

Shape-Adaptive Kernel Density Estimation

L.E.N. Baakman

September 15, 2017

Abstract

Kernel density estimation is a popular method to approximate probability densities in numerous fields. Generally these 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. We compare the performance of the shape adaptive kernels with that of an estimator that uses a symmetric kernel on simulated datasets with known density fields. No significant differences in performance between the symmetric and the shape-adaptive estimator were found. Although the former outperformed the latter on points near the boundary of the datasets. We also found some differences in performance dependent on the distance to the mean of Gaussian distributions with low values on the diagonal of the covariance matrix. In conclusion shape-adaptive kernels are a promising idea that warrants further research.

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 squared error 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 **????**.

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.

	Estimator	
	MBE	saMBE
S_1	8.306×10^{-9}	8.909×10^{-9}
S_2	1.490×10^{-8}	1.540×10^{-8}
S_3	2.937×10^{-8}	2.963×10^{-8}
S_4	5.572×10^{-8}	5.585×10^{-8}

Table 1: Performance of the Modified Breiman Estimator with fixed-shaped and shape-adaptive kernels on the datasets with a single Gaussian.

4.1 Datasets with a Single Gaussian

This section compares the performance of the Modified Breiman Estimator with symmetric and shape-adaptive kernels on datasets that contain one Gaussian. Comparing the mean squared errors of the MBE with those of saMBE in Table 1 we find that the two estimators perform comparably, but that the fixed-shape estimator always gives a slightly lower mean squared error.

This is confirmed by the visualization of the results in Figure 1 where hardly any difference is visible between Figures 1(a) to 1(d), and Figures 1(e) to 1(h), respectively. Comparing the plots associated with dataset S_1 we find that the shape-adaptive estimators tends to overestimate densities more than the symmetric observation. Based on Figures 1(c) and 1(g) the same holds for dataset S_3 . Comparing the performance within datasets between the two

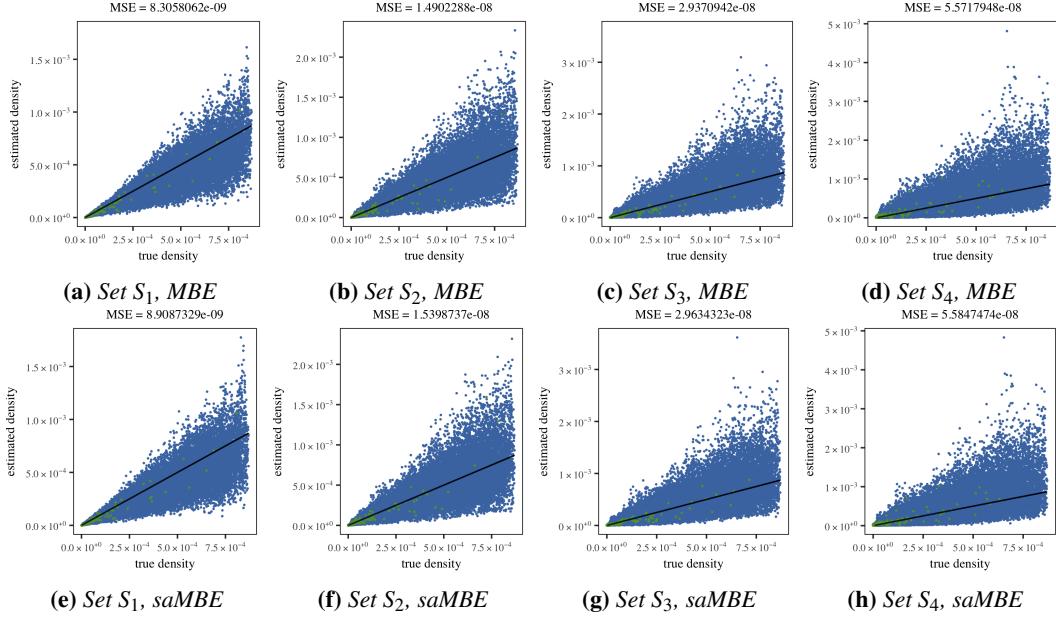


Figure 1: Plot of the density as estimated by (a)-(d) MBE and (e)-(h) saMBE as a function of the known density of the datasets with a single Gaussian.

