GLOBAL JOURNAL OF PURE AND APPLIED SCIENCES VOL. 13 NO. 1 , JANUARY, 2007 : 1 07 - 111
COPYRIGHT© BACHUDO SCIENCE CO. LTD PRINTED IN NIGERIA. ISSN 1118 - 0579

APPLICATION OF ADJUSTABLE KERNEL METHOD IN TEST EVALUATION

E. J. OSEMWENKHAE, and C. C. ISHIEKWENE

(Received 25 May, 2006; Revision Accepted 28 August, 2006)

ABSTRACT

This work is motivated by the appearance of heavy tails when estimating learners' performances in most psychological tests. Explaining this appearance of heavy tails poses a problem. Over-smoothing the target density by using large bandwidth may tend to remove this abnormal tail behaviour, but this result in masking some essential features in the distribution. Conversely, smaller bandwidth can result in under-smoothing of the true density and thus gives a biased estimate of the density under study. Therefore, in any psychological test involving heavy tails, we observe that adjustable kernel method is better for fitting this density than its fixed kernel method. However, when heavy tails are not suspected, fixed kernel method may be easier in assessing learners' performances.

KEYWORDS: Psychological test, target density, bandwidth, adjustable kernel and heavy tails

INTRODUCTION

Suppose $X_1, X_2, \ldots, X_n$ is a random sample from a probability function, $f(x)$ Generally speaking, the fixed univariate kernel estimator, $\hat{f}(x)$, introduced by Rosenblatt (1956), is given by

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

(1.1)

where the kernel function, k, satisfies $\int_{-\infty}^{\infty} k(t)dt = 1$ and h, the optional window width is assumed fixed all through the process of estimation. Literally, the kernel estimator is a sum of bumps placed at the observation X_i, i = 1(1) n. The 'shape of the bumps' are determined by the kernel function k used while the window width h determines the width of the bumps. This h has been seen to be more crucial than the shape of the kernel except in some very rare cases see, for instance, Hansen (2003), Osemwenkhae and Ogbonmwan (2003), etc.

The primary objective of this paper is to assess the performance of the fixed univariate kernel in learners' evaluation vis-à-vis the adaptive univariate kernel estimator focusing on target densities with heavy tails. This work is motivated by the appearance of this type of heavy tail density in finance, archaeology, hydrology, genetics and astronomy, see Baxter (2000) et al, Dinardo and Tobias (2001), Kim and Heo (2002), Dinardo et al (1996), etc. It is particularly necessary to find other methods that can perform well for heavy tails which most times is associated with most psychological test (see Kline (1986, 2000)).

Section two will be concerned about the nature and failure of the fixed kernel method in estimating such densities. Section three will examine briefly some other adjustable methods and their failures. Emphasis on the adjustable kernel method and how this method corrects the failures of these other methods will be examined in section 4.

The Fixed Kernel Method

The "nature" of the smoothing parameter, h, is the basis for this classification. By fixed we mean, the smoothing parameter in (1 1) is constant all through the construction of $\hat{f}$ from the sample values $X_1, \ldots, X_n$. Otherwise it is adjustable or adaptive. Generally, the fixed kernel method is insensitive to local peculiarities in the data; such as data clumping in certain areas and data sparseness in others.

The global accuracy of $\hat{f}$ can be evaluated from the Mean Integrated Square Error (MISE) given by:

$$MISE\,\hat{f}(x) = E\int\left[\hat{f}(x) - f(x)\right]^2 dx$$
$$= \int (Bias)^2 f(x)dx + \int \text{var}\, f(x)dx \quad (2\ 1)$$

Since the smoothing parameter h has played a vital role in minimizing the MISE above, so many higher order versions of (2 1) have been proposed with their accompanying window widths; all aimed at reducing the global error – MISE; see Wand and Jones (1995).

The choice of kernel to use does not pose much problem. Although, for higher order window width (say, h^4, h^6) the work of Jones and Signorini (1997) and Osemwenkhae and Ogbonmwan (2003), have shown that some kernels are preferred over the others For the second order window width, Silverman (1986) obtained that if the kernel choice is the Gaussian and the data set is normal or almost normal, the appropriate window width is

$$h_{opt} \approx 1.06\hat{\sigma}n^{-\frac{1}{5}} \quad (2\ 3)$$

where $\hat{\sigma}$ is the standard deviation of the data set

Many methods of selecting h exist in literature the plug-in bandwidth selection method by Sheather and Jones (1991) and Wand and Jones (1995), free–hand or subjective choice by Silverman (1986), cross-validation

