Economics 673: Nonparametric Econometrics
Department of Economics, Finance and Legal Studies
University of Alabama

Problem Set #1 (Density – Estimation)

1. Is it possible to derive a Silverman type rule-of-thumb bandwidth if $f(x)$ was is Cauchy (simple intuition is sufficient)?

2. Provide an intuitive explanation of why the difficulty factor, $\int f^{(2)}(x)^2\,dx$, is a reasonable measure of difficulty. Is this measure scale invariant? What types of densities are 'easy' to estimate? Why (Terrell, 1990 is a good reference for more on this issue)?

3. Attention in the growth empirics literature has focused on the fact that the distribution of output, as measured by some form of GDP, has become increasingly bimodal as time has passed. Henderson, Parmeter and Russell (2008) tested this phenomena using data from the Penn World Table version 6.2 (available on the JAE data archive website). You will be using it to analyze the robustness of their results to a variety of features. Use only the 1970 and 2000 RGDPCH (rgdpch-70 and rgdpch-00 in the file) series from their paper. For your plots please print both the 1970 and 2000 plots on the same graph but make sure that you have no more than two plots per graph. Also, prior to estimating any densities convert the data my dividing by the mean output in each year. That is instead of plotting x plot $x/\overline{x}$.

   (a) Using the Silverman rule-of-thumb (1.06) please plot the distributions of output using a Gaussian kernel. Is the bimodal feature apparent in either year?

   (b) Using Silverman's appropriate rule-of-thumb bandwidth plot out these two distributions using the Epanechnikov kernel. Is the bimodal feature still visually apparent?

   (c) Calculate the LSCV bandwidths for the 2000 density for a Gaussian kernel (using equation 2.23 from chapter 2). Is the bimodal feature still apparent. List the value of the bandwidth chosen via LSCV.

   (d) Instead of using the leave-one-out estimator, show that when you fail to use the leave-one-out estimator that the bandwidth tends towards zero. Essentially I am asking you to list the bandwidth chosen via LSCV when you do not use the leave-one-out estimator.

Economics 673: Nonparametric Econometrics
Department of Economics, Finance and Legal Studies
University of Alabama

Problem Set #2 (Density – Testing)

1. For the test of symmetry (Ahmad and Li, 1997b), show that the conditions for $H_n(x_i, x_j) = k\left(\frac{x_i - x_j}{h}\right) - k\left(\frac{x_i + x_j}{h}\right)$ being a degenerate U-statistic hold.

2. For the test of correct parametric specification (Fan, 1994), show that for a Gaussian kernel and an assumed normal distribution that

$$\begin{aligned}\widehat{ISE}_n &= \frac{1}{n(n-1)h}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n} k\left(\frac{x_i - x_j}{h}\right) + \frac{1}{2\sqrt{\pi}(h^2+\widehat{\sigma}^2)^{1/2}} \\ &\quad - \frac{2}{n\sqrt{2\pi}(h^2+\widehat{\sigma}^2)^{1/2}}\sum_{i=1}^{n}\exp\left[-\frac{(x_i-\widehat{\mu})^2}{2\left(h^2+\widehat{\sigma}^2\right)}\right],\end{aligned}$$

3. Attention in the growth empirics literature has focused on the fact that the distribution of output, as measured by some form of GDP, has become increasingly bimodal as time has passed. Henderson, Parmeter and Russell (2008) tested this phenomena using data from the Penn World Table version 6.2 (available on the JAE data archive website). You will be using it to analyze the robustness of their results to a variety of features. Use only the 1970 and 2000 RGDPCH (rgdpch70 and rgdpch00 in the file) series from their paper. For your plots please print both the 1970 and 2000 plots on the same graph but make sure that you have no more than two plots per graph. Also, prior to estimating any densities convert the data my dividing by the mean output in each year. That is instead of plotting x plot $x/\overline{x}$.

   (a) Using the Li (1996) test, test that the 1970 and 2000 distributions are equal (using a Gaussian kernel, the Silverman rule-of-thumb bandwidth and 399 bootstraps).

   (b) Using the Fan (1994) test, test the null that the 2000 distribution is Gaussian (using a Gaussian kernel, the Silverman rule-of-thumb bandwidth and 399 bootstraps).

   (c) Using the test of Ahmad and Li (1997a), test the null that the 2000 distribution of real GDP per capita (rgdpch00) is independent of real GDP per worker (rgdpwok00) (using a Gaussian kernel, the Silverman rule-of-thumb bandwidth and 399).

Economics 673: Nonparametric Econometrics
Department of Economics, Finance and Legal Studies
University of Alabama

Problem Set #3 (Regression – Estimation)

1. If the joint density of two variables (Y and X) is normal, show that the conditional expectation ($E(Y|X=x)$) must be linear in x.

2. Show that when $m(x) = \alpha + \beta x$ is linear in x, then the local-linear estimator is unbiased (note that you should not use $h = \infty$ in this problem).

