ADAPTIVE KERNEL DENSITY ESTIMATION USING BETA KERNEL

XUN-FU YIN[1], ZHI-FENG HAO[2]

[1]College of Computer Science and Engineering, South China University of Technology, Guangzhou 510640, P.R. China
[2]School of Mathematical Science, South China University of Technology, Guangzhou 510640, P.R. China
E-MAIL: xunfuyin@yahoo.com.cn, mazfhao@scut.edu.cn

Abstract:

Adaptive kernel estimation for unit interval compact bounded densities using beta kernel is considered. Beta kernel is an asymmetric kernel that has several particular properties such as variable kernel shapes and matching the unit interval support of the densities to be estimated. These properties make beta kernel appropriate to be used to estimate the densities considered here because of the free of boundary bias. However, the optimal fixed bandwidth beta kernel estimator is often under smoothed. In this paper, a local adaptive scheme is put forward to improve the performance of the estimator, amending smoothness at the cost of a little of optimality. The results of the simulation studies demonstrate the effectiveness of our proposal.

Keywords:

Kernel density estimation; Beta kernel; Boundary bias; Bandwidth selection; Bandwidth function; Asymmetric kernel

1. Introduction

When applying traditional kernel density estimation method to deal with densities with bounded support, it is always observed that the estimation can fail dramatically in the boundary region. This phenomenon is termed "boundary bias" in the context of kernel density estimation. In order to remove boundary bias, researchers have proposed some approaches such as data reflection [1], boundary kernel [2], pseudo-data [3], transformations [4] and local polynomial regression [5, 6, 7].

Well-shaped densities such as bell shaped ones can be estimated satisfactorily by conventional kernel estimators, for example using the favorite Gaussian kernel. Nevertheless, in practical experience there always exist abundant abnormal densities including U-shaped, extremely skewed, multimodal peaky (e.g. claw) etc. For instance, the densities of recovery rates (percentage) in finance tend to be skewed and have a concentration of observations near the boundaries and the U-shaped density of beta distribution with parameters (0.2812, 0.2812) has unbounded densities at both endpoints. In this instance, the performance of the original kernel density estimators using symmetric kernels deteriorated greatly, especially when the densities are bounded.

The proposition of asymmetric kernel compensates for the deficiency stated above of the conventional kernel density estimators because they are quite useful for the estimation of density functions with bounded support. Beta kernel and gamma kernel are two representative nonnegative asymmetric kernels with flexible shapes. Gamma kernel is designed to handle density functions whose supports are bounded form one end only, whereas beta kernel is particularly convenient for the estimation of density functions with compact supports. In this paper, we limit our concentration to beta kernel.

The rest of this paper is organized as follows. In section 2 beta kernel and density estimation based on it are introduced. Then the adaptive scheme to improve the performance of fixed bandwidth beta kernel density estimator is suggested in Section 3. Section 4 presents results from two simulation studies and Section 5 concludes.

2. Density estimation using beta kernel

2.1. Introduction to basic concept of kernel density estimation

Let i.i.d. data $X_1,\cdots,X_n$ be a random sample from a distribution with an unknown univariate probability density f. A standard kernel density estimator for f is

$$\hat{f}(x)=\frac{1}{nh}\sum_{i=1}^{n}K\left(\frac{x-X_i}{h}\right) \quad (1)$$

where n is the number of the observations, h is the bandwidth parameter controlling the smoothness of the estimate ($h\to 0$ with $nh\to\infty$ as $n\to\infty$), K is the kernel function, satisfying the following conditions [8, 9, 10]:

$$\int K(u)du=1,\int uK(u)du=0,\int u^2K(u)du=\sigma_K^2>0 \quad (2)$$

The standard kernel estimator (1) was developed primarily for densities with unbounded support. The kernel function K is usually symmetric and is regarded as less important than the smoothing bandwidth. While using a symmetric kernel is appropriate for fitting densities with unbounded supports, it is not adequate for densities with bounded supports as it causes boundary bias. The reason is as follows. Symmetric kernels engender boundary bias due to the allocation of non-zero weight outside the support when smoothing is carried out near the boundary. On the contrary, a proper asymmetric kernel never assigns weight outside the density support and therefore can produce better estimates of the densities near the boundary.