	● Gaussian		● Noise			
	M	SD	M	SD	M	
S_1	1.48	0.521	1.29	0.136	1.86	0.750
S_2	1.57	0.553	1.41	0.289	1.91	0.766
S_3	1.64	0.586	1.51	0.403	1.91	0.772
S_4	1.80	0.698	1.74	0.638	1.93	0.790

Table 2: The mean (M) and the standard deviation (SD) of the anisotropy of the kernels used for the datasets with a single Gaussian.

components showed no marked differences in performance between the estimators between components.

Table 2 presents the mean and the standard deviation of the anisotropy of the kernels used for the different datasets. Comparing the means we find a positive correlation between the anisotropy of the Gaussian component of the dataset and mean anisotropy of the kernels. The same positive correlation can be observed for the standard deviation. Reviewing these statistics of the components of the datasets reveals that the increase in average anisotropy is primarily caused by an increase in anisotropy of kernels of points sampled from the Gaussian component. The mean anisotropy of the noise component stays relatively constant. Furthermore as the Gaussian component is more anisotropic the variation in anisotropy of the kernels increases.

Set	Estimator	
	MBE	saMBE
M_1	5.058×10^{-8}	5.050×10^{-8}
M_2	5.147×10^{-8}	5.168×10^{-8}
M_3	4.375×10^{-6}	4.463×10^{-6}
M_4	4.189×10^{-6}	4.284×10^{-6}

Table 3: Performance of the symmetric and the shape-adaptive Modified Breiman Estimator on the datasets containing multiple Gaussian distributions.

To summarize; in spite of differences in anisotropy of the used kernels we have observed very few differences between two estimators. Using shape-adaptive kernels did not yield the expected gain in performance. We did find the expected influence of the anisotropy of the Gaussian components on that of the kernels.

4.2 Datasets with Multiple Gaussians

In this section we present the results of the two estimators on dataset M_1 , M_2 , M_3 , and M_4 . Based on the small differences between the mean squared errors of the estimators in Table 3 the estimators perform comparably on these datasets. Comparing the MSE between components within the data sets between estimators yields no differences. However within

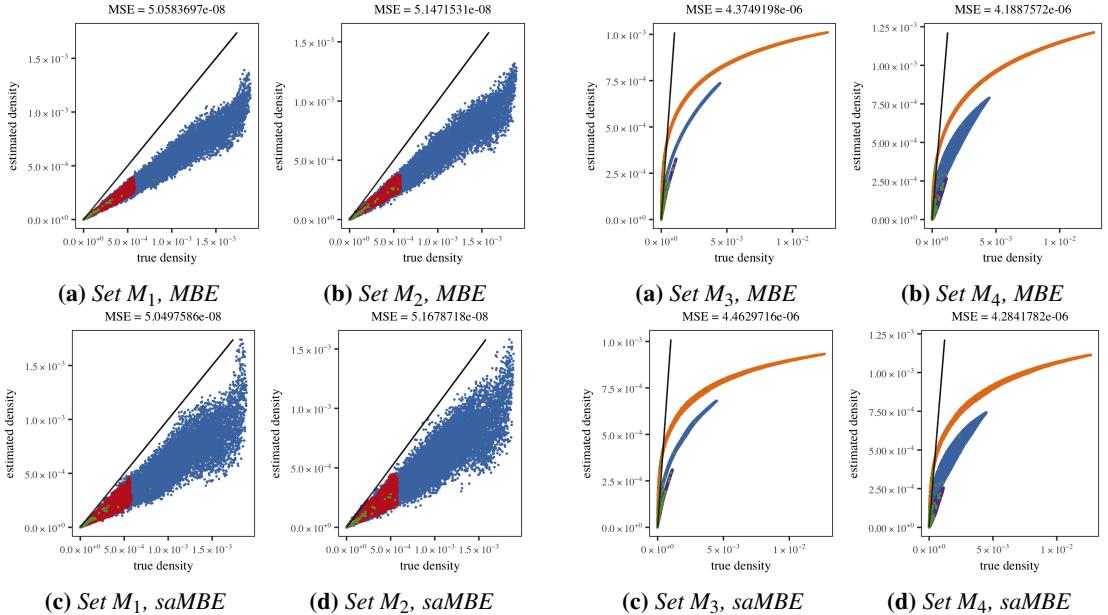


Figure 2: Plots of the true versus estimated density of datasets M_1 and M_2 for the shape-adaptive and the symmetric Modified Breiman Estimator.

