Shape-Adaptive Kernel Density Estimation

L.E.N. Baakman

September 4, 2017

Abstract

Kernel density estimation has gained popularity in the past few years. Generally the methods use symmetric kernels, even though the data of which the density is estimated are not necessarily spread equally in all dimensions. To account for this asymmetric distribution of data we propose the use of shape adaptive kernels: kernels whose shape changes to fit the spread of the data in the local neighborhood of the point whose density is estimated. We compare the performance of the shape adaptive kernels on simulated datasets with known density fields.

Results

Conclusion

	Estimator	
	MBE	saMBE
one	4.118×10^{-10}	2.983×10^{-9}
two	5.279×10^{-8}	1.001×10^{-7}
three	4.375×10^{-6}	5.484×10^{-6}
four	4.779×10^{-7}	1.231×10^{-4}
five	5.383×10^{-8}	9.425×10^{-8}
six	4.189×10^{-6}	5.454×10^{-6}
seven	7.323×10^{-7}	4.110×10^{-4}
eight	6.569×10^{-7}	3.306×10^{-4}

Table 1: *The mean squared error of the known densities and the densities estimated by the Modified Breiman Estimator (MBE) and the shape-adaptive MBE (saMBE), respectively, for the datasets in* **??**.

1 Introduction

2 Method

3 Experiment

4 Results

This section presents the results of the experiments described in Section 3. We compare the performance of the two estimators on each dataset with the mean square error, presented in Table 1, and visually with plots. All plots associated with a single dataset have the same domain and range, to allow for easy comparison of the results within a dataset. The horizontal axis is used to represent the known densities, its range is such that each known density can be shown. The estimated densities are shown on the vertical axis, the length of these axes is such that they are long enough to represent every estimated density for that dataset, independent of the used estimator. The black line in each plot illustrates the line all points would lie on if a perfect estimator was used, i.e. the line $x = x$. The colors of the points in these plot correspond to the colors of the elements of the datasets in **??** and **??**.

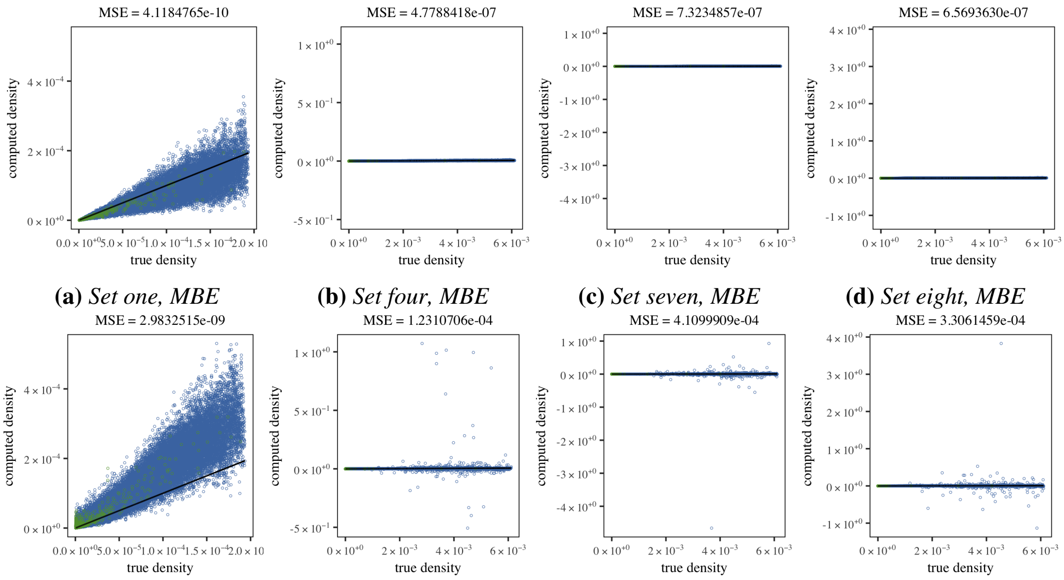
Section 4.1 presents the results of the datasets that contain a single Gaussian, in Section 4.2 the results of the datasets that consist of noise and multiple Gaussian distributions are presented.

4.1 Datasets with a Single Gaussian

This section compares the performance of the Modified Breiman Estimator and a shape-adaptive variant on dataset that contain one Gaussian, i.e. dataset one, four, seven, and eight.

Only on the dataset with a spherical Gaussian is the performance of the two estimators comparable, on all dataset with a single elliptical Gaussian the non-shape adaptive estimator performs significantly better than its shape-adaptive cousin.

Figure 1 presents the results of using the Modified Breiman Estimator and its shape adaptive variant to



(a) *Set one, MBE* **(b)** *Set four, MBE* **(c)** *Set seven, MBE* **(d)** *Set eight, MBE*

(e) *Set one, saMBE* **(f)** *Set four, saMBE* **(g)** *Set seven, saMBE* **(h)** *Set eight, saMBE*

Figure 1: *Comparative plots for dataset 1, 4, 7, and 8.*

estimate the densities of the datasets in **??** that contain a single Gaussian distribution.

Figure 1a confirms our findings from **??**, namely that the Modified Breiman Estimator gives a good approximation of the densities of dataset one. The densities estimated with the MBE both over, and undershoot the true density. Figure 1a, on the other hand, shows that shape adaptive MBE nearly always overshoots the true density.

Comparing the performance of the two estimators on dataset four with Figures 1b and 1f we find that the Modified Breiman Estimator outperforms the shape-adaptive variant. The second estimator has some extreme outliers, the most extreme of which are 1.072, and -0.5068.

