

# Shape-Adaptive Kernel Density Estimation

L.E.N. Baakman

September 13, 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

Kernel density estimation is a popular method to approximate probability densities; in the medical field it has been used to predict dose-volume histograms, which are instrumental in the determination of radiation doses [7]. Ecologists have applied it to explore the habitats of seabirds [6]. Ferdosi et al. [4] have described it as “a critical first step in making progress in many areas of astronomy.” Within this discipline density estimation is, among other things, used to estimate the density of the cosmic density field, which is required for the reconstruction of the large-scale structure of the universe.

Formally the aim of density estimation is to find the probability density  $f(\mathbf{x})$  in the  $d$ -dimensional Euclidean space underlying  $N$  points  $\mathbf{x}_1, \dots, \mathbf{x}_N$ , that have been selected independently from  $f(\mathbf{x})$ .

Kernel density estimation methods approximate  $f(\mathbf{x})$  by placing bumps, referred to as kernels, on the different observations, and summing these bumps to arrive at a final density estimate. This paper is concerned with a method to make the shape of the kernels adaptive to their local neighborhood. Before introducing the process used to determine the form of the kernel we first review different symmetric kernel density estimation methods.

The simplest of which is the Parzen approach [8]. It approximates the density of some pattern  $\mathbf{x}$  ac-

cording to:

$$\hat{f}(\mathbf{x}) = \frac{1}{N} \sum_{i=1}^N h^{-d} K\left(\frac{\mathbf{x} - \mathbf{x}_i}{h}\right). \quad (1)$$

The shape of the used bumps is determined by the kernel function  $K(\bullet)$ , their width by the bandwidth  $h$ . The Parzen approach requires the kernel to be a probability density function, i.e.  $K(\mathbf{x}) \geq 0$  and  $\int K(\mathbf{x}) = 1$  [9]. The bandwidth directly influences the result of the density estimation process; a too small bandwidth results in a density estimate with spurious fine structures, whereas kernels that are too wide can oversmooth the density estimate. Kernel estimators, such as the Parzen approach, that use kernels of the same width for all  $\mathbf{x}_i$ , are called fixed-width estimators.

One downside of these methods is that the height of the peak of the kernel is not data-responsive. Consequently in low density regions the density estimate will have peaks at the few sample points and be too low elsewhere. Whereas in areas with high density the Parzen estimate is spread out, as the sample points are more densely packed together [2]. Adaptive-width methods address this disadvantage by allowing the width of the kernel to vary per data point. For example the estimator introduced by Breiman, Meisel, and Purcell [2] uses the distance between  $\mathbf{x}_i$  and the  $k$ -nearest neighbor of  $\mathbf{x}_i$ , denoted

by  $D_{i,k}$ , to determine the width of the kernel:

$$\hat{f}(\mathbf{x}) = \frac{1}{N} \sum_{i=1}^N (\alpha \cdot D_{i,k})^{-d} K_G \left( \frac{\mathbf{x} - \mathbf{x}_i}{\alpha \cdot D_{i,k}} \right). \quad (2)$$

In this equation  $K_G$  is used to represent a Gaussian kernel, and  $\alpha$  is a multiplicative constant. The values of both  $\alpha$  and  $k$  can be determined with a minimization algorithm on a goodness of fit statistic. Comparing Equations (1) and (2) one finds that the bandwidth  $h$  of the Parzen estimator is defined as  $\alpha \cdot D_{i,k}$  in Equation (2). The factor  $D_{i,k}$  depends on the local neighborhood of  $\mathbf{x}_i$ , in low density regions this factor is large, and the kernel spreads out due to its high bandwidth. In areas with relatively many data points the converse occurs.

Silverman [9] shows that the minimization procedure used by Breiman, Meisel, and Purcell implicitly uses a  $k$ -NN pilot estimate. If pilot estimates, denoted by  $\tilde{f}(\bullet)$ , are used explicitly, the density estimation process becomes:

- (i) Compute pilot densities with some estimator that ensures that  $\forall i \tilde{f}(\mathbf{x}_i) > 0$ .
- (ii) Define local bandwidths  $\gamma$  as

$$\gamma_i = \left( \frac{\tilde{f}(\mathbf{x}_i)}{\text{GM}(\tilde{f}(\mathbf{x}_1), \dots, \tilde{f}(\mathbf{x}_N))} \right)^{-\beta}, \quad (3)$$

where GM denotes the geometric mean and the sensitivity parameter  $\beta$  must lie in the range  $[0, 1]$ .

