Theory and Application of Regression and Analysis of Variance - I
STAT 849

Miaoyan Wang

Department of Statistics
UW Madison

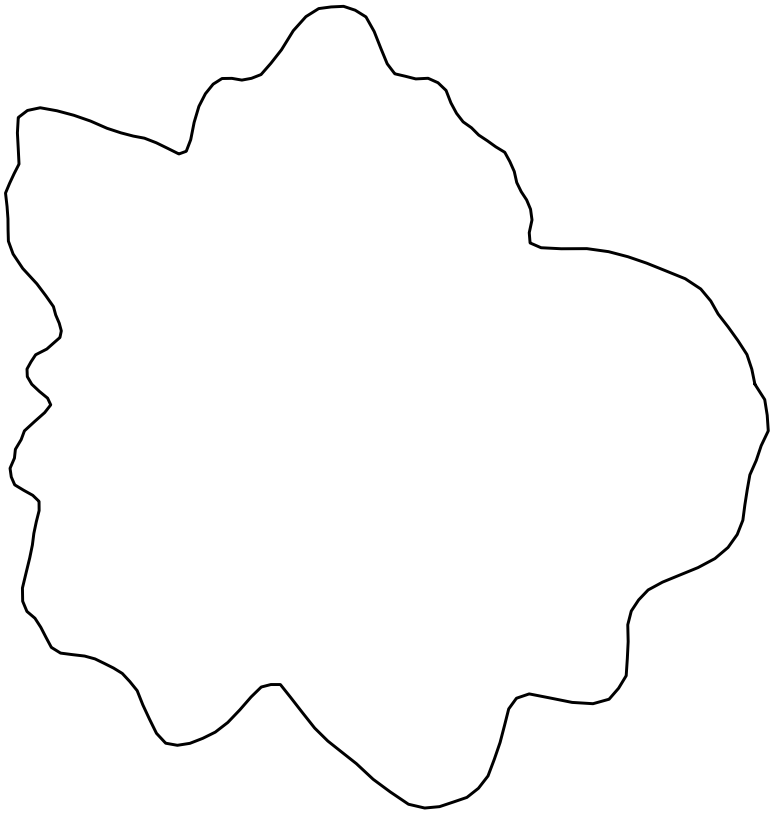
Overview

- This course is about **regression model**.
- It is designed for first or second year statistics PhD students. One of the four core courses to be tested in PhD qualifying exam.
- There are two courses in this sequence; the subsequent one is STAT 850, offered in spring semester.
- The course will be delivered **in person** through lectures and discussion sessions.
- Occasionally, we will have guest lectures by other professors, and one-to-one meetings with the Instructor.

Goals

- Learn about the linear model $\boldsymbol{Y} = \boldsymbol{X}\beta + \varepsilon$ in depth and detail.
- Cover concepts, distribution theory, computation, geometric interpretation, issues and fixed, regularization, and applications.
- Use applied statistics tools: bootstrap, cross-validation, permutations, and more.
- Touch upon connection to modern statistical research: reproducibility, random effects, sparsity, Bayes, causal inference and more.