2.2. Properties of beta kernel

Recently, Chen [11] proposed a beta kernel density estimator for densities that are defined on $[0,1]$. This estimator is based on the asymmetric beta kernel which exhibits two special appealing properties: a flexible shape and the location on the unit interval. The kernel shapes are allowed to vary with respect to the kernel center (see Figure 1) and bandwidth parameter (see Figure 2). Thus, the estimators can change the degree of smoothing in a natural way. In addition, the supports of beta kernels match the supports of the probability densities to be estimated. This leads to larger effective sample size used in the density estimation process and usually induces density estimates to have smaller finite-sample variances than other nonparametric estimators. The beta kernel density estimator is simple to implement, free of boundary bias, always nonnegative, and achieves the optimal rate of convergence ($O(n^{-\frac{4}{5}})$) for the mean integrated squared error (MISE) within the class of nonnegative kernel density estimators. Therefore, beta kernel is appropriate to be used to estimate densities defined in unit interval.

2.3. Beta kernel density estimator

Consider densities defined on the unit interval. Let i.i.d. data $X_1,\cdots,X_n$ be a random sample from an unknown univariate probability density function f defined on its support $A=[0,1]$. Chen [11] firstly introduced the beta kernel density estimator, which has the following form:

$$\hat{f}(x)=\frac{1}{n}\sum_{i=1}^{n}K_{x/h+1,(1-x)/h+1}(X_i) \tag{3}$$

where

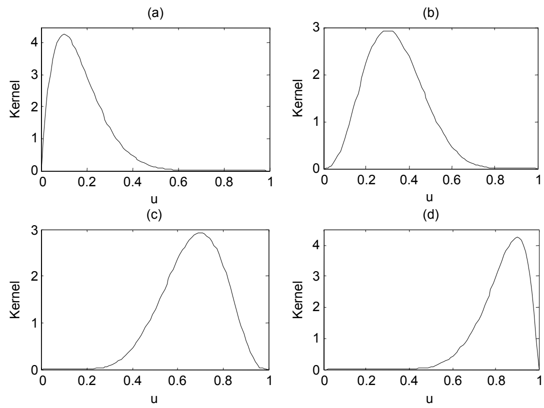

Figure 1. Flexible shapes of beta kernels with respect to variable kernel centers (*a:0.1; b:0.3; c:0.7; d:0.9*). The parameter h takes the value of 0.1

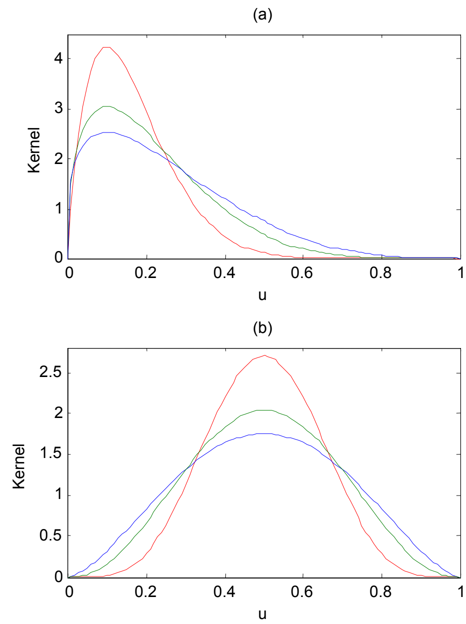

Figure 2. Variable shapes of beta kernels centered at *(a) 0.1* and *(b) 0.5* corresponding to different bandwidth parameters. (*solid-0.1; dashed-0.2; dot-dashed-0.3*)

$$K_{x/h+1,(1-x)/h+1}(u)=\frac{u^{x/h}(1-u)^{(1-x)/h}}{B(x/h+1,(1-x)/h+1)}I(0\le u\le 1) \tag{4}$$

is the beta kernel, and

$$B(a,b)=\int_0^1 t^{a-1}(1-t)^{b-1}\,dt \tag{5}$$

is the beta function. As in any kernel density estimation, the smoothing bandwidth h converges to zero as the sample size grows ($h\to 0$ as $n\to\infty$). Because arbitrary beta kernel defined as (4) centered at any point within the unit interval and with any positive parameter h ($0 < h \le 1$) integrates to 1, the kernel estimator (3) also integrates to 1. The restriction of $0 < h \le 1$ is added just because in our empirical study when $h > 1$, the shape of the beta kernel function will expand as h increases and becomes no longer appropriate to use in estimation.