E. J. Osemwenkhae, Department of Mathematics, University of Benin, Benin-City, Nigeria.
C. C. Ishiekwene, Department of Mathematics, University of Benin, Benin-City, Nigeria.

scheme suggested by Rudemo (1982) and Bowman (1984), test graph method of Silverman (1978), bootstrap choice, suggested by Taylor (1989), etc. Among these methods, the least squares cross-validation scheme is the mostly studied because of its attractive asymptotic property of giving an answer that converges to the optimum under very weak conditions – see Stone (1984) and Marron (1993). However in many simulation studies and real life examples, Hall and Marron (1987) showed that the performance of this method has been often disappointing since it suffers from sample variability.

The work of Wand and Jones (1995, pp. 60) however, suggested that a quick way of choosing the smoothing parameter especially if the data is close to normal would be to estimate $\hat{\sigma}$ from the data and then substitute it back to (2.3). This only suggests the use of a fixed bandwidth h. The consequence of using this method in (2.3) that is permitting h to be fixed in the entire course of estimating the density of the scores of 80 undergraduate students who where given some sets of validated psychological test (Kline 1986, 2000) are displayed below. For this case, the sample standard deviation is 13.03, and the optimal h corresponding to this case is given approximately as 5.6474 (Figure 1a). Figures 1b, 1c and 1d were obtained by multiplying approximate h value (5.6474) by ½, ¼ and 2 respectively. This corresponds to choosing h subjectively as suggested in Silverman (1986, pp44).

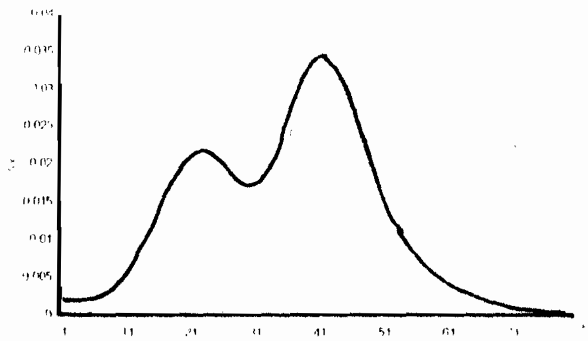

Figure 1a: Density of the scores of 80 students when h=5.6474

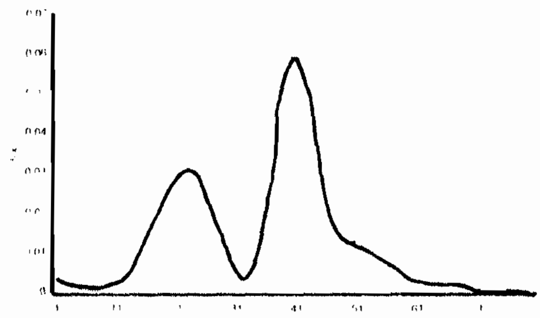

Figure 1b: Density of the scores of 80 students when h = 2.823

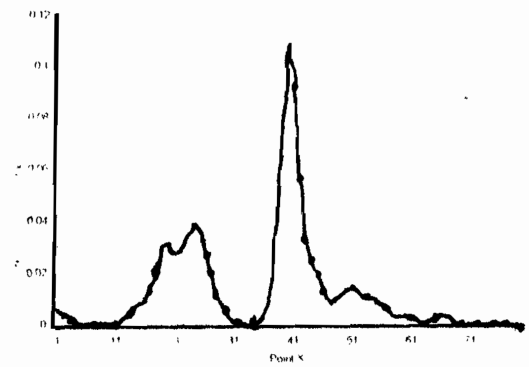

Figure 1c: Density of the scores of 80 students when h=1.4118

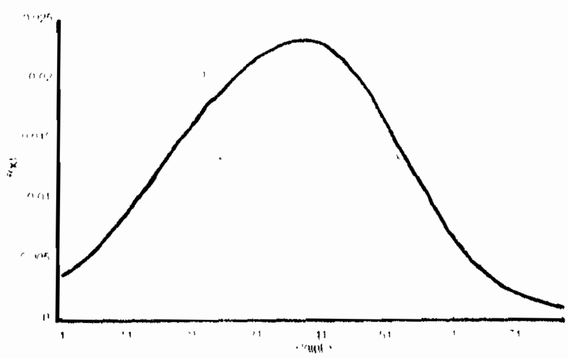

Figure 1d: Density of the scores of 80 students when h = 11.948

One setback of the fixed kernel method is the fixed nature of the window width throughout the entire process of estimation. The consequence of this is the appearance of spurious noise at the tails of the distribution where the observations corresponding to either the dull or intelligent students falls (see the tails of Figure 1c). Although over- smoothing by using large h, see Figure 1d, may tend to remove this, but this results in masking some essential details (like bimodality in 1a) in the distribution.