- (iii) Compute the adaptive kernel estimate as

$$\hat{f}(\mathbf{x}) = \frac{1}{N} \sum_{i=1}^N (h \cdot \gamma_i)^{-d} K \left( \frac{\mathbf{x} - \mathbf{x}_i}{h \cdot \gamma_i} \right) \quad (4)$$

with  $K$  integrating to unity.

Since the pilot densities computed in step (i) do not need to be sensitive to the fine details of the pilot estimate a convenient method, e.g. the Parzen approach, can be used to estimate them [9]. The local bandwidths, computed in step (ii), depend on the exponent  $\beta$ . The higher this value is the more sensitive the local bandwidths are to variations in the pilot densities. Choosing  $\beta = 0$  reduces Equation (4) to a fixed-width method. In the literature two values of  $\beta$  are prevalent. Breiman, Meisel, and Purcell [2] argue that choosing  $\beta = 1/d$  ensures that the number of observations covered by the kernel will be approximately the same in all areas of the data. Whereas Silverman [9] favors  $\beta = 1/2$  independent of the dimension of the data, as this value results in

a bias that can be shown to be of a smaller order than that of the fixed-width kernel estimate.

One disadvantage of the Breiman estimator is its computational complexity. This is partially due to the use of a Gaussian kernel. Because of the infinite base of this kernel an exponential function has to be evaluated  $N$  times to estimate the density of one data point. Wilkinson and Meijer [10] address this in their Modified Breiman Estimator (MBE) by replacing the Gaussian kernel with a spherical Epanechnikov kernel in both the computation of the pilot densities and in the final density estimate. This kernel is defined as

$$K_E(\mathbf{x}) = \begin{cases} \frac{d+2}{2c_d} (1 - \mathbf{x} \cdot \mathbf{x}) & \text{if } \mathbf{x} \cdot \mathbf{x} < 1 \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

where  $c_d$  denotes the volume of the  $d$ -dimensional unit sphere [3]. It should be noted that the kernel defined in Equation (5) does not have unit variance. This can be corrected by multiplying the bandwidth,  $h$ , with the square root of the variance of  $K_E$ , i.e.  $\sqrt{5}$ . There are two advantages to using this kernel, firstly it is computationally much simpler than the Gaussian kernel, partially due to its finite base, and secondly it is optimal in the sense of the Mean Integrated Square Error (MISE) [3]. A downside of this kernel is that it is not continuously differentiable. This is irrelevant when computing the pilot densities, however for the final densities one has to choose between a continuously differentiable density estimate and a density estimator that has a low computational complexity.

Ferdosi et al. [4] consider the application of density estimation on large datasets, i.e. sets with more than 50 000 points, where the dimension of the data points ranges from ten to hundreds of elements. They use the MBE, but introduce a computationally less complex method to estimate the bandwidth. First an intermediate bandwidth for each dimension  $l$  of the data is computed according to

$$h_l = \frac{P_{80}(l) - P_{20}(l)}{\log N}, \quad l = 1, \dots, d, \quad (6)$$

where  $P_{20}(l)$  and  $P_{80}(l)$  are the twentieth and eightieth percentile of the data in dimension  $l$ , respectively. The global bandwidth,  $h$ , is defined as the minimum of these intermediate bandwidths.

Although the widths of the kernels of the discussed adaptive-width methods are sensitive to the data, the shape of a kernel depends only on its definition, and is thus the same for all  $\mathbf{x}_i$ . To further increase the responsiveness of the estimator to the data we propose the use of shape-adaptive kernels;

not only the width but also the shape of these kernels is steered by the local neighborhood of the data.

A possible disadvantage of these shape-adaptive kernels is that in regions where the density of sample points is low, the number of data points is insufficient to reliably compute the shape of the kernel. Therefore we let the amount of influence exerted by the local data on the shape of the kernel depend on the number of data points in the local neighborhood.

This paper is organized as follows. Section 2 introduces the proposed shape-adaptive kernels. The experiment used to investigate the performance of these kernels is discussed in Section 3, the results are presented in Section 4. They are discussed in Section 5, and the reached conclusions can be found in Section 6.