data sets the difference in mean squared errors are quite large. Within dataset M_1 and M_2 both estimators perform significantly better on the more sparse component ‘Trivariate Gaussian 2’. Both estimators show the same positive correlation between the density of the Gaussian components and the MSE between the components of dataset M_3 and M_4 . Additionally contrary to our expectation both estimators performed slightly better on dataset M_4 than on the set it was derived from.

Figure 2 shows the estimated density as a function of the known density for both estimators and both datasets with two Gaussians. These plots show that both estimators underestimate the density, MBE more than saMBE. Furthermore the shape-adaptive estimator shows more variation in the densities it estimates than the symmetric estimator. Comparing Figures 2(a) and 2(b) with Figures 2(c) and 2(d) suggest that saMBE hardly underestimates the densities of the most anisotropic component, i.e. ‘Trivariate Gaussian 2’ in bot dataset M_1 and M_2 . However the difference in mean squared error of this component between MBE and saMBE is pretty small, a large difference in standard deviation of the squared error suggests that the seemingly better performance of the shape-adaptive estimator in Figures 2(c) and 2(d) is mostly due to the higher spread of densities estimated by the shape-adaptive estimator. Furthermore the difference in mean squared error between com-

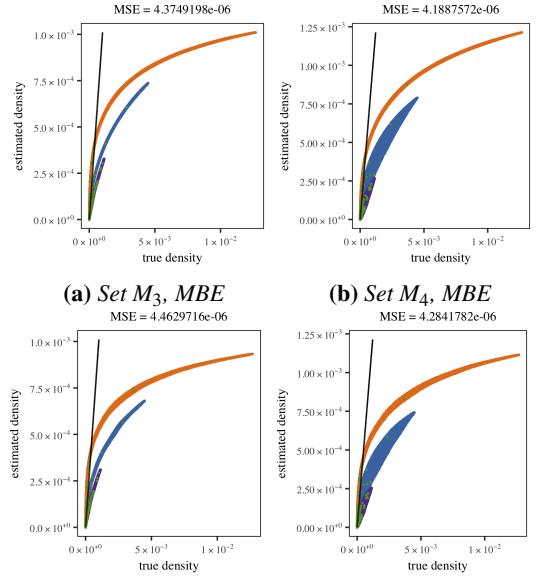


Figure 3: The estimated density plotted as a function of the true density for datasets M_3 and M_4 for MBE and saMBE.

ponents within estimators and dataset M_1 and M_2 shows that both estimators perform worse on the denser Gaussian component, i.e. ‘Trivariate Gaussian 1’. The plots in Figure 3 confirm the large difference in performance of both estimators between datasets with two and four Gaussian components. Moreover they show that both MBE and saMBE underestimate densities, especially on the points with a high density. Figure 3 also shows that the spread of the densities estimated by saMBE is higher than of those estimated by MBE.

Table 4 presents the mean and standard deviation of the anisotropy of the kernels used for the points from dataset M_1 through M_4 and from their components. Although the anisotropy of the kernels used for the datasets with anisotropic Gaussian components is higher on average and more varied than the anisotropy of the kernels used for dataset M_1 and M_3 , the differences are small. As in Table 2 the kernels associated with the points sampled from the component ‘Uniform random background’ have the highest anisotropy, and vary the most in how anisotropic they are. Contrasting the mean anisotropy of the kernels used for points drawn from the different components we find a positive correlation between the mean anisotropy of the kernel and the anisotropy of the Gaussian components. In the kernels used for dataset M_2 we observe the same positive correlation between the variation of

	• Gaussian 1		• Gaussian 2		• Gaussian 3		• Gaussian 4		• Noise	
	M	SD	M	SD	M	SD	M	SD	M	SD
M_1	1.50	0.531	1.32	0.175	1.30	0.143			1.89	0.759
M_2	1.61	0.570	1.41	0.278	1.49	0.345			1.95	0.783
M_3	1.46	0.551	1.29	0.189	1.27	0.130	1.29	0.210	1.28	0.165
M_4	1.53	0.571	1.31	0.219	1.49	0.339	1.29	0.210	1.40	0.285
									1.82	0.811
									1.85	0.820

