Advances and Applications in Statistics

Published Online: April 2016
http://dx.doi.org/10.17654/AS048020109
Volume 48, Number 2, 2016, Pages 109-121 **ISSN: 0972-3617**

AN APPROACH FOR BIAS REDUCTION IN DENSITY ESTIMATION

Kee-Hoon Kang

Department of Statistics
Hankuk University of Foreign Studies
Yongin 449-791, Korea
e-mail: khkang@hufs.ac.kr

Abstract

Bias reduction in nonparametric function estimation is an important issue and enough work has been done on this topic. In this paper, we consider the data sharpening approach, which is a useful tool for enhancing statistical properties of conventional nonparametric curve estimators. We propose a simple numerical algorithm for data sharpening in density estimation to reduce biases. It does not require the estimation procedures of density and its derivatives. Numerical results reveal that our algorithm works well and is computationally efficient.

1. Introduction

Bias reduction in kernel density estimation is an important issue and there is an extensive literature on this matter. Jones and Foster [9, 10] stated

Received: December 3, 2015; Accepted: January 11, 2016
2010 Mathematics Subject Classification: 62G07.
Keywords and phrases: bandwidth, data sharpening, EM algorithm.
This work was supported by the Hankuk University of Foreign Studies Research Fund of 2015.
Communicated by K. K. Azad

that much of it can be seen as special cases of generalized jackknifing in which linear combinations of kernels are constructed to reduce bias. In addition, there are other approaches to bias reduction in density estimation. Jones et al. [11] proposed two-stage multiplicative bias correction method. DiMarzio and Taylor [4] used boosting approach. Mynbaev and Martins-Filho [14] proposed imposing global Lipschitz conditions on f. Hirukawa and Sakudo [7] applied two classes of multiplicative bias correction to density estimation.

Data sharpening for density estimation, which was initially named by Choi and Hall [1], is currently a popular research area owing to its simplicity with good properties. This method may be employed to perturb data so as to enhance properties of relatively conventional statistical procedures without modifying the estimator itself. This concept has been motivated by a geometric consideration in order to reduce biases in nonparametric curve estimation. The basic idea of this approach in density estimation is to move data values closer together near peaks and further apart near troughs. As a result, the majority of bias problems would vanish. Similar approaches were also considered in Samiuddin and El-Sayyad [16], Jones and Signorini [12] and Park et al. [15] in view of varying location.

Choi et al. [2] proposed the data sharpening method in nonparametric regression problems. They found some good properties of it, which involves adjusting both the explanatory and the response variables prior to substitution into a local linear estimator. It enhances resistance of the estimator to design sparseness problems and reduces bias by an order of magnitude. Hall and Minnotte [5] proposed the higher order data sharpening method in density estimation, which tends to result in arbitrarily high orders of bias reduction. Hall and Kang [6] used the cross-validation approach to get the sharpened data set from the original one, as well as a bandwidth, which is used in constructing the estimates with sharpened data.

In this paper, we will propose a simple numerical approach to data sharpening based on a parametric idea. We successively estimate the perturbation amount of each datum through EM like iteration. Finally, we

reconstruct the density estimate with the sharpened data, which can be obtained by adding perturbation amount to the original data. Unlike Hall and Kang [6], our approach does not need an additional tuning parameter for the determination of perturbation amounts. In addition, our approach does not need an estimation procedure for density and its derivatives in Samiuddin and El-Sayyad [16]. Unlike using higher order kernel, the resulting estimates by what we have proposed guarantee the nonnegativity of a density. Numerical properties of the proposed method reveal that our approach works quite well with larger bandwidth than that for ordinary kernel density estimator.

The rest of the paper is organized as follows. Section 2 gives a brief introduction of bias reduction through data sharpening in density estimation. An iterative algorithm for reducing bias is proposed in Section 3. Numerical properties of the proposed method are investigated in Section 4. Some discussion and concluding remarks are summarized in Section 5.