## 2 Method

We use shape adaptive kernels in combination with the Modified Breiman Estimator introduced by Wilkinson and Meijer [10], the resulting estimator is henceforth referred to as the shape-adaptive Modified Breiman Estimator (saMBE). For its low computational complexity we use the method defined in Equation (6) to compute the general bandwidth. Pilot densities are computed according to Equation (1), with an Epanechnikov kernel. Since using  $\beta = 1/2$  instead of  $\beta = 1/d$  results in a final density approximation with a lower mean squared error for most of our datasets we use the first. We compute the final density estimate according to:

$$\hat{f}(\mathbf{x}) = \frac{1}{N} \sum_{i=1}^N \frac{1}{\det(\mathbf{H}_i)} K_{\mathcal{E}}(\mathbf{H}_i^{-1}(\mathbf{x} - \mathbf{x}_i)). \quad (7)$$

The shape of the kernel  $K_{\mathcal{E}}(\bullet)$  is determined by the bandwidth matrix  $\mathbf{H}_i$  [5]. If  $\mathbf{H}_i = h \cdot \gamma_i \cdot \mathbb{I}_{d \times d}$ , Equation (7) reduces to Equation (4).

For each data point  $\mathbf{x}_i$  that is used in the density estimation of some pattern  $\mathbf{x}_j$ , the bandwidth matrix is determined according to these steps:

- (i) Find  $C_{\mathbf{x}_i}$ , the  $k$ -nearest neighbors of  $\mathbf{x}_i$ .
- (ii) Compute  $\Sigma$ , the unbiased covariance matrix of the local neighborhood  $C_{\mathbf{x}_i}$ .
- (iii) Determine  $\mathbf{H}_i$  by scaling  $\Sigma$  with

$$s = h \cdot \gamma_i \left( \prod_{l=1}^d \lambda_l \right)^{-\frac{1}{d}} \quad (8)$$

where  $\lambda_1, \dots, \lambda_d$  are the eigenvalues of  $\Sigma$ .

Step (i) determines the local neighborhood of  $\mathbf{x}_i$  with a  $k$ -nearest neighbors search in a KD-tree [1], with Euclidean distance as the distance metric. We follow Silverman's [9] recommendation of choosing  $k = \sqrt{N}$ . To ensure that  $\Sigma$  is nonsingular we also need  $k > d$ , therefore

$$k = \max \left( \left\lfloor \sqrt{N} \right\rfloor, d \right) + 1.$$

Using a KD-tree for the  $k$ -nearest neighbors search instead of the naive implementation, significantly improves the time complexity of finding  $\mathbf{H}_i$ . The downside of using a space partitioning tree is that  $C_{\mathbf{x}_i}$  is an approximation of the actual neighborhood, as long as  $k$  is rather large the use of an approximation instead of the exact  $k$ -nearest neighbors should not impact the final kernel result too strongly. We use  $k$ -NN rather than a fixed-radius neighborhood to ensure that, independent of the sparsity of the data, the kernel shape is always based on a reasonable number of data points.

The basic shape of the kernel is determined in step (ii). The covariance matrix ensures that the major axis of the kernel has the same direction as the maximum variance of the data.

The scaling factor computed in step (iii) ensures that the kernels used in the density estimation of different patterns have a comparable domain. Equation (8) scales the bandwidth matrix in such a way that the volume of the ellipsoid defined by the eigenvectors and values of  $\mathbf{H}_i$  is equal to that of the eigenellipsoid of the bandwidth matrix that is implicitly used in Equation (4).

## 3 Experiment

We contrast the performance of the shape-adaptive and the symmetric Modified Breiman Estimator on simulated datasets with known density fields. This allows us to test how well the proposed method recovers simple density distributions in comparison to an existing method. The mean squared error (MSE) is used to quantify the performance of the estimators. We use

$$\frac{\max(\lambda_1, \dots, \lambda_d)}{\min(\lambda_1, \dots, \lambda_d)}$$

with  $\lambda_1, \dots, \lambda_d$  the eigenvalues of the bandwidth matrix, to express how anisotropic a kernel is. Two different types of datasets can be distinguished: datasets consisting of a single Gaussian distribution and noise, defined in Section 3.1 and datasets containing multiple Gaussian distributions embed in noise, these sets are presented in Section 3.2.

