).

Simulated annealing.

Numerical integration, with applications to Bayesian inference and simulation. Related mathematical topics, such as orthogonal polynomials of many types; how orthogonal polynomials are related to standard probability distributions.

The issue of limited machine precision: the IEEE format for floating point numbers, and how it affects the accuracy of numerical computations.

Sampling from a Bayesian posterior distribution: MCMC (Markov chain Monte Carlo), etc. Much background information on Markov chains and their structure.

Smoothing and nonparametric density estimation; mathematical aspects.

Information criteria for model selection, i.e., deciding between fitted statistical models on the basis of data; connections with information theory.

"Algorithmic and Computational Statistics"
MATH 577-1
Fall 2017

Place and Time: Room 307, Social Sciences Bldg., TTh 2:00–3:15.

Instructor: Prof. Rob Maier (Professor, Mathematics and Physics; affiliate member, Statistics GIDP). My office is Math 609B, and my phone numbers are 621-2617 (office) and 621-6893 (Mathematics Department). My current office hours are Tuesday 11:00–11:50, Wednesday 2:00–2:50, and Thursday 11:00–11:50, though these may change. My email address is `rsm@math.arizona.edu`.

Primary Audience: Students in the Mathematics and Applied Mathematics graduate programs who would like to see statistical applications of analysis and numerical analysis, and students in the Statistics graduate program.

Prerequisites: Preferably, any of the core graduate courses MATH 523ab, 527ab, 575ab, or 564/566. Formal exposure to probability and/or statistics and statistical inference is not necessarily required.

Overview: This course will cover many algorithms now used in statistical inference and machine learning; and even more, their mathematical underpinnings. It should be be especially valuable to those who may implement such algorithms from scratch, rather than relying on standard statistical software.

What This Course Is *Not*: A course in the use of standard statistical software, whether it be R, STATA, SAS, etc. Other courses at UA exist for that purpose. They include STAT 675 ("Statistical Computing"), EPID 503 ("Statistical Analysis Using STATA"), and EPID 504 ("Statistical Analysis Using SAS"). For teaching yourself R, the book by Maria Rizzo, "Statistical Computing with R," is recommended.

Course Textbook: "Computational Statistics", by Geof Givens and Jennifer Hoeting (2nd edition, Wiley, 2013), which is on sale in the bookstore. (Be warned: though popular, this resembles a reference book more than it does a textbook.) To bring everyone up to speed, the first two or three weeks of the course will be a review of basic probability and parametrized families of probability distributions, taken from G. Casella and R. L. Berger, "Statistical Inference."

Detailed Content: Aspects of maximum likelihood and Bayesian inference. (Review; see Chapter 1.) Numerical optimization and the solution of nonlinear equations; statistical applications. (See Chapter 2.). Introductory combinatorial optimization and simulated annealing. (See sections 3.1–3.4.) Numerical integration, with applications to Bayesian inference and simulation. (See Chapter 5. Related topics to be covered include orthogonal polynomials and how they are related to standard probability distributions, and to special functions; these are not in the text.) The issue of limited machine precision: the IEEE format for floating point numbers, and how it affects the accuracy of numerical computations. (Not in the text; a relevant [but highly technical!] reference is the book of P. Kornerup and D. W. Matula, "Finite-Precision Number Systems and Arithmetic.") Sampling from a Bayesian posterior distribution: MCMC (Markov chain Monte Carlo). (See section 7.1; additional background information on Markov chains will be supplied.) Smoothing and nonparametric density estimation; mathematical aspects of same. (See sections 10.1–10.2.) Information criteria such as AIC and BIC for model selection, i.e., deciding between fitted statistical models on the basis of data; connections with information theory. (Covered only briefly in the text; see, e.g., section 10.3.)

Grade Computation: This course will not make excessive demands on your time. There will be (1) as many as half a dozen homework assignments, comprising mathematically-oriented problems similar to those in the text, and (2) a mathematically-oriented final examination. The course grade will be based on (1) and (2), weighted in the ratio 2:1.

Classroom Etiquette (Technology-Related): This class is a technology-free zone. All mobile electronic devices, such as laptops, tablets, cellphones, and iPods, must be turned off during class meetings.

Classroom Etiquette (Not Technology-Related): Standard academic rules apply (there should be no eating or drinking, etc.).

Attendance Policy: Students are expected to attend class. If important circumstances prevent this, it is the student's responsibility to find out what was covered in class, what was assigned for reading or homework, and what special announcements (if any) were made. "Excessive absence" in this class will be construed to be absence from more than 10 percent of the scheduled class sessions, whether excused or unexcused, and will be subject to Administrative Drop as per UA policies.

University Policies: Students are expected to be familiar with and to abide by the UA's Code of Academic Integrity and Student Code of Conduct. UA's policy concerning Class Attendance, Participation, and Administrative Drops is available on-line, as is its policy on accommodation, where reasonable, of absences for any sincerely held religious belief, observance or practice. Absences pre-approved by the UA Dean of Students (or Dean Designee) will be honored.

Students with Disabilities: If you expect problems related to the course format or requirements, please register with the DRC (621-3268), and notify me in a timely way of your need for formal accommodations.

Final Note: Any information in the body of this handout, other than the grading and absence policies, may be subject to change with advance notice, as deemed appropriate by the instructor.