The bias of the beta kernel estimator is

$$Bias[\hat{f}(x)] = [(1-2x)f'(x) + \tfrac{1}{2}x(1-x)f''(x)]h + O(h) \quad (6)$$

It is obvious that the bias is $O(h)$ uniformly for all $x \in [0,1]$, which means the free of boundary bias.

3. Adaptive beta kernel density estimator

Nonparametric density estimation requires the specification of the smoothing parameter. The performance of all kernel density estimation methods is crucially dependent on the specification of the smoothing parameter. For the kernel density estimator with a fixed bandwidth, it is difficult to estimate heavy-tailed densities (has spurious bumpiness in the tail region) and highly variable densities. Many theoretical and practical studies have been carried out for choosing the optimal smoothing parameter for the standard kernel estimators.

Interactive analysis of a data set always results in a desire to use different amounts of smoothing in different locations. It is often the case that estimators with different bandwidths reveal different interesting structure. So, intuitively, an estimator that uses different amounts of smoothing in different locations could mimic each of these estimates where they perform well. Many approaches concerning local adaptive kernel density estimation have been developed, including the nearest neighbor method [12], the transformation method [13], the shifted kernel method [14] and the variable kernel method [15], etc.

In the context of the beta kernel density estimation, the choice of the parameter h is a key issue. As shown in Figure 2, the parameter h has great impact on the shape of the beta kernel not in width (all the widths of any beta kernels are 1) but in height. This is different form the bandwidth parameter of symmetric kernels which only affects the width within which the kernel works (the heights are invariable). Hence in the context of beta kernel estimate, it is more exactly to term h as bandheight parameter than bandwidth parameter. But according as convention, we don't change the term. Chen [11] derived that the optimal bandwidths of the beta kernel estimator are $O(n^{-\frac{2}{5}})$ as compared with $O(n^{-\frac{1}{5}})$ for the other kernel estimators. In this section, we try to use varying bandwidths rather than a fixed bandwidth to do estimation.

In general, we suggest that applying smaller parameter in steep regions, whereas applying larger parameter in flat regions. We detect via experiments that when h becomes larger (restricting h varies within an appropriate range), the beta kernel estimator can catch the overall characteristics of the real density more exactly and becomes more smooth. However, the bias between the estimator and the real density in some regions (especially in the regions which have local or global extrema) increase along with h. On the other hand, though we can seek out the optimal value of h (through data based cross validation method or minimizing the MSE) when applying the fixed bandwidth beta kernel estimator, the value of the parameter is always so small that the estimator is under smoothed and has undesirable bumpiness. Consequently, based on the above discussion, we propose a novel scheme to apply the technique of beta kernel estimate. Firstly, produce a pilot estimate via a fixed bandwidth beta kernel density estimator using a relatively biggish h in order to catch the global characteristics of the real density appropriately. Secondly, we collect some information about the real density from the pilot estimate. The information includes the extrema and the steepness which can be used to calibrate the fixed bandwidth beta kernel estimator. Lastly, we get the adaptive beta kernel density estimator which incorporates the above useful information about the real density. In essence, the scheme aforementioned is a comparatively appropriate trade-off between smoothness and optimality (in the sense of the minimization of MSE).

The algorithm is as follows:

Step 1: Obtain a pilot estimate $\tilde{f}(x)$ by beta kernel density estimator with a global constant bandwidth which has a appropriate large value h_0.

Step 2: Set local bandwidth factors $\lambda(x)$ as

$$\lambda(x) = \gamma\tilde{f}(x) + \eta\left|\tilde{f}(x+\alpha) - \tilde{f}(x-\alpha)\right| / 2\alpha \quad (7)$$

where γ and η are weight parameters satisfying $\gamma + \eta = 1$, α is a little positive number (e.g. 0.01).

Step 3: Compute the adaptive beta kernel density estimate defined by

$$\hat{f}(x) = \frac{1}{n}\sum_{i=1}^{n} K_{x/\bar{h}\lambda(x)^{-1}+1,(1-x)/\bar{h}\lambda(x)^{-1}+1}(X_i) \quad (8)$$

where $\bar{h}$ can take the same value as the one used in the pilot estimate or another relatively biggish value.

4. Simulation studies

4.1. Case 1: Kernel estimate of Nassau County racial distribution data

We firstly apply the adaptive scheme of beta kernel density estimator discussed above to a practical case. The data set that reports the racial distribution of students in the 56 public school districts in Nassau County in the 1992-1993 school year used here comes from Chatterjee et.al [16].