Intuitively, the larger the bandwidth (lower variance), the smoother the resulting estimates. Conversely, smaller bandwidth (see Figure 1c) can result in under smoothing of the true density and thus gives a biased estimate of the density under study. This makes the fixed kernel method a 'very bad' estimator at the tails of a distribution. This work would tend to handle this pitfall.

The Adaptive Methods

Convincingly, the importance of the tails of a distribution cannot be overemphasized. Silverman (1986, Section 4.5) showed that in a ten-dimensional standard normal distribution, 99% of the mass of the distribution is at points whose distance from the origin is greater than 1.6 of the standard normal distribution. So, the need to have a better way of analyzing the behaviour of densities at the tails becomes imperative.

The failure of the fixed method in estimating densities shown above at the tails (Figures 1c and 1b necessitated a study into finding a way of handling this. This method suggests the use of a varying bandwidth instead of the traditional fixed method of Section two.

Some adjustable methods of interest to us in this work are the adjustable histogram, the nearest neighbourhood, the length biased data approach and the adaptive kernel schemes.

The adjustable histogram has been used to provide a visual clue to underlying distribution of f, see Izenman (1991). Suppose f has support $\Omega = [a,b]$, partition $[a,b]$ into non-overlapping bin widths given by h_n, where $h_n = (t_{n,i+1} - t_{n,i})$ for $i = 1(1)m$ and m is the number of observations in bin i. If $I_i(x)$ is the indicator function for the i^{th} bin, then the histogram estimator is given as

$$\hat{f}(x) = \frac{1}{nh_n}\sum_{i=1}^{m} N_i I_i(x) \quad (3.1)$$

where $n = \sum_{i=1}^{m} N_i$ is the sample size, N_i is the size of the ith sample.

Basically, the choice of origin and the length of the bin h_n, affect its smoothing procedure. The histogram estimator (3.1) lacks accuracy when used in cluster analysis and nonparametric discriminant analysis, see Silverman (1986), and also lacks continuities at cell boundaries when derivatives of estimates are required, see Hand (1982). Another major pitfall of the histogram estimator is that it does not allow the drawing of contour diagram in the representation of data and so it does not work well in multivariate data, see Tukey and Tukey (1981). The sensitivity of the histogram's shape to the choice of origin is a more serious defect as stated in Silverman (1986, pp10) and Devroye and Lugosi (1997, 2001).

Other methods of adaptive origin include the nearest neighbourhood (NN), the maximum penalized likelihood (MPL) and the length biased data approach (LBDA). The failure of the nearest neighbourhood method, as pointed out in Bowman and Foster (1993), is the tendency for the estimators to exhibit jagged peaks, and pitifully the complicated function of x obtained does not integrate to 1. Fundamentally, these methods among other things, failed to be a proper probability density function, see Silverman (1986) and Patil et al (1991). So, using it in a long tailed data is risky.

The Adaptive Kernel Scheme

From Section 2, the problem of either over smoothing or under smoothing whether at the main part or tails of the distribution is evident. This tends to affect the variance and bias in one form or the other. In this section, we suggest varying the bandwidth along the support of the sample data to allow flexibility which in turn will reduce the variance of the estimates in areas with fewer observations, and reducing the bias of the estimates in areas with many observations (see an earlier suggestion in Silverman (1986), Wand and Jones (1995) and Salgado – Ugarte and Perez Hermandez (2003) and Sheather (2004).

The Adaptive Kernel Density Estimator (AKDE) is obtained by modifying (1.1) to reflect and correct the pitfall of the fixed kernel method. The Adaptive Kernel Density Estimator method is given by

$$\hat{f}(x) = \frac{1}{nh\lambda_i}\sum_{i=1}^{n} k\left(\frac{x - X_i}{h\lambda_i}\right) \quad (4.1)$$

where $\lambda_i = \left(\tilde{f}(X_i)/g\right)^{-\alpha}$, k( ) the kernel function, g the geometric mean of the pilot estimates, $\tilde{f}(X_i)$, corresponding to (1.1), α the sensitivity parameter and h the window width as usual.