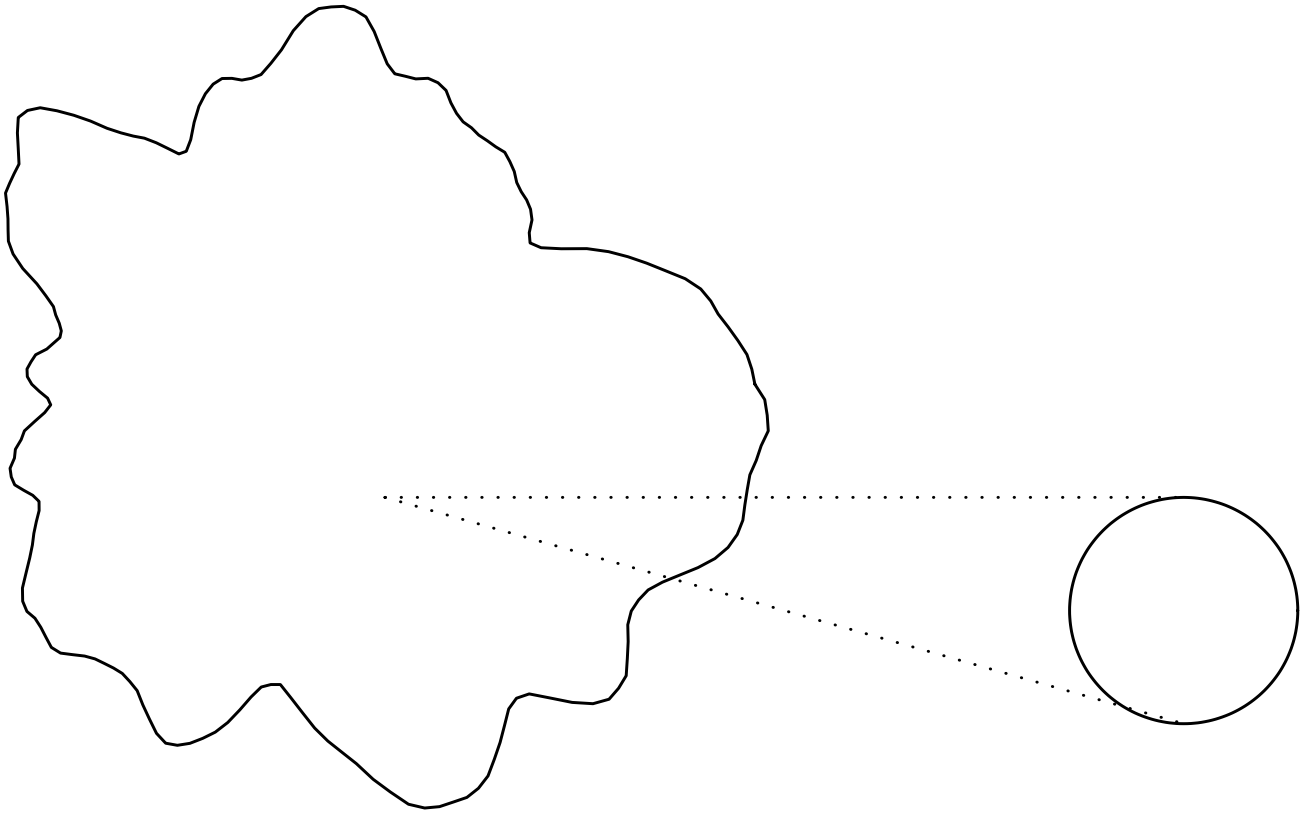
Probability vs. Statistics



POPULATION

parameters
(unknown)