Table 4: The mean (M) and standard deviation (SD) of the anisotropy of the kernels used for points from the datasets with multiple Gaussians, split per component and for the full dataset.

the anisotropy of the kernels and the anisotropy of the associated kernel as we observed in dataset S_2 , S_3 , and S_4 . Comparing the standard deviation of the anisotropy of the kernels used for the points sampled from the components of dataset M_4 does not reveal this relation: the densest component, Trivariate Gaussian 3, does not have the highest variation in kernel anisotropy. One thing that stands out when comparing dataset M_3 and M_4 in Table 4 is the lack of difference in the mean and standard deviation in the anisotropy of the kernels associated with the component ‘Trivariate Gaussian 3’. Reviewing the raw data reveals that the largest difference in anisotropy between any point drawn from that component is 3.109×10^{-15} .

To summarize the results of the datasets with multiple Gaussian components: we have found that the differences in performance are very small between the two estimators, although saMBE underestimates less than MBE, the later performs slightly better. Furthermore both estimators show a positive correlation between the density of the Gaussian component and their performance on that component. Regarding the anisotropy of the kernels we have found only a small increase between dataset with spherical kernels and dataset M_2 and M_4 . Zeroing in on the components we have observed that the anisotropy of kernels associated with points drawn from the noise is strongest. Lastly a positive correlation between the mean anisotropy of the kernels of the points sampled from a component and the anisotropy of the Gaussian those points were drawn from was found.

In conclusion for all datasets we have found very small differences in mean squared error between the two estimators. Furthermore both estimators perform better on points drawn from Gaussian components with a smaller eigensphere. Concerning the anisotropy of the kernels we have observed that in all datasets this statistic is highest for the noise component. Comparing the mean anisotropy of the kernels of the points drawn from the Gaussian components showed a positive correlation with the anisotropy of the component the point was sampled

from.

5 Discussion

This section discusses the results presented in Section 4. We consider the lack of difference in performance between the estimators in Section 5.1. The next section discusses the anisotropy of the kernels used by the shape-adaptive estimators.

5.1 Performance

One of the most striking observations from Section 4 is that the differences in performance between the two estimators are minimal.

Plotting the densities as estimated by saMBE as a function of those estimated by MBE shows no interesting differences between the two estimators for dataset S_1 through S_4 . However for dataset M_1 through M_4 these plots are interesting. As can be seen in Figure 4 using shape-adaptive kernels results in estimated densities that are generally higher than those estimated with a fixed-shape kernel for dataset M_1 and M_2 . A review of the raw data shows that the points whose density is estimated by saMBE to be significantly higher than MBE, and nearer to the true density, can all be found near the mean of ‘Trivariate Gaussian 1’. The kernels in this neighborhood are all slightly anisotropic, and have allowed the shape-adaptive estimator to use more data points, to better approximate the local densities. Thus showing that at least for some points estimating the density with shape-adaptive kernels is advantageous.

Figure 5 shows the opposite effect; the densities estimated by saMBE for dataset M_3 and M_4 are generally lower than those estimated by MBE. Reviewing the raw data shows that the points with the greatest differences between the two estimators lie near the mean of the ‘Trivariate Gaussian 3’ in both dataset M_3 and M_4 . The number of points used in the density estimate by saMBE is consistently lower and than those used by the fixed-shape estimator. Given