2. Bias Reduction in Density Estimation

Let $\mathcal{X} = \{X_i, i = 1, ..., n\}$ be a given random sample of size n from a smooth density f. The standard kernel density estimator of f at the point x takes the form

$$\hat{f}_{h\mathcal{X}}(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right), \tag{1}$$

where h is a bandwidth and K is a kernel function, which integrates to 1. Typically, K is assumed to be a smooth, bounded unimodal probability density function symmetric about zero. Then K satisfies the following conditions:

$$\int K(x)dx = 1,$$

$$\int xK(x)dx = 0,$$

$$\int x^2 K(x)dx = \kappa_2 > 0$$

and we assume the sample size $n \to \infty$ and the bandwidth $h = h(n) \to 0$. Then the basic properties of such a traditional estimator are well known and these include

$$E\{\hat{f}_{h\mathcal{X}}(x)\} = f(x) + \frac{1}{2}h^2\kappa_2 f''(x) + o(h^2),$$

$$Var\{\hat{f}_{h\mathcal{X}}(x)\} = (nh)^{-1}R(K)f(x) + o\{(nh)^{-1}\},$$

where $R(K) = \int K(z)^2 dz$. See for example, Wand and Jones [17, pp. 20-21].

Motivated by the principle in Samiuddin and El-Sayyad [16], consider perturbing X_i to Y_i for $1 \le i \le n$, forming a new data set $\mathcal{Y} = \{Y_i, i = 1, ..., n\}$, which we shall call the *sharpened data set*. Let $A(\cdot)$ be a function of each datum, which determines the perturbation amount and each Y_i has the following form:

$$Y_i = X_i + \frac{1}{2}h^2\kappa_2 A(X_i). \tag{2}$$

Define

$$\hat{f}_{h\mathcal{Y}}(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - Y_i}{h}\right) = \frac{1}{n}\sum_{i=1}^{n} K_h(x - Y_i),$$

which is the kernel density estimator based on the sharpened sample $\mathcal{Y}$ from (2) rather than the original sample $\mathcal{X}$. Then the expected value of the kernel estimator $\hat{f}_{h\mathcal{Y}}(x)$ has the following form:

$$\int K_h\left(x - t - \frac{1}{2}h^2\kappa_2 A(t)\right)f(t)dt$$

$$= \int K\left(u - \frac{1}{2}h\kappa_2 A(x - hu)\right)f(x - hu)du. \tag{3}$$

Taylor expansion of integrand in (3) and using $\int uK'(u)du = -1$ when we take Gaussian kernel gives

$$E\hat{f}_{h\mathcal{Y}}(x) = f(x) + \frac{1}{2}h^2\kappa_2[f''(x) - A'(x)f(x) - A(x)f'(x)] + O(h^4).$$

Hence, with $A(x)$ chosen as $(Af)' = f''$, which corresponds to $A(x) = f'(x)/f(x)$, we can achieve bias reduction up to order of h^4. This implies that bias reduction is achieved by moving data points in the direction of the sign of slope of the density at that point.

It is well known that if we use kernel of order *r* in equation (1), that is, *K* satisfies $\int_{-\infty}^{\infty} t^j K(t)dt = 0$ for $j = 1, ..., r-1$, then its bias is of order h^r. Therefore, using fourth order kernel results in the same bias reduction as before. However, this can be inconvenient because the resulting density estimate can no longer be nonnegative everywhere.

Then one can think of various versions of the estimator of $A(x)$, which is actually the derivative of $\log f$. One of the natural estimators is $\hat{A}(x) = \hat{f}_h'(x)/\hat{f}_h(x)$, which was considered in Samiuddin and El-Sayyad [16] and Jones and Signorini [12]. Park et al. [15] considered an estimator of $(\log f)'$ using local likelihood principle and plugged it into (2) in place of *A*. It was shown that the proposed approach has a better boundary property comparing to the former one. For constructing local likelihood function and its properties, one may refer to Loader [13] and Hjort and Jones [8].

