Econ 883-06: Problem Set 1

Federico A. Bugni*

October 22, 2020

Due date: Monday, November 2, 4 p.m., as an upload in Sakai

1. Let $\{X_i\}_{i=1}^n$ be an i.i.d. sample from a discretely distributed random variable $X : \Omega \to \mathbb{R}$ with probability mass function (PMF) $f_X(x) = P(X = x)$. Suppose that we are interested in estimating $f_X(x)$ at a particular value of $x \in \mathbb{R}$.

   As shown in Handout 2, we can estimate $f_X(x)$ by the sample analogue estimator, i.e.,

   $$\tilde{f}_X(x) \;\equiv\; n^{-1} \sum_{i=1}^{n} 1[X_i = x]. \tag{1}$$

   Another possibility could be to use nonparametric kernel density estimation. In particular, consider the nonparametric kernel density estimator $\hat{f}_X(x)$ with a bounded and second-order kernel function $K : \mathbb{R} \to \mathbb{R}$ with $\int |K(v)|v^2 dv < \infty$, $\int K(v)v^2 dv = A \neq 0$, and $\int K(v)^2 dv = B$.

   Are there conditions on $\{h_n\}_{n\geq 1}$ s.t. $\hat{f}_X(x) \xrightarrow{p} f_X(x)$? Justify your answer.

2. Consider the following alternative to Theorem 3.1 in Handout 2.

   Theorem 1 (Asymptotic MSE and consistency, version 2)**.** *Let $\{X_i\}_{i\leq n}$ be an i.i.d. sample distributed according $f_X(x)$. Furthermore, assume that:*

   (a) $K : \mathbb{R} \to \mathbb{R}$ is a bounded function that satisfies $\int K(v)dv = 1$, $\int vK(v)dv = 0$, $\int v^2|K(v)|dv < \infty$, $\int v^2 K(v)^2 dv < \infty$, $\int v^2 K(v)dv = A > 0$, $\int K(v)^2 dv = B < \infty$ (i.e. a bounded second-order kernel). Also, $K(u) = 0$ for all $|u| > C$ for some $C < \infty$.

   (b) The sequence of bandwidths $\{h_n\}_{n\geq 1}$ satisfies $h_n \to 0$ and $nh_n \to \infty$.

   (c) Within an arbitrary small neighborhood of x, f_X is bounded and the second-order derivatives of f_X are continuous.

   Then,

   (a) $E[\hat{f}_X(x) - f_X(x)] = h_n^2 \times [f''_X(x)A/2 + o(1)]$.

   (b) $V[\hat{f}_X(x)] = (nh_n)^{-1} \times [f_X(x)B + o(1)]$.

   (c) $E[(\hat{f}_X(x) - f_X(x))^2] = h_n^4 (f''_X(x))^2 A^2/4 + (nh_n)^{-1} f_X(x)B + o(h_n^4 + (nh_n)^{-1})$.

   (d) $\hat{f}_X(x) - f_X(x) \xrightarrow{p} 0$.

   Relative to Theorem 3.1, this result strengthens the assumptions on the kernel function K and weakens the assumptions on the density function f_X.

*Email: federico.bugni@duke.edu. Please let me know if you think you found a typo or a mistake.

(a) Prove part (a). Be as detailed as possible.

Comment: An analogous argument can be used to prove (b). From here, the proofs of (c) and (d) follow from the same arguments used to show Theorem 3.1.

(b) Evaluate the validity of the assumptions of Theorem 1 for the following kernel functions: uniform, Gaussian, and Epanechnikov.

3. Reproduce the analysis in DiNardo and Tobias (2001). To this end, download the Current Population Survey (CPS) Merged Outgoing Rotation Groups (MORS) dataset for 1979 and 1989.[1] To make the data comparable across years, use the consumer price index (CPI) to express wages in U.S. dollars of the year 2000.[2]

Based on this data, complete the following questions.

(a) Estimate the density function of female log wages for each year assuming that these are normally distributed. Compare results across years.

(b) Estimate the density function of female log wages for each year using nonparametric kernel density estimation. Implement the estimation using a second-order Gaussian kernel and the following bandwidth choices: plug-in based on Gaussian model and cross-validation.

(c) Briefly compare results across years and bandwidth choices. Also, compare nonparametric estimation with parametric results under normality.

4. Li and Racine (2007), Exercise 1.7.

5. Li and Racine (2007), Exercise 1.8.

6. Let $\{X_i\}_{i=1}^n$ be an i.i.d. sample from a one dimensional distribution $f_X(x)$ with support $S_X(x) = [0, \infty)$ and $f_X(0) > 0$ (e.g. exponential distribution). By definition, the density function $f_X(x)$ is discontinuous at $x = 0$. In this problem, we are interested in estimating $f_X(x)$ at the point of discontinuity $x = 0$.

Suppose that we estimate $f_X(0)$ using a nonparametric kernel density estimator $\hat{f}_X(0)$ with "standard" choices of bandwidth and second-order kernel functions, i.e.,

(i) $K : \mathbb{R} \to \mathbb{R}$ is a bounded function that satisfies $\int K(v)dv = 1$, $\int vK(v)dv = 0$, $\int v^2|K(v)|dv < \infty$, $\int v^2K(v)dv = A > 0$, and $\int K(v)^2dv = B < \infty$.

(ii) $\{h_n\}_{n\geq 1}$ satisfies $h_n \to 0$ and $nh_n \to \infty$.

Based on this setup, answer the following questions.

(a) Characterize the asymptotic behavior of the mean and the variance of $\hat{f}_X(0)$.

(b) Is $\hat{f}_X(0)$ a consistent estimator of $f_X(0)$?

(c) The results covered in class do not apply. Explain why.

Comment: This issue is commonly known as "boundary bias".

7. Let $\hat{f}_X(x)$ be a one-dimensional kernel density estimator of with kernel function $K : \mathbb{R} \to \mathbb{R}$.

(a) First, assume that K is non-negative and $\int K^2(u)du < \infty$.

i. Show that $\hat{f}_X$ is nonnegative, i.e., $\hat{f}_X(x) \geq 0$ for all $x \in \mathbb{R}$.

[1] http://www.nber.org/morg/annual/

[2] http://www.usinflationcalculator.com/inflation/consumer-price-index-and-annual-percent-changes-from-1913-to-2008/

ii. Show that $\hat{f}_X$ integrates to one, i.e., $\int \hat{f}_X(x)dx = 1$.

iii. Conclude that $\hat{f}_X$ is a density function.

(b) Second, assume that K is a symmetric kernel function. Show that if $\int |u^s K(u)|du < \infty$ then $\int u^h K(u)du = 0$ for any odd h such that $h \leq s$.

Show that $\int u^h K(u)du = 0$ for any odd h such that $h \leq s$.

(c) Third, assume that K is a kernel function of order $s > 2$.

i. Show that K cannot be non-negative.

ii. Conclude that $\hat{f}_X$ may not be density function.

References

DiNardo, J. and J. L. Tobias (2001): "Nonparametric Density and Regression Estimation," *The Journal of Economic Perspectives*, 15, 11–28.

Li, Q. and J. S. Racine (2007): *Nonparametric Econometrics: Theory and Practice*, Princeton University Press.