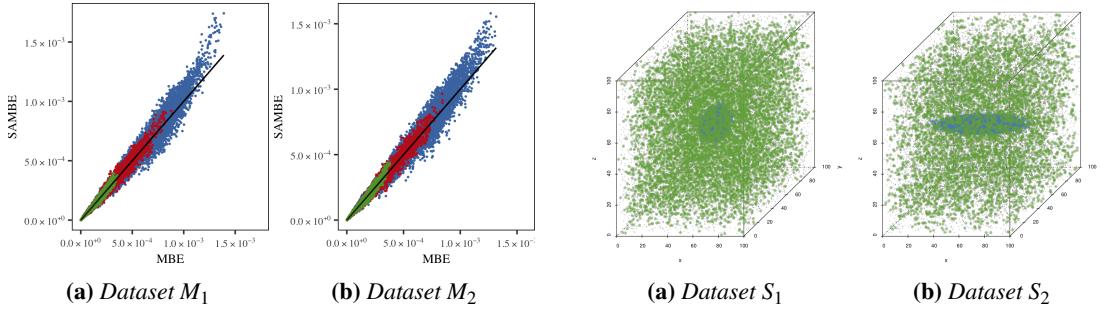


Figure 4: Plots of the density estimated by saMBE as a function of those estimated by MBE for dataset (a) M_1 and (b) M_2 .

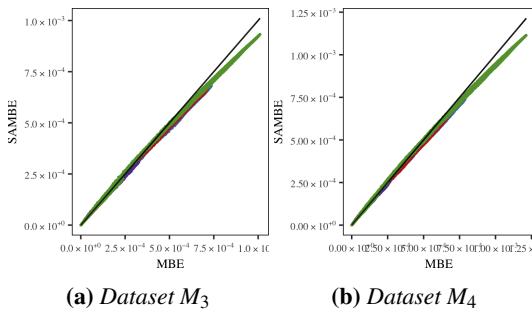


Figure 5: Plots of the density estimated by saMBE as a function of those estimated by MBE for dataset (a) M_3 and (b) M_4 .

the relatively high anisotropy of the kernels in that area we expect that this is due to the kernels reflecting fine local structures, that are not relevant to the more global neighborhood.

The plots in Figure 6 emphasizes the points in dataset S_1 through S_2 where the absolute error of MBE is smaller than that of saMBE. These plots show that the shape-adaptive kernels outperforms the symmetric kernels near the borders of the datasets. We expect that the boundary effect is due to the strong anisotropy of the local neighborhood of the points near the boundaries. After all these neighborhoods all point strongly towards the dataset. Consequently the domain of the shape-adaptive kernels extends less outside of the boundaries dataset than the domains of the symmetric kernels. This is results in less underestimating of densities near the boundary of the dataset if shape-adaptive kernels are used. Furthermore the strength of the boundary effect seems to increase as the Gaussian component of the dataset is more anisotropic. However the seemingly better performance of saMBE is due to an increase in the number of points where the density estimated by saMBE equals that estimated by MBE. In dataset S_1 the density estimates of the two esti-

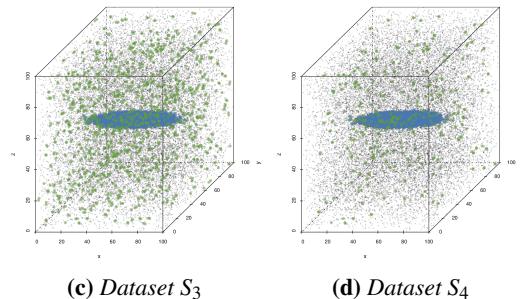


Figure 6: Low opacity scatter plot of dataset (a) S_1 , (b) S_2 , (c) S_3 , and (d) S_4 , with an overlay of larger colored points where the absolute error of saMBE is larger than the absolute error of MBE.

mators differ on all points, however this percentage increases as the anisotropy of the Gaussian component increases, to 35.4 % in dataset S_4 . As the Gaussian component becomes more anisotropic there are more points whose local neighborhood consists only of uniform noise. On average the covariance matrix of neighborhoods that contain primarily points sampled from the noise component should be scalar. Consequently the shape-adaptive kernels of a large number of points sampled from the noise component are symmetric, resulting in points where the estimators give the same approximated density.