3. An Iterative Algorithm

The approach in the last paragraph of previous section, which uses $\hat{A}(x) = \hat{f}_h'(x)/\hat{f}_h(x)$, has a drawback. It gives up working in the region where the density is very low since it needs estimates of density in the denominator. One possible alternative is simply finding δ_i, $i = 1, ..., n$,

which corresponds to the perturbation amount $X_i \to X_i + \delta_i$ through a simple computational algorithm. The proposed approach is based on regarding the problem at hand as a big parametric model. That is to say, when we consider the data sharpening approach, we can fit a model

$$\hat{f}(x \mid \delta_1, ..., \delta_n) = n^{-1} \sum_{i=1}^{n} K_h(x - X_i - \delta_i)$$

to data, as this may be regarded as a parametric approach.

Then log-likelihood based on this density can be represented

$$l(\underline{\delta}) = -n \log n + \sum_{i=1}^{n} \log\left\{ \sum_{j=1}^{n} K_h(X_i - X_j - \delta_j) \right\}. \tag{4}$$

In order to get maximum likelihood estimates of δ_i, $i = 1, ..., n$, we set the partial derivatives of (4) equal to zero. When using the Gaussian kernel function, this leads to

$$\sum_{i=1}^{n} \frac{(X_i - X_k - \delta_k) K_h(X_i - X_k - \delta)}{\sum_{j=1}^{n} K_h(X_i - X_j - \delta_j)} = 0 \tag{5}$$

for $k = 1, .., n$.

Let us denote

$$p(k \mid X_i) = \frac{K_h(X_i - X_k - \delta_k)}{\sum_{j=1}^{n} K_h(X_i - X_j - \delta_j)}, \tag{6}$$

which can be interpreted as the probability for X_i to belong to mixture component k. From (5), δ_k can be represented as

$$\delta_k = \frac{\sum_{i=1}^{n} p(k \mid X_i)(X_i - X_k)}{\sum_{i=1}^{n} p(k \mid X_i)}, \quad k = 1, ..., n. \tag{7}$$

This is not a closed form solution, in that $p(k \mid X_i)$ at each stage depends on all the δ_js, but iterating them in an EM manner would give the required $\hat{\delta}_1, \ldots, \hat{\delta}_n$. One may start this iteration with initial values of δs being set to be 0 and iterate until a convergence criterion is satisfied or one can try finite number of iterations to explore the amounts of changes.

4. Numerical Properties

This section is devoted to the comparison of the small sample performance of the ordinary kernel density estimator and two data sharpening approaches, one using the explicit formula in (2) by replacing $A(X_i)$ with $\hat{A}(X_i)$ and the other using the proposed algorithm in (7). We drew data sets of size 50 from the standard Gaussian density (#1), Student's *t*-distribution with degree of freedom 3(#2), skewed normal mixture (#3) and bimodal normal mixture (#4). Figure 1 depicts these true densities. Integrated squared errors (ISEs) were calculated by using the trapezoidal rule and their average (mean) values were obtained from 200 simulations. The Gaussian kernel function was used throughout and 21 bandwidths were chosen logarithmically equidistant in proper ranges.

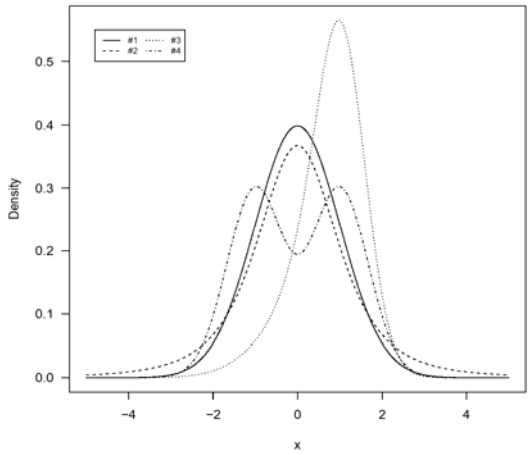

Figure 1. Four true densities: #1: Standard Gaussian, #2: *t*(3), #3: Skewed normal mixture, and #4: Bimodal normal mixture.