Probability vs. Statistics



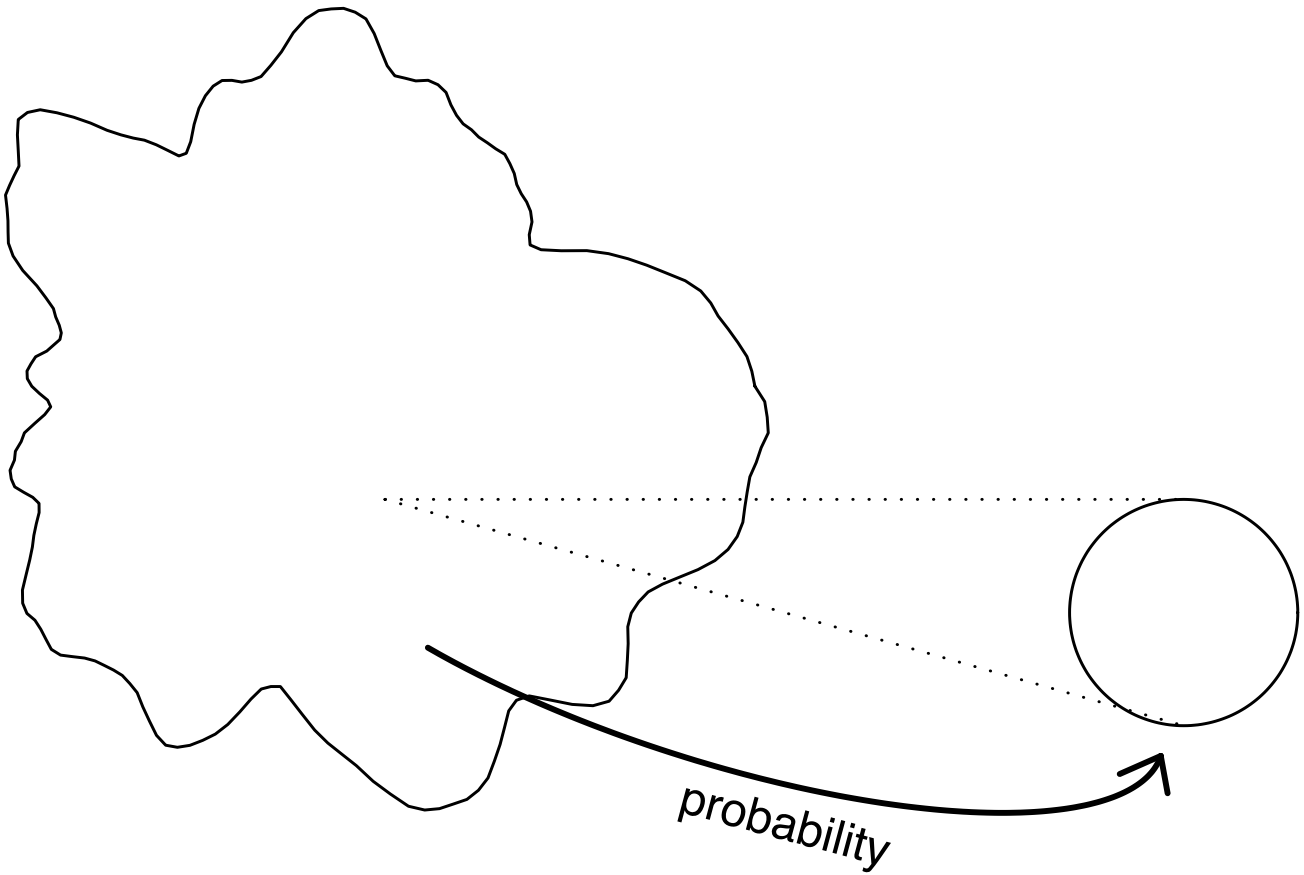
POPULATION

parameters
(unknown)

Sample

statistics
(from data)