The points where using symmetric kernels results in a smaller error in datasets M_1 through M_4 are emphasized in Figure 7. We contribute the boundary effect in these datasets to the same cause as the boundary effects in the datasets with a single Gaussian component. Interestingly the points dataset M_3 and M_4 where the absolute error of MBE is lower approximate a sphere, contrary to the cube they define for dataset M_1 and M_2 . It is our expectation that this is caused by smaller distance between the means of the Gaussian components and the boundary of the dataset. To test this we define dataset M'_3 , this dataset replaces the uniform random background of M_3 with $\mathcal{U}([-20, -20, -20], [120, 120, 120])$. To ensure that the density of that components is equal

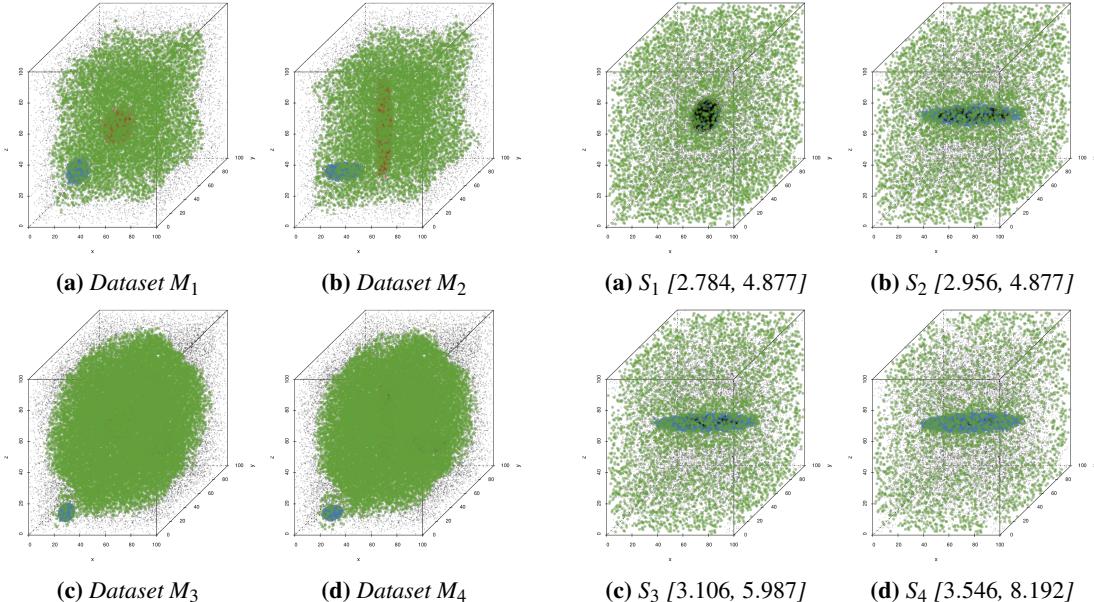


Figure 7: Low opacity scatter plot of dataset (a) M_1 , (b) M_2 , (c) M_3 , and (d) M_4 with an overlay of high opacity larger points where the absolute error of MBE is smaller than that of saMBE.

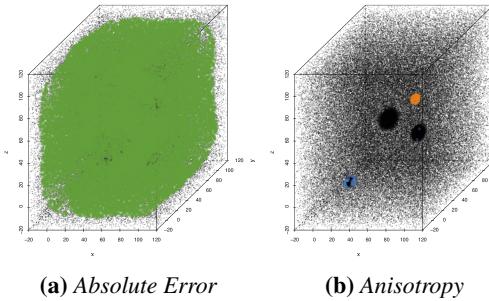


Figure 8: Low opacity scatter plot of dataset M'_3 with (a) points where the absolute error of MBE is lower than that of saMBE and (b) points sampled from the Gaussian components with kernels whose anisotropy falls in the 95th percentile of the complete dataset emphasized.

to that of the noise of M_3 we adjust the number of points of the component. Compared to M_3 , the overall mean squared error of both estimators has decreased slightly for M'_3 , however the mean squared error of ‘Trivariate Gaussian 1’ and 3 shows a small increase. Figure 8(a) confirms that the spherical shape in Figures 7(c) and 7(d) is caused by the Gaussian near the boundary of the dataset.