The results of data set seven, shown in Figures 1b and 1f respectively, are comparable to those of four: the original estimator approximates the density pretty well, the shape-adaptive variant has some extreme outliers, the densities estimated by saMBE fall in the range, $[-4.661, 0.9283]$, whereas the true densities all lie within $[5.000 \times 10^{-7}, 6.108 \times 10^{-3}]$.

Figures 1d and 1h compare the performance of respectively MBE with saMBE on data set eight. We once again observe that the non-shape adaptive estimator approximates the known densities pretty well. Whereas the shape-adaptive estimator returns extreme results with densities that are estimated to be as high as 3.827 and as low as -1.134.

In general we have found that the Modified Breiman estimator works pretty well for data sets that contain a single Gaussian, especially if the Gaussian is spherical. Since the mean square error for dataset eight is lower than the MSE of dataset seven the ellipticalness of the distribution does not seem to influence the performance of this estimator. The shape adaptive MBE results in some extremely high and low estimated densities if used to estimate the densities of non-spherical Gaussian. saMBE overestimated some of the densities of the spherical Gaussian compared to the Modified Breiman Estimator. The range of the values estimated by the shape-adaptive estimator does not seem to be influenced by how electricalness of the Gaussian distribution.

4.2 Datasets with Multiple Gaussians

Mention MSE results

MSE for the different components

Point to Figure 2

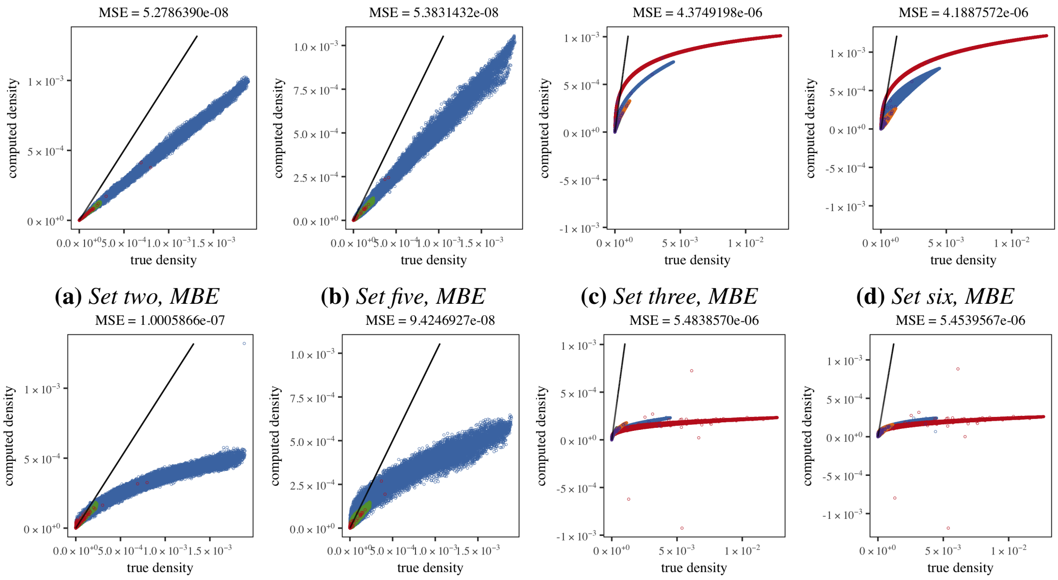
.

Report on Figure 2a and Figure 2b.

Report on Figure 2c and Figure 2d.

General observation of multi sphere datasets.

General observation of the results.



(a) *Set two, MBE* **(b)** *Set five, MBE* **(c)** *Set three, MBE* **(d)** *Set six, MBE*

(e) *Set two, saMBE* **(f)** *Set five, saMBE* **(g)** *Set three, saMBE* **(h)** *Set six, saMBE*

Figure 2: *Comparative plots for dataset 2, 3, 5, and 6.*

5 Discussion

This section discusses the results presented in Section 4 using the structure used in **??**.

5.1 Datasets with a Single Gaussian

The difference in results between the Modified Breiman Estimator and its shape adaptive variant on dataset one, four, seven, eight raise several questions. This section attempts to answer them.

In Figure 1 we observed that saMBE overestimates the densities of dataset one. This could indicate that the kernels are too small, resulting in a too high contribution to the density estimate. Since the Modified Breiman estimator uses the same general and local bandwidth as the shape adaptive version the likely culprit is the shape of the kernel.

Another strange result observed in Figure 1 is that a large number of estimated densities were not in the expected range $[0,1]$ when saMBE was used. Strangely this effect does not occur when the shape adaptive matrix is not used in a dataset that contains spherical data, i.e. in dataset one. Looking back to **??** we find that since $\forall \mathbf{x}\ K(\mathbf{x}) \in [0,1]$, $\det(\mathbf{H}_i)$ must be smaller than zero to cause a negative density estimate. For the same reason the density estimates that are greater than zero must be the result of $\det(\mathbf{H}_i) < 1$ for some $\mathbf{H}_i$.

The issues above probably explain why the shape adaptive Modified Breiman Estimator does not outperform the non-shape adaptive variant on these data sets.

5.2 Datasets with Multiple Gaussians

What does this section do?

Discuss Figure 2a and Figure 2b.

Discuss Figure 2c and Figure 2d.

General discussion of multi sphere datasets.

General Discussion

6 Conclusion