MA 750 - Final Project

Nate Josephs Matthew Wiens Ben Draves

November 11, 2017

Abstract

One of the most fundamental tasks in Statistics is to understand the relationship between two random variables, X, Y, via an unspecified function $Y = f(X)$. Typically, $f(\cdot)$ is unknown and must be estimated from data relating X and Y. Estimating $f(\cdot)$ using maximum likelihood yields no meaningful solution when we consider *all* functions. Hence statisticians turn to estimating $f \in \mathcal{F}$ where $\mathcal{F}$ is a function space with some structure that provides meaningful solutions to the problem at hand. In most cases however, these function spaces are fixed, with no regard to the sample from which we are trying to infer f. In order to utilize all information inherent in the data while still imposing structure on $\mathcal{F}$, *Sieve Estimation* allows $\mathcal{F}$ to grow in complexity as n increases. Heuristically, as n increases, we attain a more robust understanding of f and should allow our modeling procedure to consider more complex forms of f. Sieve achieves this by introducing more complex functions to $\mathcal{F}$ as n increases. Here, we consider the function space

$$\mathcal{F}_n = \left\{ g(x) : g(x) = \sum_{d=1}^{D(n)} \beta_d x^d \right\}$$

where $D(n) \to \infty$ as $n \to \infty$. We focus our efforts on estimating $D(n)$ as a function of the data. This report is organized as follows; in sections 1 and 2 we summarize some of the foundational results on Sieve estimation and introduce notation used throughout the report. In section 3 we introduce some theoretical applications and in section 4 we offer some methodologies of estimating $D(n)$. Lastly, in section 5 we analyze our methods via intensive simulation study and a real data application.

1 Introduction

2 Developing Sieve Estimation

- Define series estimator
- Parameters - choice of basis function and dimensionality D
- Compare to Kernel Density estimates (Kernel and h)
- Suppose D is fixed - what is Bias, Variance, MSE, MISE
- develop PSE
- Connect MISE and PSE

3 Estimating $D(n)$

- Connect Cross Validation to PSE
- CV and other data dependent methods for choice of D
- Asymptotic Behavior of Cross Validation
- Convergence rates

4 Theoretical Applications

- Choosing basis function sets
- Relationship to penalization schemes and information criterion

5 Simulation and Applications

- Simulation - effect of basis sets and optimal dimension on the project
- Application

6 Conclusion