Figure 2 contains the mean integrated squared error (MISE), integrated variance (IVAR) and integrated squared bias (IBIAS) for densities #1. The data sharpening method based on the proposed algorithm shows considerable improvements over ordinary kernel estimates and data sharpening using the formula in (2). But, our approach shows poor performance when the bandwidth is small. This is mainly from the IVAR although it shows improvement in terms of IBIAS, which is negligible relative to IVAR. Optimal bandwidths for two versions of the data sharpening method are generally larger than that for an ordinary estimate.

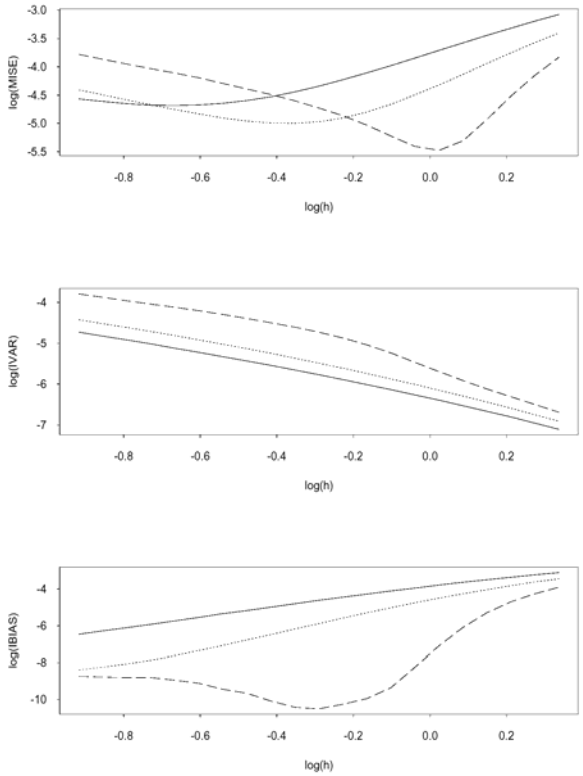

Figure 2. Plots of the MISE and its two components IVAR, IBIAS for density #1: from the top, *y*-axes correspond to logarithms of MISE, IVAR and IBIAS and *x*-axis denotes logarithm of bandwidth. Solid, dashed and dotted lines correspond to ordinary estimates, estimates using the proposed algorithm and estimates based on explicit formula, respectively.

Figure 3 depicts pointwise mean squared errors (MSEs) as well as variances and biases for density #1 when the optimal bandwidths in terms of minimizing each MISE are used. It clearly demonstrates that the proposed algorithm works better than the others in terms of both variances and biases. Figures 4 and 5 illustrate the performance of estimators for density #2. We get a very similar impression to Figures 2 and 3. This seems to be the typical pattern for simple unimodal densities, which include density #3. For density #4, our approach shows improvements over the other two estimates but not quite reaching a satisfactory level relative to the other cases.