Probability vs. Statistics



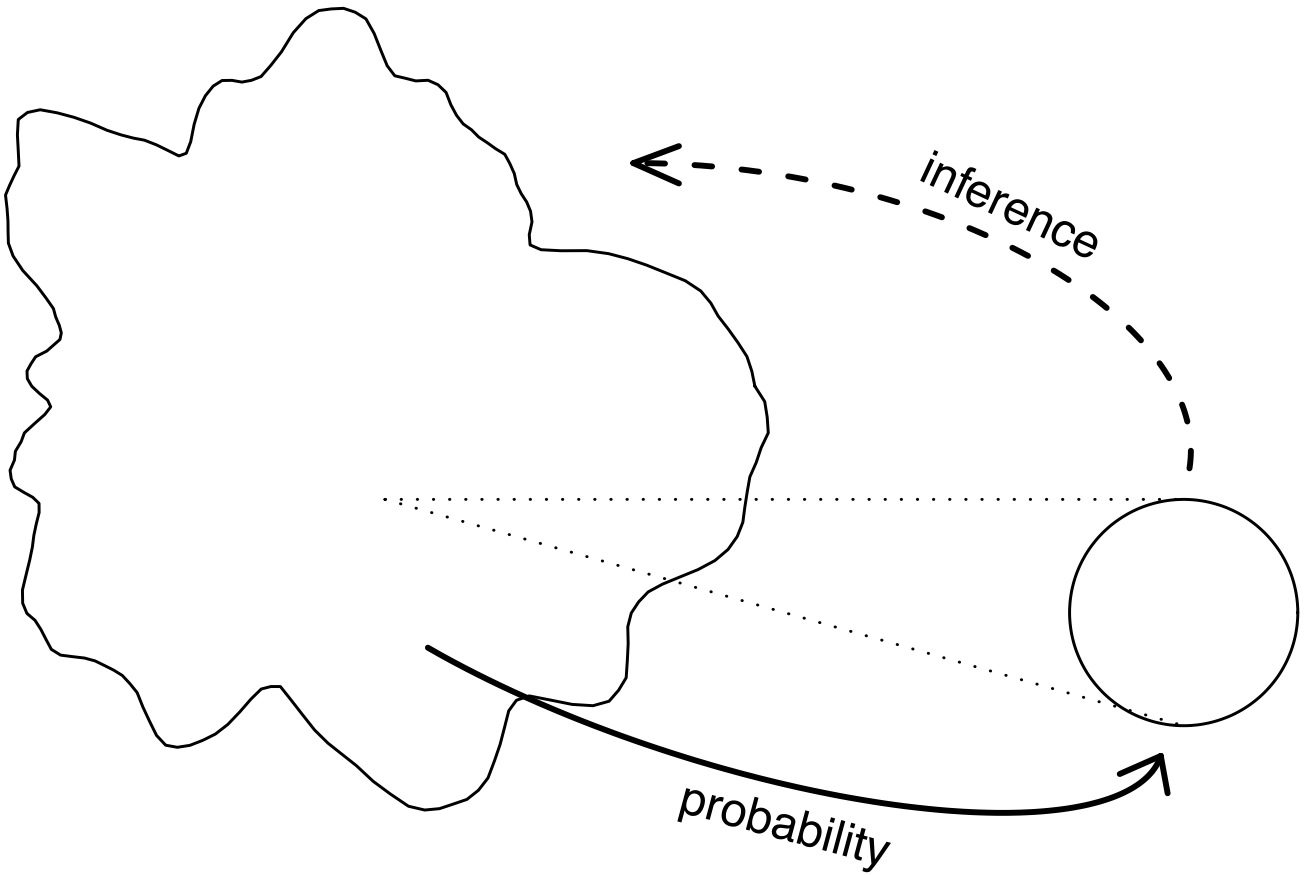
POPULATION

parameters
(unknown)

Sample

statistics
(from data)