There are only two variables in this data set. One is the proportion of white students; the other is the corresponding district. Figure 3 represents an overlay of three fixed bandwidth kernel density estimates of the racial distribution data applying beta kernel with different bandwidths. The points along the bottom of the plot give the positions of the observations. Obviously, the density is highly left skewed with heavy left tail. Moreover, the use of relatively small bandwidth has caused apparent spurious bumpiness within the tail and a little boundary bias near 1 on the right, whereas the use of relatively large bandwidth may induce great bias in regions nearby the extrema. Fortunately, the scheme recommended in Section 3 can conquer the above deficiencies. Figure 4 illustrates an instance of the realization of the scheme with parameters $h_0 = 0.2$, $\overline{h} = 0.1$, $\alpha = 0.005$, $\gamma = 0.9$, $\eta = 0.1$.

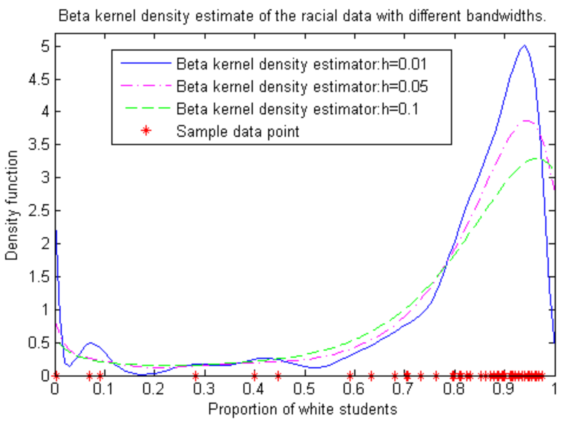

Figure 3. Beta kernel density estimators for the racial data

4.2. Case 2: Monte Carlo simulation of a beta distribution

In Case 1, the data are obtained from the real world and the underlying density is unknown, so it is somewhat difficult to evaluate the performance of variable estimators.

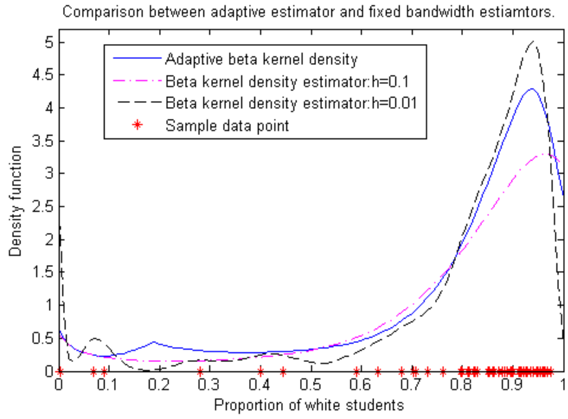

Figure 4. Comparison between an adaptive beta estimator and two fixed bandwidth estimators

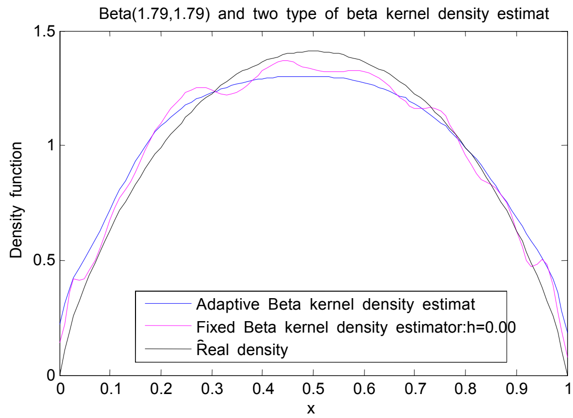

Figure 5. Comparison between an adaptive beta estimator, a fixed bandwidth estimator and real density

However, our interest is to acquire the overall appropriate information about the distribution of the proportion of white students rather than to get the mathematical formula of the density. In this sense, the estimators which catch the important characters such as boundary property and multimodality can be judged to be appropriate and effective ones and can be accepted. In Case 2, we generate 1000 random sample data points from the Beta distribution with parameters (1.79, 1.79). The method that minimizing pointwise MSE was utilized to choose the optimal smoothing bandwidth in the fixed bandwidth beta kernel density estimate, and prescribed $h = 0.008$. As case 1, the optimal parameter for fixed bandwidth beta kernel estimator is so small that some undesirable bumpiness is induced. Again, we apply the proposed scheme. This time, the specification of parameters is as follows: $h_0 = 0.05$, $\overline{h} = 0.05$, $\alpha = 0.01$, $\gamma = 0.5$, and $\eta = 0.5$. Figure 5 is a overlay of three curves, the solid line represents the adaptive beta kernel estimate with the above specification, the dashed line represents the real density, and the dot-dashed line represents the fixed bandwidth beta kernel estimate with the optimal parameter. We can assert that the adaptive beta kernel estimator in Figure 6 is smooth enough and have obtained adequate characteristics of the real density.