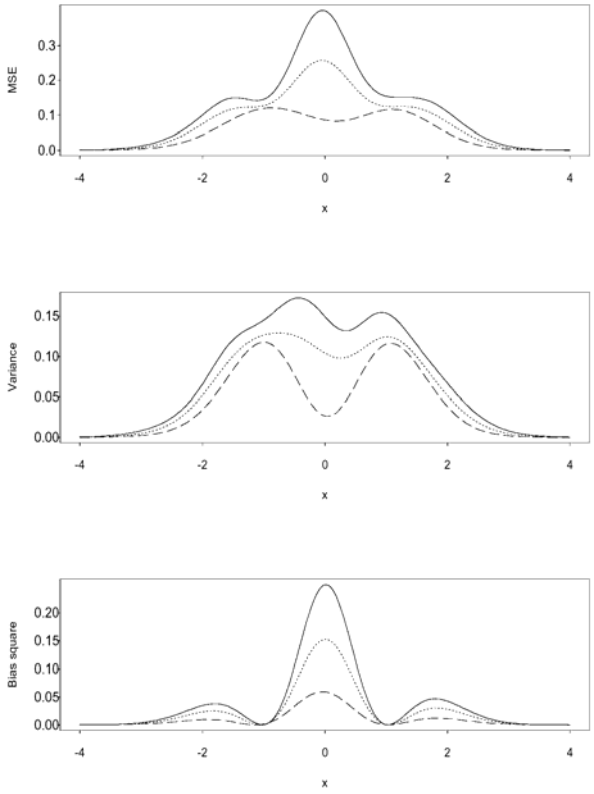

Figure 3. Pointwise mean squared errors, variances and biases plots for density #1: values in *y*-axes are multiplied by 10^2. Solid, dashed and dotted lines correspond to ordinary estimate, estimates using the proposed algorithm and estimates based on explicit formula, respectively. Optimal bandwidths which minimize MISE for each method are used.

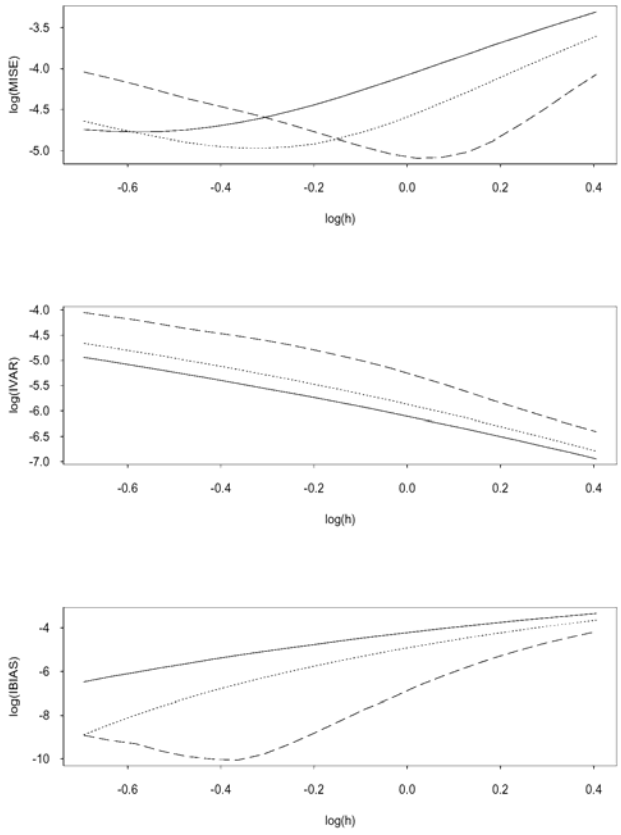

Figure 4. Plots of the MISE and its two components IVAR, IBIAS for density #2: Axes and line types are as for Figure 2.