Probability vs. Statistics



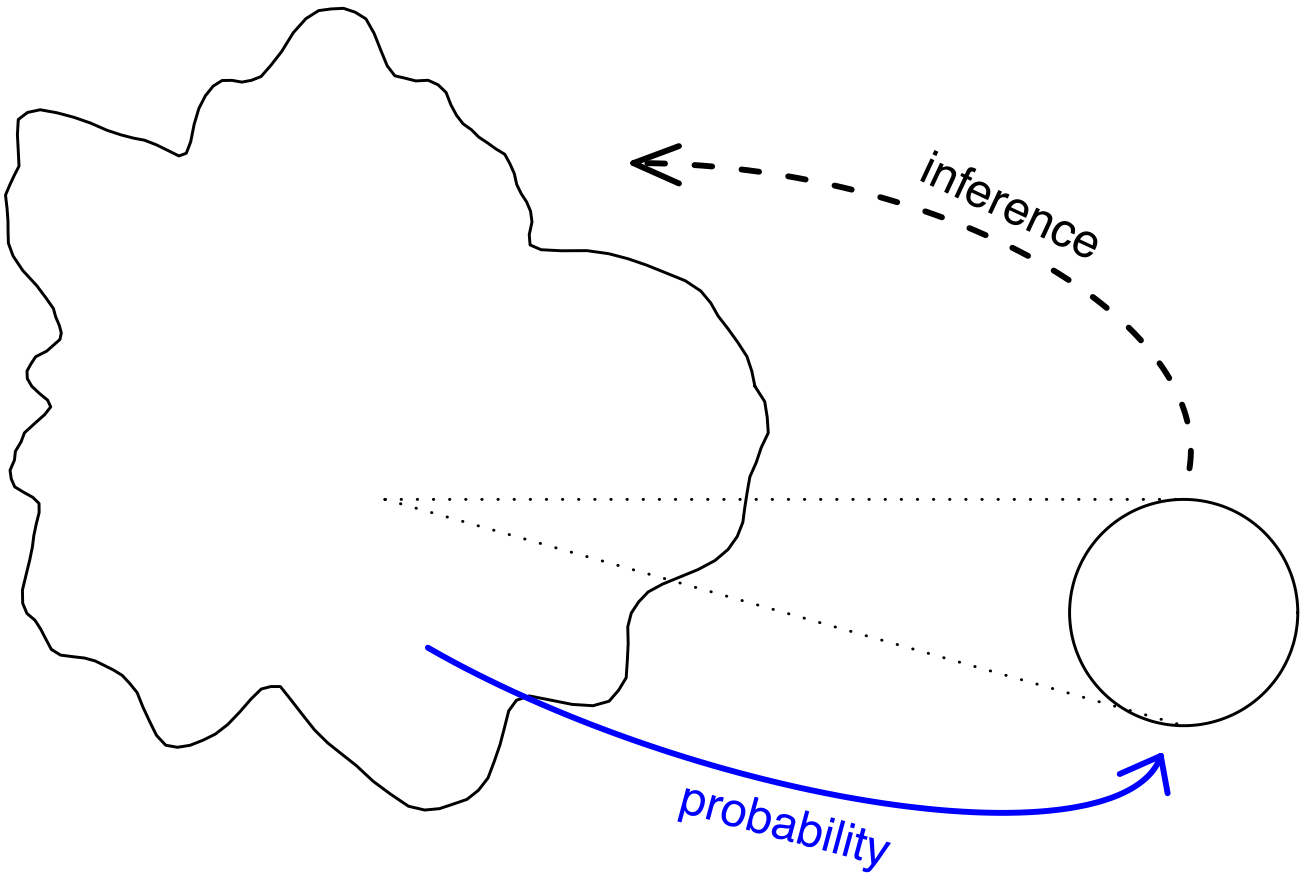
POPULATION

parameters
(unknown)

Sample

statistics
(from data)

Probability vs. Statistics



POPULATION

parameters
(unknown)

Sample

statistics
(from data)

Discussion

- Peter Norvig has a nice blog (`http://norvig.com/21-days.html`) on why it takes a long time to get good at something. If you're talented and work really hard at it, you can get very good at applied statistics in about 10 years. This course is designed to get you started and speed you along the way.
- Andrew Gelman's blog `http://statmodeling.stat.columbia.edu`. Good statistical practice is not just about the math or the computing but about how they interact with the underlying science and goals.
- Cautions about statistics and data science. You could be too data driven. An example?

Introduction

- A demanding course. Requires a general ability to do rigorous mathematical proofs and hands-on data analyses.
- Most students already have:
  - strong preparation in probability
  - linear algebra and analysis
  - basic statistics theory
  - hands on experience modeling data

  With hard work, you can make up one or one and a half deficits. More than that, you will feel lost.
- The course assumes a significant level of mathematical maturity. The minimal levels of math preparation can be found in the book "Plane Answers to Complex Questions" (on canvas).
  - Linear Algebra: Appendix B1-B4 and B7
  - Probability: Appendix C, D