3. Heckman and Polachek (1974) suggest a quadratic parametric relationship between earnings and age

$$y_i = \alpha + \beta z_i + \gamma x_i + \delta x_i^2 + u_i$$

where y_i is the logarithm of earnings, z_i is education and x_i is age. Mincer (1974) finds that earnings increase with age through much of the working life but the rate of increase diminishes with age. Pagan and Ullah (1999) present a local-constant kernel estimate of an age earnings profile based on Canadian data (cps71 – available in the np package in R) for $n = 205$ males having common education (high school)

$$y_i = m(\overline{z}, x_i) + u_i.$$

   (a) Compute and plot the parametric quadratic as well as local-constant and local-linear estimates using a standard normal kernel with
      1. Rule of thumb bandwidth $h = 1.06\sigma_x n^{-1/(4+q)}$
      2. Bandwidth calculated using least-squares cross-validation

   (b) Is the dip present in the resulting nonparametric estimates?

   (c) Plot the nonparametric estimates along with their error bounds (use a wild bootstrap). Without conducting a formal test, does the dip appear to be significant?

   (d) Which nonparametric estimator appears to provide the most "appropriate" fit to this data?

Economics 673: Nonparametric Econometrics
Department of Economics, Finance and Legal Studies
University of Alabama

Problem Set #4 (Regression – Testing)

1. For the test of correct parametric specification (Li and Wang, 1998), show that the conditions for $H_n(x_i, x_j) = \widehat{u}_i \widehat{u}_j k\left(\frac{x_i - x_j}{h}\right)$ being a degenerate U-statistic hold.

2. For the test of correct parametric specification (Ullah, 1985), show that the following hold for the wild bootstrapped residuals:

   (a) $E[u_i^*] = 0$

   (b) $E\left[(u_i^*)^2\right] = \widehat{u}_i^2$

   (c) $E\left[(u_i^*)^3\right] = \widehat{u}_i^3$

3. Heckman and Polachek (1974) suggest a quadratic parametric relationship between earnings and age

$$y_i = \alpha + \beta z_i + \gamma x_i + \delta x_i^2 + u_i$$

where y_i is the logarithm of earnings, z_i is education and x_i is age. Mincer (1974) finds that earnings increase with age through much of the working life but the rate of increase diminishes with age. Pagan and Ullah (1999) present a local-constant kernel estimate of an age earnings profile based on Canadian data (cps71 – available in the np package in R) for $n = 205$ males having common education (high school)

$$y_i = m(\overline{z}, x_i) + u_i.$$

   (a) Using the Li and Wang (1999) test, test that the quadratic parametric specificiation is appropriate (use a wild bootstrap — 399 bootstraps). Use the

      1. Rule of thumb bandwidth $h = 1.06\sigma_x n^{-1/(4+q)}$
      2. Cross-validated bandwidth from the local-constant least-squares regresion

   (b) Using the local-constant least-squares version of the Ullah (1985) test, test that the quadratic parametric specificiation is appropriate (use a wild bootstrap — 399 bootstraps). Use

      1. Rule of thumb bandwidth $h = 1.06\sigma_x n^{-1/(4+q)}$
      2. Cross-validated bandwidth from the local-constant least-squares regresion

   (c) Using the Zheng (2009) test, test that the errors from the nonparametric model are homoskedastic (note: estimate the residuals via LCLS with a rule-of-thumb bandwidth and use a wild bootstrap — 399 bootstraps). Use the

      1. Rule of thumb bandwidth $h = 1.06\sigma_x n^{-1/(4+q)}$
      2. Cross-validated bandwidth from the local-constant least-squares regresion

Economics 673: Nonparametric Econometrics
Department of Economics, Finance and Legal Studies
University of Alabama

Problem Set #5 (Discrete Data)

1. Consider the following kernel functions

$$
\begin{aligned}
l\left(x_i, x, \lambda\right) &= \begin{cases} 1-\lambda \text{ if } x_i \neq x \\ \lambda/C \text{ if } x_i = x \end{cases} \\
l\left(x_i, x, \lambda\right) &= \binom{C}{\left|x_i - x\right|} \lambda^{\left|x_i - x\right|} \left(1-\lambda\right)^{C-\left|x_i - x\right|}
\end{aligned}
$$

   (a) Show that the kernels sum to 1.

   (b) Show that the density estimator

$$
\widehat{f}\left(x\right) = \frac{1}{n} \sum_{i=1}^{n} l\left(x_i, x, \lambda\right)
$$

   sums to 1.

2. Consider the following kernel functions

$$
\begin{aligned}
l\left(x_i, x, \lambda\right) &= \lambda^{1\{x_i \neq x\}} \\
l\left(x_i, x, \lambda\right) &= \lambda^{\left|x_i - x\right|}
\end{aligned}
$$

   (a) Show that each kernel does not sum to 1.

   (b) Show that the normalized density estimator

$$
\widetilde{f}\left(x\right) = \frac{\widehat{f}\left(x\right)}{\sum_{z \in S} \widehat{f}\left(z\right)}
$$

   does sum to 1.

3. Using your own code, replicate Table 8.1 in Henderson and Parmeter (2015, pp 223).