Many suggestions exist as to the choice of the sensitivity parameter, α, ranging about 0.25 to 0.75, see Silverman (1986, pp100), Kerm (2003). Nevertheless, it is clear that if $\alpha = 1$, (4.1) reduces to (1.1). The sensitivity parameter together with the varying scaling parameter, $h\lambda_i$, play a significant role in this process. Here h controls the overall degree of smoothing while λ_i stretches or shrinks the sample points bandwidth so as to adapt to the density of the data.

This scheme attempts to investigate the lapses experienced by the fixed kernel method and many other adaptive schemes especially at the tails. This in turn will aid educational evaluators in estimating the very bright or the very dull students which usually constitute learners at both tails.

With the same data generating figures 1a – d, (4.1) is applied in estimating the density of this same data. We consider cases when α =0.5 and 0.25. Figures 2a – 2c are obtained when the above sensitivity parameters are used (where Figure 2c is the combination of Figures 2a & 2b).

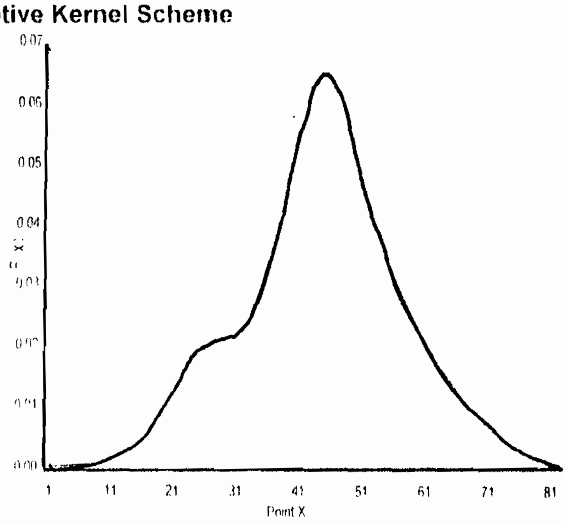

Figure 2a. Density of 80 students when α = 0.25

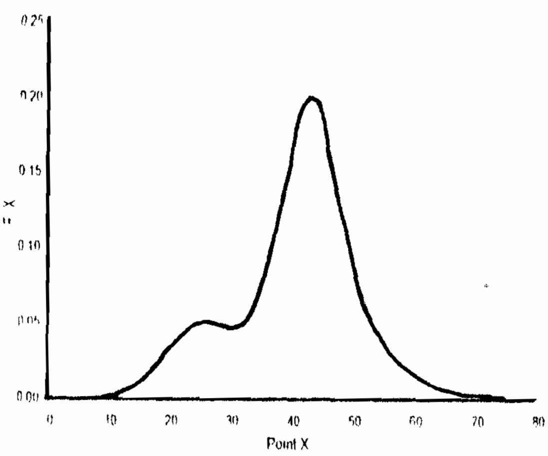

Figure 2b. Density of 80 students when α = 0.5

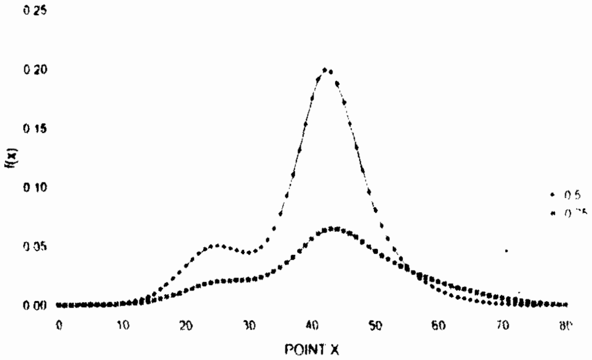

Figure 2c Density of 80 students when α = 0.50 and α = 0.25

DISCUSSION OF RESULTS

First, the excitement about this work is on its inherited pattern of being a proper probability density function (this is evident from figures 2a and 2b above). This is unlike most adjustable methods listed above in Section three. Comparing Figures 1b and 1c with Figures 2a and 2b, the spurious noise evident in the tails of Figures 1b and 1c, occasioned by the fixed nature of the smoothing parameter *h*, has been seriously handled. Therefore, (4.1) will be a better alternative at the tails when many observations are suspected to be around the tails.

So, in a psychological test, where learners' performance are independent and there is a heavy suspicion of very intelligent or very dull ones (necessitating heavy tails), a better estimation scheme would be the adjustable kernel method considering its huge benefits at the tails.