### 3.1 Datasets with a Single Gaussian

Figure 1 shows a scatter plot representation of the datasets defined in Table 1.

The Gaussian components of these datasets progress from a sphere, i.e. dataset  $S_1$ , to an increasingly more elongated ellipsoid. This makes it possible to investigate the influence of how strongly elongated the distribution is, on the density estimate. The first dataset is a simple spherical Gaussian distribution centered in a uniform random background. The covariance matrix of the Gaussian component in  $S_2$  is created from  $S_1$  by squaring one of the eigenvalues of the covariance matrix, and taking the square root of the other two eigenvalues, without changing the eigenvectors. The resulting covariance matrix defines an eigenellipse with the same volume as the one defined by  $S_1$ . The Gaussian component of dataset  $S_3$  changes the shape of the eigenellipse of the Gaussian component by lengthening one of the minor axes, and shortening the other. In dataset  $S_4$  the Gaussian component is spread out more along the y-axis and less along the z-axis, than the Gaussian component in dataset  $S_3$ .

We expect the Modified Breiman Estimator and its shape-adaptive cousin to perform comparably on dataset  $S_1$ , since due to the symmetric shape of the Gaussian distribution no advantage should be gained by using a shape-adaptive kernel. As the Gaussian distribution is more and more elongated, the advantages of using saMBE should become more pronounced.

### 3.2 Datasets with Multiple Gaussians

Table 2 defines the datasets that consist of uniform random noise and multiple Gaussian distributions, a scatter plot representation of these sets is shown in Figure 2. Dataset  $M_1$  consists of two Gaussian distributions, that are unlikely to overlap, embedded in noise. The first Gaussian component is significantly denser than the second. The procedure outlined in Section 3.1 for the creation of dataset  $S_2$  was used to derive dataset  $M_2$  from  $M_1$ . Dataset  $M_3$  embeds four non-overlapping Gaussians, with eigenspheres with notably different radii, in the uniform random background. The last dataset,  $M_4$ , is a variation on  $M_3$ , created with the method that was used for the definition of dataset  $S_2$  from  $S_1$ .

Due to the spherical nature of the Gaussian components we expect hardly any difference in performance between the estimators on dataset  $M_1$  and  $M_3$ . Given the shape of the Gaussian distributions embedded in dataset  $M_2$  and  $M_4$  we hypothesize that

saMBE outperforms MBE on these sets.

Ferdosi et al. [4] found that the Modified Breiman Estimator resulted in lower integrated squared errors if fewer Gaussian distributions were present in the datasets. Since the presented datasets are comparable to those used by Ferdosi et al. we expect to find the same influence of the number of distributions on the error.

## 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 Tables 1 and 2.

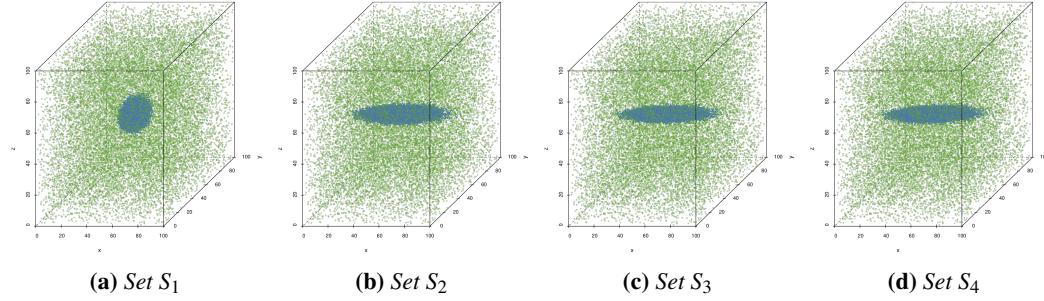
Section 4.1 presents the results of the datasets that contain a single Gaussian, in Section 4.2 the results of the datasets that consist of noise and multiple Gaussian distributions are presented.

### 4.1 Datasets with a Single Gaussian

This section compares the performance of the Modified Breiman Estimator 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 3 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 3 where hardly any difference is visible between Figures 3(a) to 3(d), and Figures 3(e) to 3(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 3(c) and 3(g) the same holds for dataset  $S_3$ . Comparing the performance within datasets between the two

**Before Final Version:** Remove ticks and labels.



**Figure 1:** Scatter plot representation of the datasets defined in Table 1. The used colors correspond to those associated with the different components in Table 1.