To conclude, by investigating the differences in performance between the two estimators on various datasets we have found that shape-adaptive ker-

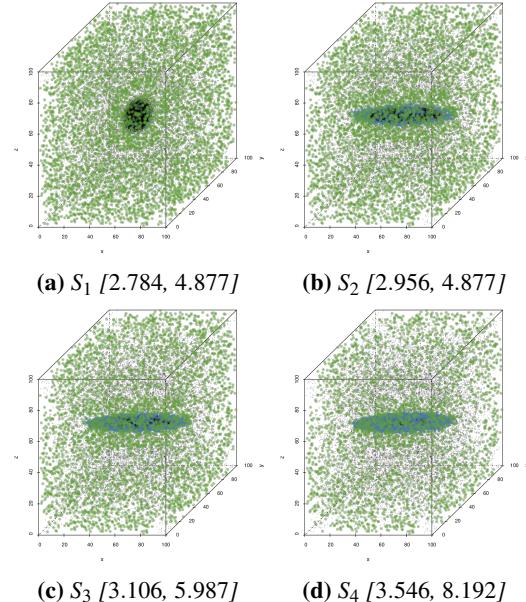


Figure 9: Scatter plot of the data sets (a) S_1 , (b) S_2 , (c) S_3 , and (d) S_4 . The points with kernels whose anisotropy lies in the 90th percentile are shown larger and in the color of the component they were drawn from. The range of the anisotropy of the kernels of the emphasized points is shown below each plot.

nels definitely improve performance in some cases, e.g. near the boundary of the datasets and near the mean of some Gaussian components. Unfortunately in other cases the anisotropic kernels are detrimental, the difference in mean squared error between the two estimators show that on these datasets using symmetric kernels is slightly advantageous.

5.2 Kernel Anisotropy

In Section 4 some differences in anisotropy of the kernels were observed, however these differences were relatively small. This section investigates the anisotropy of the kernels.

Figure 9 shows the datasets with a single Gaussian with points whose anisotropy lies in the 90th percentile emphasized. In this figure hardly any difference is visible between Figures 9(a) and 9(b). In Figures 9(a) and 9(b) 0.518 % and 11.7 %, respectively of the emphasized points are sampled from the Gaussian component of the datasets. Illustrating that the kernels in dataset S_2 are influenced by the anisotropy of the Gaussian component of that dataset. Furthermore a shell of points with kernels with relative high anisotropy, sampled from the noise, seems to surround the Gaussian component,

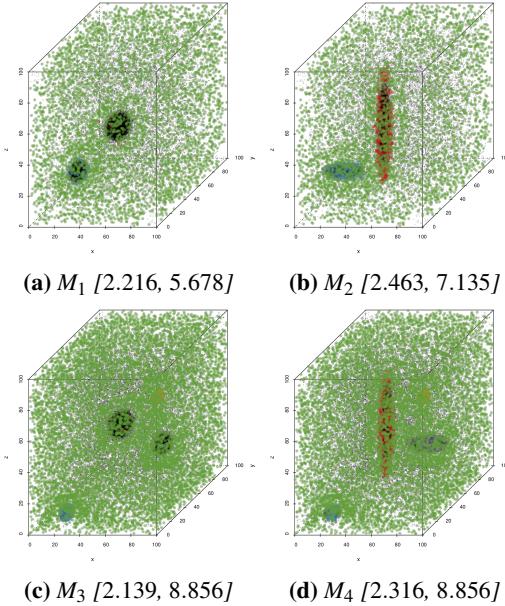


Figure 10: Scatter plot of datasets (a) M_1 , (b) M_2 , (c) M_3 , and (d) M_4 . The points that have an anisotropy in the 90th percentile are shown larger and in the color of the component they were drawn from. The range of the anisotropy of the kernels of the emphasized points is shown below each plot.

it is quite likely that the shape of the kernel of these points is influenced by the Gaussian component. We expect that nearer to the mean of the Gaussian component fewer kernels are influenced by its anisotropy as the physical density of points is quite high in that area. Consequently the volume of the local neighborhood is quite small, and is therefore not representative of the shape of the Gaussian component. In dataset S_3 and S_4 the number of points with a kernel with a high anisotropy sampled from the Gaussian components is higher than in dataset S_1 and S_2 , to be exact 21.9 % and 42.1 %, respectively. Thus as the anisotropy of the Gaussian component increases, the anisotropy of the kernels of the points near that component increase.