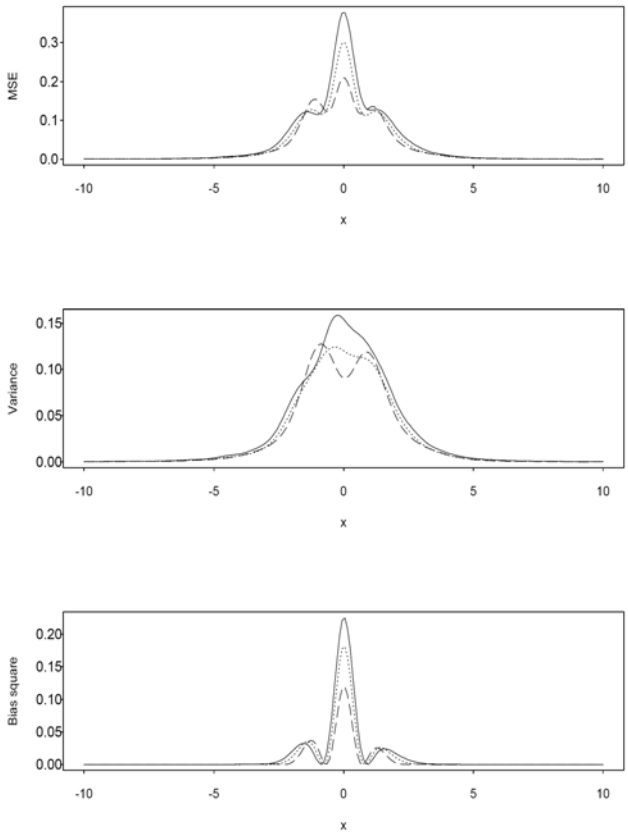

Figure 5. Pointwise mean squared errors, variances and biases plots for density #2: Axes and line types are as for Figure 3.

5. Concluding Remarks

Note that the performance of data sharpening methods highly depends on smoothing amount. Bandwidth selection for the sharpened data set is important and should be larger than that of the original data. But, it is not an easy problem to resolve because it also relates to the sharpening procedure. And hence, few works have been published on this topic. One may use cross-validation for bandwidth selection such as in Hall and Kang [6]. The proposed algorithm converges relatively fast and hence it is computationally efficient compared to the cross-validation approach in Hall and Kang [6].

Future works are needed to develop the theoretical properties of the proposed approach, which is not considered in this paper.

References

[1] E. Choi and P. Hall, Data sharpening as a prelude to density estimation, Biometrika 86 (1999), 941-947.

[2] E. Choi, P. Hall and V. Rousson, Data sharpening methods for bias reduction in nonparametric regression, Ann. Statist. 28 (2000), 1339-1355.

[3] A. P. Dempster, N. M. Laird and D. B. Rubin, Maximum likelihood from incomplete data via the EM algorithm (with discussion), Journal of the Royal Statistical Society Ser. B 39 (1977), 1-38.

[4] M. DiMarzio and C. C. Taylor, Boosting kernel density estimates: a bias reduction technique? Biometrika 91 (2004), 226-233.

[5] P. Hall and M. C. Minnotte, High-order data sharpening for density estimation, Journal of the Royal Statistical Society Ser. B 64 (2002), 141-157.

[6] P. Hall and K. H. Kang, Using cross-validation to alter the data as well as the bandwidth, Unpublished Manuscript, 2002, Available from the second author.

[7] M. Hirukawa and M. Sakudo, Nonnegative bias reduction methods for density estimation using asymmetric kernels, Comput. Statist. Data Anal. 75 (2014), 112-123.

[8] N. L. Hjort and M. C. Jones, Locally parametric nonparametric density estimation, Ann. Statist. 24 (1996), 1619-1647.

[9] M. C. Jones and P. J. Foster, Generalized jackknifing and higher order kernels, J. Nonparametr. Statist. 3 (1993), 81-94.

[10] M. C. Jones and P. J. Foster, A simple nonnegative boundary correction method for kernel density estimators, Statist. Sinica 6 (1996), 1005-1013.

[11] M. C. Jones, O. Linton and J. P. Nielsen, A simple bias reduction method for density estimation, Biometrika 82 (1995), 327-338.

[12] M. C. Jones and D. F. Signorini, A comparison of higher-order bias kernel density estimators, J. Amer. Statist. Assoc. 92 (1997), 1063-1073.

[13] C. R. Loader, Local likelihood density estimation, Ann. Statist. 24 (1996), 1602-1618.

[14] K. Mynbaev and C. Martins-Filho, Bias reduction in kernel density estimation via Lipschitz condition, J. Nonparametr. Stat. 22 (2010), 219-235.

[15] B. U. Park, S. O. Jeong, M. C. Jones and K. H. Kang, Adaptive variable location kernel density estimators with good performance at boundaries, J. Nonparametr. Statist. 15 (2003), 61-75.

[16] M. Samiuddin and G. M. El-Sayyad, On nonparametric kernel density estimates, Biometrika 77 (1990), 865-874.

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