This is Math 445: Statistical Theory

Math 445

March 28, 2016

Overview: Statistical Inference

3 Prongs of Statistics

1. Design
   The design of experiments and collection of data to more efficiently/correctly address scientific questions

2. Exploratory statistics
   Understand the major features of and detect patterns in data

3. Inferential statistics
   Account for randomness, variability, and bias in a sample in order to draw reasonable and correct conclusions about a population

Statistics vs. Probability

Probability (Math 440)

We learned how to calculate the probability of seeing a result (data) given a specific probability model (e.g., a specific distribution)

Statistics (Math 445)

We will learn how to make statements about the underlying probability models given the data we see

Example: Spies vs. Statisticians

During WWII, the Allies wanted to determine production rates of tanks (and airplanes, missiles, etc.)

Spies

Gathered intelligence (intercepted messages, interrogated of prisoners, etc.) and made the following estimates:

June 1940: 1000

June 1941: 1550

August 1942: 1550

Example: Spies vs. Statisticians

Statisticians

- The Allies had a sample of serial numbers (via capture, photography, etc.), $X_1, X_2, \ldots, X_n$, and there were N produced.
- Allied statisticians needed to devise an estimator to obtain N
- Ultimately, they used

$$\widehat{N} = X_{\max} + \frac{X_{\max}}{n} - 1$$

to get estimates

June 1940: 169

June 1941: 244

August 1942: 327

Example: Spies vs. Statisticians

After the war, the Allies discovered documents revealing the true number of tanks produced:

Month	Truth	Statisticians	Spies
June 1940	122	169	1000
June 1941	271	244	1550
August 1942	342	327	1550

Statistical Inference

Statistical inference

"A statistical inference is a procedure that produces a probabilistic statement about some or all parts of a statistical model" (Morris and DeGroot, 378).

Statistical model

A statistical model consists of

- a collection of random variables to describe observable data,
- the possible joint distribution(s) of the random variables,
- and the parameters, $\boldsymbol{\theta}$, that define those distributions

(Morris and DeGroot, 377)

Types of Inference in Math 445

Nonparametric

"The basic idea of nonparametric inference is to use data to infer an unknown quantity while making as few assumptions as possible. Usually, this means using statistical models that are infinite-dimensional." (Wasserman, 2006)

Parametric

A parametric inference uses models that consist of a set of distributions/densities that can be parameterized by a finite number of parameters.

Types of Inference in Math 445

Frequentist Paradigm

- Probability refers to limiting relative frequencies $\Longrightarrow$ "objective"
- Parameters are fixed, unknown constants
- Statistical procedures are designed to have well-defined long-run frequency properties

Bayesian Paradigm

- Probability describes a degree of belief $\Longrightarrow$ "subjective"
- Parameters are random variables because they are quantities about which we are uncertain
- Inferences for a parameter are made by producing a probability distribution for it

Tentative Schedule

Topic	Chapters	Approx. Duration
Exploratory Data Analysis	1–2	1 week
Nonparametric inference	3–5	2 weeks
Frequentist inference	6–8	4 weeks
Bayesian inference	10, supplements	3 weeks

Course Logistics

math445-lu.github.io

My Info

- email: adam.m.loy@lawrence.edu
- Office: Briggs 410
- Office hours:
  - MF 11:10am–12:20pm
  - MT 3:00–4:00pm
  - W 3:00–4:30pm
  - and by appointment

Required Materials

- Textbook:
  Mathematical Statistics with Resampling and R, Laura M. Chihara and Tim C. Hesterberg, 2011, Wiley, ISBN 978-1-118-02985-5.

- Access to R
  - You can download your own version of R and RStudio
  - You can access our RStudio server: `rstudio.lawrence.edu`

Grading

- Homework (50%)
  - Due (most) Wednesdays by 4:30 pm
  - No late work accepted without a valid excuse
  - Mix of theoretical and applied problems
  - Most applications will use R

- Exams (50%)
  - Midterm (25%) and a final (25%)
  - Each will have in-class and take-home components
  - Midterm tentatively scheduled for 5/4 during class
  - Final will be held on 6/6 from 3–5:30 pm