CONCLUDING REMARKS

When heavy tails are suspected in a set of any psychological test, a better estimation scheme would be the adjustable kernel method. Conversely, when heavy tails are not suspected, the adjustable kernel method easily decomposes to the fixed scheme if the sensitivity parameter, $\alpha = 1$.

REFERENCES

Baxter, M. I., Beardah, C. C. and Westwood, S., 2000. Sample Size and related issues in the analysis of lead isotope data. Journal of Archaeological Science 27: 973 – 980

Bowman, A. W., 1984. An alternative method of cross-validation for the smoothing of density estimates. Biometrika, 71: 353 – 360.

Bowman, A. W and Foster, P J., 1993. Adaptive smoothing and density-based tests of multivariate normality. Journ. of Amer Stat. Ass; 88: 529 –537.

Devroye, L. and Lugosi, G., 1997. Nonasymptotic universal smoothing factors, kernel complexity and yatracos classes. Ann. of Statist, 25: 2626 – 2637.

Devroye, L. and Lugosi, G., 2001 Combinational methods in density estimation. Springer – Verlag, New York.

Dinardo, J. and Tobias, J. L., 2001 Nonparametric density and regression estimation. Journal of Economic Perspectives, 15: 153 – 158.

Dinardo, J. Fortin, N. M and Lemieux, J., 1996. Labour market institutions and the distribution of wages, 1973 – 1992 A semiparametric approach. Econometrica, 64. 1001 – 1044

Hall, P. and Marron, J. S., 1987, Extent to which least-square cross-validation minimizes integrated square error in nonparametric density estimation. Prob. Theory and Related field, 74: 567 – 581

Hand, D., 1982. Kernel Discriminant Analysis. Research Studies, New York.

Hansen, 2003. Exact Mean Integrated Square Error of higher order kernel estimators. Joun. of Amer Stat. Ass., 87 1 – 16.

Izenman, A J., 1991. Recent developments in nonparametric density estimation. Journal of Amer. Stat. Ass., 86: 205 – 221

Jones, M. C. and Signorini, D. F., 1997 A Comparison of higher – order bias kernel density estimator Journ. of Amer Stat. Ass., 92: 1063 – 1073.

Kerm, V P., 2003. Adaptive kernel density estimation. 9th UK Stata users meeting, Royal Statistical Society, London.

Kim, K. D. and Heo, J. H., 2002. Comparative Study of flood quartiles estimation by nonparametric models. Journals of Hydrology, 260. 176 – 193.

Kline, P., 1986 Handbook of Test Construction London, Methuen

Kline, P., 2000. Handbook of Psychological Testing. London, Routledge

Marron, J. S., 1993. Discussion of practical performance of several data driven bandwidths selectors by Park and Turlach Comp Stat 8: 17 -19

Osemwenkhae J. E. and Ogbonmwan S M., 2003 Generalized efficiencies for higher- order symmetric univariate kernel. Journal of the Nig. Ass. of Math. Physics, 7: 89 – 96.

Patil, P. N., Wells, M. J. and Marron, J. S., 1991 Kernel based estimations of ratio functions Journal of Nonparametric Statistics, 1 223 – 231

Rosenblatt, M., 1956. Remarks on some nonparametric estimates of a density function Berlin, Springer-verlag, 181 – 190.

Rudemo, M., 1982. Empirical choice of histograms and kernel density estimators. Scand. Journal of Statistics, 9: 65 – 78

Salgardo – Ugarte, I. H and Perez Hermandez, M. A, 2003. Exploring the use of variable bandwidth kernel density estimators Stata Journal, Statacorp LP, 3:133 –147

Sheather, S J., 2004 Density Estimation Statistical Science, 4 588 – 597

Sheather, S J. and Jones, M C., 1991 A reliable data-based bandwidth selection method for kernel density estimation. Journ of Roy. Statist Soc. B, 53 683 – 690.

Silverman B. W. 1978 Choosing the window width when estimating a density. Biometrika, 65 1-11

Silverman B W., 1986 Density estimation for statistics and data analysis London, Chapman & Hall

Stone, C. J. 1984 An asymptotically optimal window selection rule for kernel density estimates Ann of Stat., 12 1285 – 1297

Taylor, C., 1989. Bootstrap choice of the smoothing parameter in kernel density estimation Biometrika, 76: 705 – 712

Tukey, P A. and Tukey, J. W, 1981 Graphical display of data sets in 3 or more dimensions Interpreting Multivariate Data. Chichester: Wiley 189 – 275

Wand, M. P and Jones, M C., 1995. Kernel Smoothing Chapman and Hall, London