Shape-Adaptive Kernel Density Estimation

L.E.N. Baakman

September 1, 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

1 Introduction

2 Method

3 Experiment

4 Results

This section presents the results of the experiments described in Section 3. In Section 4.1 the mean squared errors of the estimated densities are presented. The comparative plots are presented in Section 4.2.

4.1 Mean Squared Error

Table 1 presents the mean squared error of the different dataset for the two estimators. Comparing the two columns of mean squared errors we find that the Modified Breiman Estimator outperforms the shape-adaptive variant on every dataset.

The differences in MSE between the two datasets on dataset 1 is negligible. On dataset 2, and 10, the performances of the estimators differ more, but the results are comparable. The biggest difference in performance between estimators is found when comparing the results of dataset 4, 7, and 8.

	Estimator	
	MBE	saMBE
1	4.118×10^{-10}	2.983×10^{-9}
2	5.279×10^{-8}	1.001×10^{-7}
3	4.375×10^{-6}	5.484×10^{-6}
4	4.779×10^{-7}	1.231×10^{-4}
5	5.383×10^{-8}	9.425×10^{-8}
6	4.189×10^{-6}	5.454×10^{-6}
7	7.323×10^{-7}	4.110×10^{-4}
8	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* **??**.

4.2 Comparative Plots

This section presents a visual representation of the results of the estimators that allows us to compare the performance of the two estimators on a single dataset. All plots associated with a single dataset have the same axes. The horizontal axis is used to represent the known densities, this axis has the same range as the known densities. 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 **??**.

Figure 1 presents the results of using the Modified Breiman Estimator and its shape adaptive variant to estimate the densities of the datasets in **??** that contain a single Gaussian distribution.

Figure 1a confirms our findings from Section 4.1, namely that the Modified Breiman Estimator gives a good approximation of the densities. The estimated densities both over, and undershoot the true density. Figure 1a, on the other hand, shows that shape adaptive MBE consistently overshoots the true density.

Comparing the performance of the two estimators on dataset four with Figures 1c to 1d 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 $1.071\,674\,654\,143\,224$, and $-0.506\,829\,052\,779\,317$. Plotting only the results where the density as estimated by saMBE is in the range $(-0.07, 0.07)$ results in Figure 2. The mean square errors of this subset are $4.775\,603\,4 \times 10^{-7}$ and $4.258\,066\,7 \times 10^{-6}$ for MBE and saMBE, respectively. Observing Figure 2b we find that using saMBE results in a lot of estimated densities that are smaller than zero and in underestimated densities, especially when compared to Figure 2a.

The results of data set seven, shown in Figures 1c to 1d are comparable to those of four: the original estimator approximates the density pretty well, the shape-adaptive variant has some extreme outliers.

Extra plaatje met de outliers eruit gehaald, kijken wat het dan geeft.

Figures 1g to 1h compares the performance of MBE with saMBE on data set eight. We once again observe that the shape adaptive variant results in extreme outliers, and that the non-shape adaptive estimator approximates the known densities pretty well.

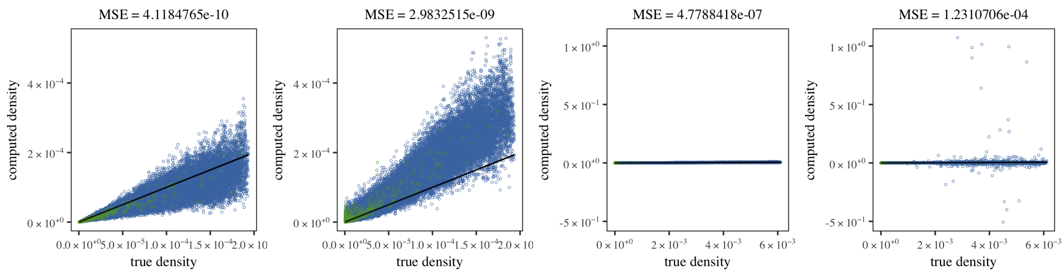
Extra plaatje met de outliers eruit gehaald, kijken wat het dan geeft.

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 of the MBE is lower for dataset 8, than for dataset 7 how elongated the sphere is does not seem to influence the performance of this estimator. The shape adaptive MBE results in some extremely high and low values if it is used to estimate the densities of elongated Gaussians, when applied to a spherical Gaussians it overestimates the densities. The value of these extreme 'densities' does not seem to be influenced by how elongated the Gaussian distribution is.