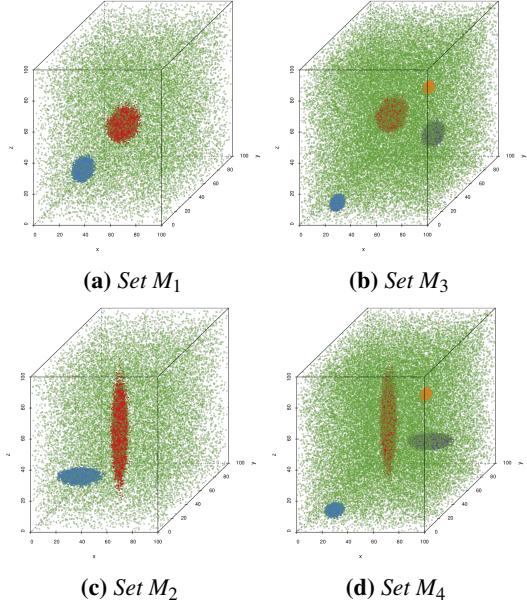
	Component	Number	Distribution
$S_1$	• Trivariate Gaussian	40000	$\mathcal{N}([50, 50, 50], \text{diag}(11))$
	• Uniform random background	20000	$\mathcal{U}([0, 0, 0], [100, 100, 100])$
$S_2$	• Trivariate Gaussian	40000	$\mathcal{N}([50, 50, 50], \text{diag}([11, \sqrt{11}, \sqrt{11}]))$
	• Uniform random background	20000	$\mathcal{U}([0, 0, 0], [100, 100, 100])$
$S_3$	• Trivariate Gaussian	40000	$\mathcal{N}([50, 50, 50], \text{diag}([11, 2 * \sqrt{11}, 1/2\sqrt{11}]))$
	• Uniform random background	20000	$\mathcal{U}([0, 0, 0], [100, 100, 100])$
$S_4$	• Trivariate Gaussian	40000	$\mathcal{N}([50, 50, 50], \text{diag}([11^2, 11, 1]))$
	• Uniform random background	20000	$\mathcal{U}([0, 0, 0], [100, 100, 100])$

**Table 1:** The datasets containing a single Gaussian distribution embedded in uniform noise. The column ‘Number’ indicates for each component the number of patterns sampled from it.  $\mathcal{N}(\mu, \Sigma)$  denotes a Gaussian distribution with mean  $\mu$  and covariance matrix  $\Sigma$ . A diagonal matrix with the values  $x_1, \dots, x_d$  on the diagonal is represented as  $\text{diag}([x_1, \dots, x_d])$ , a scalar matrix with  $x$  on the diagonal is shown as  $\text{diag}(x)$ .  $\mathcal{U}(a, b)$  denotes a uniform distribution with its minimum and maximum set to  $a$  and  $b$ , respectively. The second column presents the symbol used to represent this component in plots throughout the paper.

	Component	Number	Distribution
$M_1$	• Trivariate Gaussian 1	20000	$\mathcal{N}([25, 25, 25], \text{diag}(5))$
	• Trivariate Gaussian 2	20000	$\mathcal{N}([45, 45, 45], \text{diag}(11))$
	• Uniform random background	20000	$\mathcal{U}([0, 0, 0], [100, 100, 100])$
$M_2$	• Trivariate Gaussian 1	20000	$\mathcal{N}([25, 25, 25], \text{diag}([5^2, \sqrt{5}, \sqrt{5}]))$
	• Trivariate Gaussian 2	20000	$\mathcal{N}([45, 45, 45], \text{diag}([\sqrt{11}, \sqrt{11}, 11^2]))$
	• Uniform random background	20000	$\mathcal{U}([0, 0, 0], [100, 100, 100])$
$M_3$	• Trivariate Gaussian 1	20000	$\mathcal{N}([24, 10, 10], \text{diag}(2))$
	• Trivariate Gaussian 2	20000	$\mathcal{N}([33, 70, 40], \text{diag}(10))$
	• Trivariate Gaussian 3	20000	$\mathcal{N}([90, 20, 80], \text{diag}(1))$
	• Trivariate Gaussian 4	20000	$\mathcal{N}([60, 80, 23], \text{diag}(5))$
	• Uniform random background	40000	$\mathcal{U}([0, 0, 0], [100, 100, 100])$
$M_4$	• Trivariate Gaussian 1	20000	$\mathcal{N}([24, 10, 10], \text{diag}([4, \sqrt{2}, \sqrt{2}]))$
	• Trivariate Gaussian 2	20000	$\mathcal{N}([33, 70, 40], \text{diag}([\sqrt{10}, \sqrt{10}, 100]))$
	• Trivariate Gaussian 3	20000	$\mathcal{N}([90, 20, 80], \text{diag}(1))$
	• Trivariate Gaussian 4	20000	$\mathcal{N}([60, 80, 23], \text{diag}([25, \sqrt{5}, \sqrt{5}]))$
	• Uniform random background	40000	$\mathcal{U}([0, 0, 0], [100, 100, 100])$

