MA 750 - Final Project

Nate Josephs   Matthew Wiens   Ben Draves

November 12, 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

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

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 (**M**)

  Recall that a sieve estimator for the unknown mean function $g(x)$ in the regression is a sequence of models. The motivation of sieve estimators leads to additional constraints, namely that the models should be have finite dimension, and that the complexity (i.e. dimension) increases as n increases. Therefore the most natural choice for a sieve estimator and the choice we will focus on is a series estimator:

  $$g_m(x) = \sum_{d=0}^{D} \alpha_{dm} \phi_{dm}(x)$$

  [Hansen] Such a formulation leads to traditional problems in linear models, so given the dimension of the model and a choice of basis function, finding the coefficients α has well known solutions. The constraints in the formulation are light; namely the choice of basis function must approximate functions in L^2, among other natural constraints for the regression problem. For completeness, examples of non linear sieves are Neural Networks, or penalized sieve estimators. [Chen]

- Parameters - choice of basis function and dimensionality D (**M**)

  Given the sieve estimator defined as a series, there are two choices to be made: the dimension of the model, and the choice of basis function. D must be data dependent to scale the complexity of the estimator with the data. Therefore, for now we will assume there is an algorithm to choose $D(n)$ which produces a D_{opt}, for some sense of optimality. There are a number of choices for the basis function, and we will highlight just a few possibilities here:

  - Polynomials

    $$\phi_d(x) = \sum_{i=0}^{d} c_d x^d$$

    Within the polynomial class, choices of c determine the exact form. For example $\mathbf{c} = [0, 0 \ldots 1]$ would correspond to a basis of

$1, x, x^2, x^3....$ The Hermite Polynomials would correspond to another choice of $\mathbf{c}$, which have nice theoretical properties because the Hermite Polynomials are orthogonal.

– Fourier Series

$$\phi_d(x) = a \ cos(\pi dx) + b \ sin(\pi dx)$$

– Gaussian

$$\phi_d(x) = \varphi^{(d)}(x)$$

Where $\varphi^{(d)}(x)$ is the *dth* derivative of the Normal density, which will be the Normal density with coefficients of the *dth* Hermite polynomial. Like the Fourier Series, the Gaussian series estimator has the nice property that it goes to zero at $\pm\infty$.

Other common choices for the basis function are Splines and Wavelets. A univariate polynomial spline has dimension of the number of partitions plus the degree of the spline. The choice of basis in the univariate case extends naturally to the multidimensional case, where the multivariate basis is constructed as a tensor produce of the univariate basis. [Chen]

• Compare to Kernel Density estimates (Kernel and h) (**M**)

The sieve estimator has a number of similarities to the Kernel Density estimation problem. In both cases there are two choices to be made, one parameter that controls the bias variance tradeoff, and a second parameter that is related to underlying beliefs about the model. So the bandwidth in the Kernel estimation problem is analogous to the choice of dimension. Intuitively, as the dimension increases or the bandwidth decreases the estimator is more sensitive to local behavior and produces a rougher estimate. Similarly, the choice of the Kernel is analogous to the choice of basis function; the choice impacts the final model and exact statistical properties of the estimator, yet is less interesting because any reasonable choice of basis function or kernel function produces a similar estimate.

• Suppose D is fixed - what is Bias, Variance, MSE, MISE (**B**)

- develop PSE (**N**)

- Connect MISE and PSE (**N**)

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