Figure 10 emphasizes the points with the most anisotropic kernels in the dataset M_1 through M_4 . In the plot associated with dataset M_1 we observe the same shells of high anisotropy kernels of points sampled from the noise around the Gaussian components as in dataset S_1 . Another similarity between these two datasets is that very few points sampled from the Gaussian component have a kernel with high anisotropy; 1.45 % and 0.117 % of the first and second Gaussian component, respec-

tively. We contribute the difference in the number of highly anisotropic kernels associated with data points sampled from the two Gaussian components to the difference in density between the two Gaussian components. Comparing Figures 10(a) and 10(b) we find that the increase in anisotropy of the kernel has causes 4.33 % and 8.66 % of the kernels with high anisotropy to be associated with a point sampled from ‘Trivariate Gaussian 1’ and 2, respectively. Interestingly in dataset M_3 , two of the four spherical Gaussians, i.e. ‘Trivariate Gaussian’ are associated with respectively 1.40 % and 2.22 % of the highly anisotropic kernels. Comparing the mean kernel anisotropy in Table 4 we find that compared to the other Gaussian components in M_3 , these two component have a relatively low mean. However their standard deviations are higher, suggesting that in these denser components some kernels are extremely anisotropic, whereas others are near to isotropic. We contribute this difference within the points sampled from a Gaussian component to difference in physical density of the data points nearer and farther away from the mean as we did for dataset S_1 and S_2 . Component one and four of dataset M_3 differ from the other Gaussian components in two aspects: firstly they have a relatively dense and secondly they are placed near the boundaries of the dataset. Figure 8(b) shows that distance to the boundaries do not explain the difference in anisotropy of the kernel as in that figure the first and third component of M'_3 have more anisotropic kernels than the other components. The first explanation fits with the observations from Section 4 that components with a higher density have a higher anisotropy. In dataset M_4 we observe the same effect as in M_3 but stronger; from the points with the most anisotropic kernels more are sampled from the two densest Gaussian component than from the others.

Figures 9 and 10 show that the overwhelming majority of the points with a relative highly anisotropic kernel are sampled from the ‘Uniform random background’. We contribute this to the covariance detecting fine, random structures in the noise, that give the impression of anisotropy where there is none.

In Section 4 we observed that both the density and the anisotropy of the Gaussian component influenced the anisotropy of the kernels. To test if the difference was caused by the anisotropy or the density of the dataset we introduce two datasets, A_1 and A_2 , with the same anisotropy but the same density. To create dataset A_1 the covarianceMatrix used in S_1 was replaced by $\text{diag}([4, 3, 1])$, dataset A_2 is the same as set A_1 but uses $3 \cdot \text{diag}([4, 3, 1])$. The mean squared errors of the two estimators on dataset A_1 and A_2 are

Estimator		
	MBE	saMBE
A_1	1.240×10^{-6}	1.297×10^{-6}
A_2	4.713×10^{-8}	4.941×10^{-8}

Table 5: Performance of the Modified Breiman Estimator with fixed-shaped and shape-adaptive kernels on the datasets A_1 and A_2 .

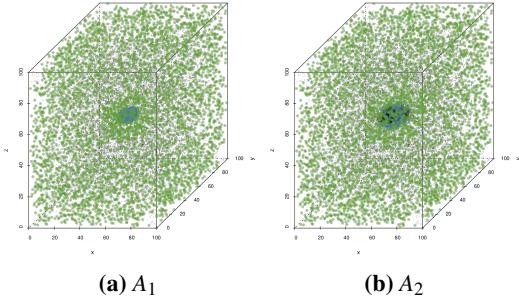


Figure 11: Scatter plot of the data sets (a) A_1 and (b) A_2 with emphasized the points whose kernels with anisotropy in the 90th percentile.

shown in Table 5, clearly both estimators perform better on the dataset with higher values on the diagonal of the covariance matrix of the Gaussian component. Figure 11 illustrates the influence of the density of the Gaussian component on the anisotropy of the kernels near the Gaussian component. 6.23 % of the points with a highly anisotropic kernel in dataset A_1 is sampled from the Gaussian component, in dataset A_2 this is 4.16 %. Although the difference is small, this does illustrate that the density of a component, irrespective of its anisotropy, influences the anisotropy of the kernels.

Suggest increasing K: which problems would that solve, discuss experiment with K10

Suggest not always using a shape-adaptive kernel, only if the local neighbourhood has an anisotropy that is greater than x. Which problems might that solve?

Influence of density suggest on the performance of both estimators suggests that the local bandwidth computation is far from ideal

6 Conclusion