

Lecture Notes in Statistics

Kaizhao Liu

November 2023

Contents

1	Introduction	5
1.1	In the Heartbeat of Our Real World	5
1.2	The Enchanting Tapestry of Interdisciplinary Insights	5
1.3	Organization	6
2	Three Basic Objectives of Statistics	7
3	The Curse of Dimensionality	9
4		11

Chapter 1

Introduction

1.1 In the Heartbeat of Our Real World

Statistics is a discipline where we harness the power of numbers to understand data. Its relevance and applications are deeply intertwined with real-world scenarios. Therefore, a solid statistical analysis must first prove its merit in addressing real-world challenges before we delve into its mathematical underpinnings. This mirrors the approach in computational mathematics and probability theory.

However, this might appear starkly different from what you've heard about pure mathematics. Pure math often delves into areas like algebra, number theory, and combinatorics, or it might explore more philosophical realms like mathematical logic, or even bridge with theoretical physics through geometry and analysis. But at their core, these are all frameworks for modeling the world quantitatively. Take combinatorics, for instance: its essence would be lost if we were merely counting systems without origin or purpose. Similarly, what would be the point of studying geometry or topology if they didn't in some way reflect the world around us? A classic example is the advent of analysis. It began as a means to describe object motion but has since permeated countless fields. Even disciplines like number theory, which at first glance seem esoteric, have their roots in deeper philosophical pursuits, aiming for comprehensive understanding.

In essence, while pure mathematicians might appear to be delving into the profound, they're fundamentally exploring challenges that, albeit sometimes abstract, have ties to our tangible world. This interconnectedness between various fields and the real world underscores the essence of our research. After all, the real world remains the ultimate benchmark for the validity and relevance of our academic endeavors.

1.2 The Enchanting Tapestry of Interdisciplinary Insights

Statistics, inherently applied, finds its inspirations from the needs across diverse fields. This includes the sciences – spanning computational science, physics, chemistry, biology, medicine, psychology, and geography – as well as the liberal arts, encompassing economics, finance, sociology, semantics, aesthetics, and numerous others I haven't detailed here. Consequently, our research is deeply interdisciplinary. We value the influence of these diverse disciplines, incorporating their challenges and ideas, which enrich and elevate the realm of statistics. We keep with the pace of them. Moreover, we recognize and harness profound mathematical advancements, arming ourselves with sophisticated tools and methods to address these multifaceted problems.

For instance, one of the early impacts of computer science on statistics was the development of bootstrap resampling. Prior to the computational revolution, statisticians heavily relied on stringent assumptions and theoretical derivations to make population inferences. These theoretical methods often imposed conditions that might not align with real-world data, such as assuming normal distributions or large sample sizes. However, with the advent of more powerful computers and advancements in computer science, Bradley Efron introduced the method of **bootstrap resampling** in the late 1970s. This innovative approach allows statisticians to resample or draw repeated samples from observed data, enabling the estimation of variability and confidence intervals for statistics without necessitating strict theoretical assumptions.

As another illustration, the influence of neuroscience on statistics is notably exemplified by the introduction of neural networks. Traditional statistical methods often make assumptions about data, such as linearity, independence, or homoscedasticity. While these methods have been robust

and enduring, they can face challenges when confronted with complex, non-linear data structures prevalent in today's vast datasets. Neural networks, designed to mimic the intricate workings of the human brain's neurons, exhibit the capacity to model and capture intricate, non-linear relationships in data.

1.3 Organization

Let us restate, **the goal of statistics** is to deal with data encountered in our real world. According to this goal, this lecture note focuses on the interplay between real world applications and theoretical development. As this subject is applied in nature, many of our theoretical investigations are motivated by demand in various related areas, such as sciences like computational science, physics, chemistry, biology, psychology, or liberal arts like

Chapter 2

Three Basic Objectives of Statistics

Chapter 3

The Curse of Dimensionality

Parallel to the development of the early twentieth century statistics, a new branch of science emerges, namely what we called computer science now. They were first used for calculating bullets. The theory of computational mathematics also revives as this new computational tool appeared. Numerical methods for calculus, linear algebra, and differential equations are extensively studied and implemented in the computers.

Chapter 4