APPM Thesis Draft

K. McLean

June 12, 2019

Introduction

The parameter λ for the class of linear predictors $\hat{\mathbf{Y}} = \mathbf{S}(\lambda)\mathbf{Y}$ where $\hat{\mathbf{Y}}$ is a column vector of estimates, $\mathbf{S}(\lambda)$ is an $n \times n$ linear smoother matrix dependent on λ and $\mathbf{Y}$ is a column vector of regressands, has different domains depending on the non-parametric estimator type of $\mathbf{S}(\lambda)$. For kernel density estimators, λ is frequently denoted as h, with $h \in [0, \infty)$. For spline estimators, λ is called k, with $k \in \mathbb{N}$. So comparing h to k directly is not feasible, especially considering the effects of choosing specific kernels and spline order.

λ is commonly chosen by cross-validation of the specific estimator. In Monte Carlo simulations, the mean squared error (MSE) of the estimators are compared with no attention paid to how much smoothness that choice of λ imparts on the data above and beyond other estimator choices. Cross-validation also introduces a dependency of λ on $\mathbf{X}$ (the $n \times m$ matrix of regressors) and $\mathbf{Y}$.

To maintain the exogeneity of the choice of λ (and therefore the independence of $\mathbf{S}(\lambda)$), let us establish a function indexed by λ that maps $\mathbf{S}(\lambda) \to \mathbb{R}$, $f_h : \mathbf{S_1}(h) \to \mathbb{R}$ and $f_k : \mathbf{S_2}(k) \to \mathbb{R}$ where the images to be the same when $\mathbf{S_1}(h)$ and $\mathbf{S_2}(k)$ impart the same amount of smoothness. After fixing the value of the image, we invert f_h and f_k to find the specific h and k that would have the specified degree of smoothness. Now comparing MSEs between estimators of equivalent smoothness is a fair comparison. We choose f_λ to be the trace of $\mathbf{S}(\lambda)$, which is the number of parameters estimated in linear models. The estimator therefore imparts more smoothness as $f_\lambda \to \infty$ and less smoothness as $f_\lambda \to 0$. The best estimator (the one with the lowest MSE) may not remain the same for different amounts of smoothness.

Model Set-Up

For linear predictors $\hat{Y} = \mathbf{S}(\lambda)Y$, where λ is the pertinent smoothing parameter (either number of knots for splines or bandwidth for kernel density estimators).