5. Conclusions

In this paper, we have examined a nonparametric estimation of densities defined in the unit interval utilizing a local adaptive scheme of beta kernel density estimate. Firstly, the advantage of beta kernel density estimator consists in the ability to appropriately grip on the overall characteristics of the densities. Secondly, in comparison with the conventional kernel density estimator based on symmetric kernels, the estimator based on beta kernel is free of boundary bias and spurious bumpiness.

A respect deserving attention is that fixed bandwidth beta kernel estimators can produce considerable variation in the shapes of the estimates owing to the change of parameter h. This is due to the intrinsical characteristic of beta kernel which shape is intimately connected with h. The densities considered here are limited to be bounded in the unit interval. In fact, the adaptive scheme of beta kernel density estimator can be generalized to dispose any density with compact support.

Acknowledgements

This work has been supported by the National Natural Science Foundation of China (10471045, 60433020), Natural Science Foundation of Guangdong Province (970472, 000463, 04020079), Excellent Young Teachers Program of Ministry of Education of China, Fok Ying Tong Education Foundation (91005), Social Science Research Foundation of MOE (2005-241), Key Technology Research and Development Program of Guangdong Province (2005B10101010), Key Technology Research and Development Program of Tianhe District (051G041) and Natural Science Foundation of South China University of Technology (D76010).

References

[1] Schuster, E.F., "Incorporating support constraints into nonparametric estimators of densities." Commun. Statist, Vol 14, pp. 1123-1136, 1985.

[2] Müller, H.G., "Smooth optimum kernel estimators near endpoints", Biometrika, Vol 78, pp. 521-530, 1991.

[3] Cowling, A. and Hall, P, "On pseudodata methods for removing boundary effects in kernel density estimation." J. Roy. Statist. Soc., Vol 58, pp. 551-563, 1996.

[4] Marron, J.S. and Ruppert, D, "Transformation to reduce boundary bias in kernel density estimation." J. Roy. Statist. Soc., Vol 56, pp. 653-671, 1994.

[5] Lejeune, M. and Sarda, P., "Smooth estimators of distribution and density functions.", Comput. Statist. Data. Anal, pp. 457-471, 1992.

[6] Jones, M.C., "Simple boundary correction for kernel density estimation.", Statist.Comp.3, pp. 135-146, 1993.

[7] Jones, M.C. and Foster, P.J., "A simple nonnegative boundary correction method for kernel density estimation.", Statist. Sin., Vol 6, pp. 1005-1013, 1996.

[8] Akaike, H., "An approximation to the density function.", Annals of the Institute of Statistical Mathematics, Vol 6, pp. 127-132, 1954.

[9] Rosenblatt, M., "Remarks on some nonparametric estimates of a density function.", Annals of Mathematical Statistics, Vol 27, pp. 832-837, 1956.

[10] Parzen, E., "On the estimation of a probability density function and the mode.", Annals of Mathematical Statistics, Vol 33, pp. 1065-1076, 1962.

[11] Chen S., "Beta kernel estimators for density functions.", Computational Statistics & Data Analysis, Vol 31, pp. 131-145, 1999.

[12] Loftsgaarden, D.O. and Quesenberry, C.P., "A nonparametric estimate of a multivariate density function.", Annals of Mathematical Statistics, Vol 36, pp. 1049-1051, 1965.

[13] Wand, M.P., Marron, J.S. and Ruppert, D.R., "Transformations in density estimation.", Journal of the American Statistical Association, Vol 86, pp. 343-361, 1991.

[14] Samiuddin, M. and El-Sayyad, G.M., "On nonparametric kernel density estimates.", Biometrika, Vol 77, pp. 865-874, 1990.

[15] Breiman, L., Meisel, W. and Purcell, E., "Variable kernel estimates of probability densities.", Technometrics, Vol 19, pp. 135-144, 1977.

[16] Chatterjee, S., Handcock, M.S., and Simonoff, J.S., A Casebook for a First Course in Statistics and Data Analysis, John Wiley, New York, 1995