STA355H1F
Theory of Statistical Practice

Instructor: K. Knight (office: Sidney Smith 5016G; e-mail: keith@stat.utoronto.ca) My office is at the west end of the 5th floor.

Office hours: Thursdays 10 to noon, or by appointment. Do not hesitate to contact me by e-mail as many problems you might encounter can be easily resolved this way.

Goal: The main goal of this course is to provide students with the necessary tools of mathematical statistics necessary to be a good applied statistician. The focus of the course will be on the theory behind statistical methodology (from exploratory data analysis to formal statistical inference) and there will be a substantial data analytic component.

Textbook: The required textbook is Statistical Models by A.C. Davison (Cambridge University Press); we will not make extensive use of this book although it will serve as a valuable reference in subsequent courses. The textbook will be supplemented with a number of handouts and journal articles; most of these are already on Blackboard and more will be added as the course progresses. Some other good references are:


J. Rice: Mathematical Statistics and Data Analysis (3rd edition). (Duxbury)

(I definitely recommend buying the Nolan/Speed book.)

Other resources: There are a number of interesting blogs that deal with statistics and related topics. Some of my favourites are:

Statistical Modeling, Causal Inference, and Social Science: andreugelman.com

Big Data, Plainly Spoken: junkcharts.typepad.com/numbersruleyourworld

Simply Statistics: simplystatistics.org

Also, there is a recently published book “Past, Present, and Future of Statistical Science” that can be downloaded for free at nisla05.niss.org/copss/past-present-future-copss.pdf. This book contains (very eclectic) contributions from 50 prominent statisticians (including three from the University of Toronto) on various topics related to statistics. If you are considering a career in statistics or a related field, this book is definitely worth reading and even the more technical chapters are quite accessible.

Computing: To recognize the role of computing in mathematical statistics as well as to emphasize the connections between applied and mathematical statistics, we will use R extensively in this course both for data analysis as well as for carrying out simple Monte Carlo (simulation) experiments. R is free software and can be downloaded (for Windows, Mac, and Linux operating systems) from
cran.utstat.utoronto.ca. Documentation for R can also be found at www.r-project.org and this site also lists some books related to R. A useful book that gives a good introduction to R programming is

*A First Course in Statistical Programming with R* by Braun and Murdoch (Cambridge University Press)

**Evaluation:** The course grade will be based on four homework assignments (7.5% each for a total of 30%), a midterm exam (25%), and a final exam (45%).

- Homework assignments will involve both mathematical exercises as well as some computing (using R). Two assignments will be handed in before the midterm and two after.
- The midterm exam is scheduled for Monday November 2 from 2:10pm to 4pm (the usual lecture time) at a location to be announced later.
- The final exam will be held during the December exam period at a date and time to be announced later.
- **Students should familiarize themselves with the University’s policies on academic integrity, which can be found at www.artsci.utoronto.ca/osai/students.**

**Syllabus**

The following topics will be covered in the course:

**Short probability review.** Random variables, probability distributions and expected values, convergence in distribution and in probability, related theorems (CLT, WLLN etc), distribution theory for normal samples.

**Statistical models.** Sampling variation and uncertainty in estimation, order statistics, spacings, standard errors, jackknife estimates of bias and variance, density estimation, introduction to goodness-of-fit.

**Point and interval estimation.** Substitution principle, likelihood estimation, more on standard errors and their estimation, introduction to Bayesian estimation, confidence intervals, pivots (exact and approximate), credible intervals, bias/variance tradeoffs (in density estimation and non-parametric regression), robustness, methods for “big data”.

**Hypothesis Testing.** Elements of hypothesis testing, Neyman-Pearson Lemma and its consequences, p-values (and their behaviour under the null and alternative hypotheses), goodness-of-fit testing, multiple tests (“p-hacking” and false discovery rate).