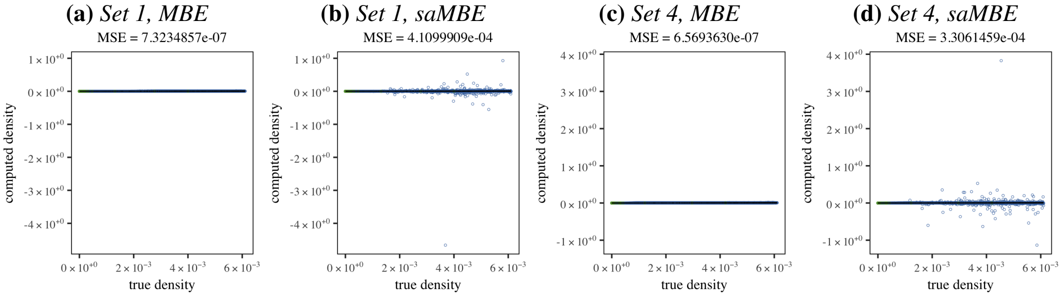
What do we observe for dataset ferdosi 2, baakman 2, ferdosi 3, baakman 3

5 Discussion

6 Conclusion

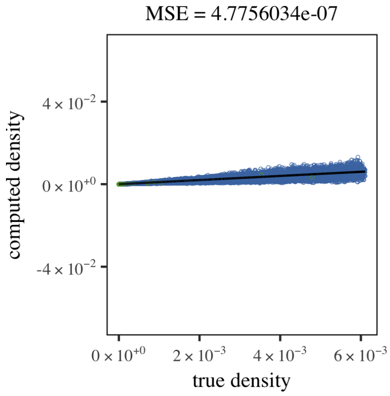


(a) *Set 1, MBE* **(b)** *Set 1, saMBE* **(c)** *Set 4, MBE* **(d)** *Set 4, saMBE*

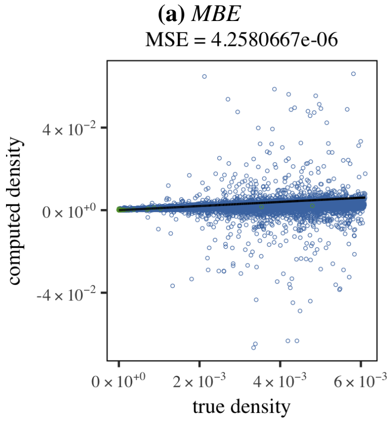


(e) *Set 7, MBE* **(f)** *Set 7, saMBE* **(g)** *Set 8, MBE* **(h)** *Set 8, saMBE*

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

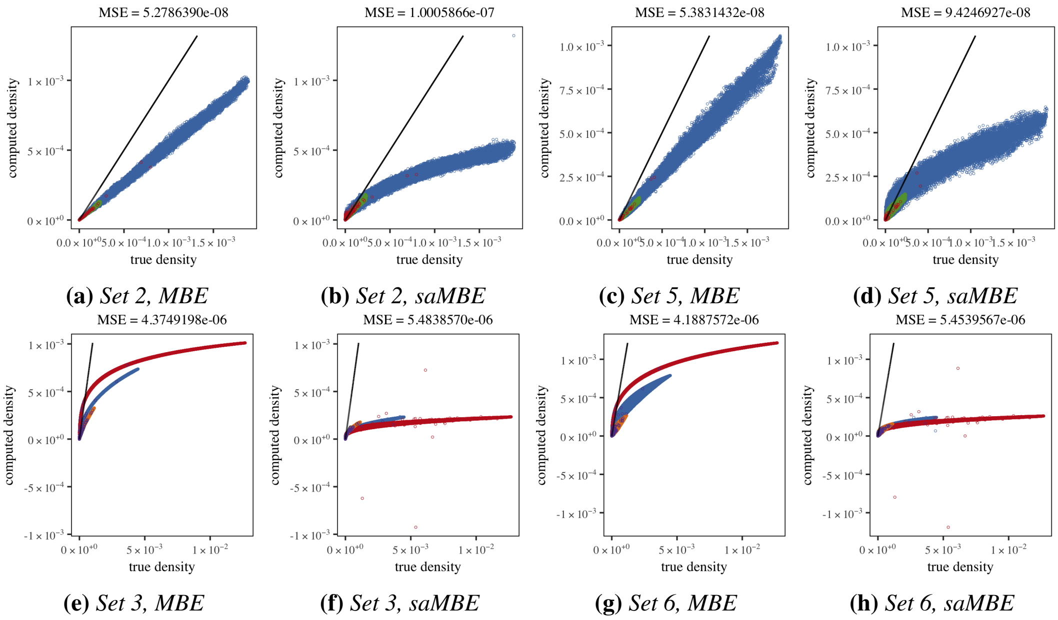


(a) *MBE*



(b) *saMBE*

Figure 2: *Comparative plots between the true densities of dataset 4 as estimated by a MBE and b saMBE, with only the points whose density is estimated by saMBE to lie in* $(-0.07, 0.07)$.



(a) *Set 2, MBE* **(b)** *Set 2, saMBE* **(c)** *Set 5, MBE* **(d)** *Set 5, saMBE*

(e) *Set 3, MBE* **(f)** *Set 3, saMBE* **(g)** *Set 6, MBE* **(h)** *Set 6, saMBE*

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