Economics 597: Nonparametric Econometrics

Department of Economics, Finance and Legal Studies
University of Alabama
Spring 2015

Midterm I

The exam consists of three questions on three pages. Each question is of equal value.

1. Consider the bias and variance formulas for the univariate kernel density estimator

$$
\begin{aligned}
Bias\left[\widehat{f}(\mathbf{x})\right] &\simeq \frac{h^2}{2} f^{(2)}(x)\,\kappa_2(k) \\
Var\left[\widehat{f}(\mathbf{x})\right] &\simeq \frac{1}{nh} f(x) R(k)
\end{aligned}
$$

   (a) Give the formula for the asymptotic mean intergrated squared error (AMISE).

   (b) Explain the relevance of each term in part (a).

   (c) Using the formula given in part (a), derive the optimal bandwidth.

   (d) Explain the relevance of each term in part (c).

   (e) Explain how you could make the optimal bandwidth in part (c) functional.

2. For the test of equality (Fan, 1994), show that the conditions show that the conditions for

$$H_n\left(z_i, z_j\right) = k\left(\frac{x_i - x_j}{h}\right) + k\left(\frac{y_i - y_j}{h}\right) - k\left(\frac{x_i - y_j}{h}\right) - k\left(\frac{y_i - x_j}{h}\right)$$

being a second-order degenerate U-statistic hold. For simplicity, assume a scalar variable and that the sample sizes and bandwidths are the same (i.e., $q = 1$, $n_1 = n_2$ and $h_x = h_y$).

3. Consider the lines of R code below. Write a brief note next to each line to explain what is happening. What is this code attempting to accomplish?

```
 rm(list=ls())
set.seed(123456)

library(stats)

 dat <- read.csv("data2000hpr.csv", header=TRUE)
 dat <- subset(dat, complete.cases(dat))
 attach(dat)

rgdpch00 <- as.matrix(rgdpch00/mean(rgdpch00))
rgdpch00 <- rgdpch00[order(rgdpch00),]

n <- length(rgdpch00)

KK.store.loo <- matrix(0,nrow=n,ncol=n)
KK.store.convolution  <- matrix(0,nrow=n,ncol=n)

h <- seq(0.001,0.5,0.001)

lscv.loo <- matrix(0,nrow=length(h),ncol=1)

for (jj in 1:length(h)){

 for (j in 1:n){

  dx <- (rgdpch00 - rgdpch00[j])/h[jj]
  KK <- solve(sqrt(2*pi))*exp(-0.5*dx^2)
  KK.convolution <- solve(sqrt(4*pi))*exp(-0.25*dx^2)

  KK.store.loo[,j] <- KK
  KK.store.loo[j,j] <- 0
  KK.store.convolution[,j] <- KK.convolution

 }

 lscv.loo[jj] <- solve(n^2*h[jj])*sum(KK.store.convolution) - 2*solve(n*(n-1)*h[jj])*sum(KK.store.loo)

}
```