**Table 2:** The datasets with multiple Gaussian distributions embedded in uniform noise. This table has the same structure and uses the same notation as Table 1.

**Before Final Version:** Remove ticks and labels.



**Figure 2:** Scatter plot representation of the datasets defined in Table 2, the colors used for the different components correspond to those in Table 2.

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

**Table 3:** Performance of the Modified Breiman Estimator with fixed-shaped and shape-adaptive kernels on the datasets with a single Gaussian.

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

Table 4 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

	• Gaussian		• Noise	
	$\mu$	$\sigma$	$\mu$	$\sigma$
$S_1$	1.48	0.521	1.29	0.136
$S_2$	1.57	0.553	1.41	0.289
$S_3$	1.64	0.586	1.51	0.403
$S_4$	1.80	0.698	1.74	0.638
			1.93	0.790

**Table 4:** The mean and the standard deviation of the anisotropy of the kernels used for the datasets with a single Gaussian.

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

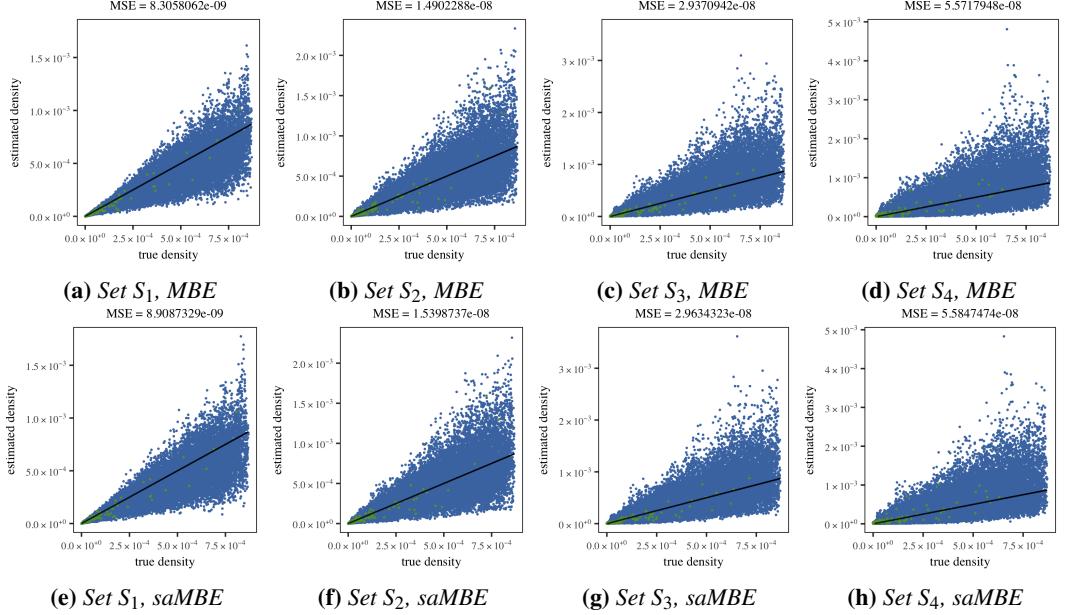
**Table 5:** Performance of the symmetric and the shape-adaptive Modified Breiman Estimator on the datasets containing multiple Gaussian distributions.

stays relatively constant. Furthermore as the Gaussian component is more anisotropic the variation in anisotropy of the kernels increases.

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

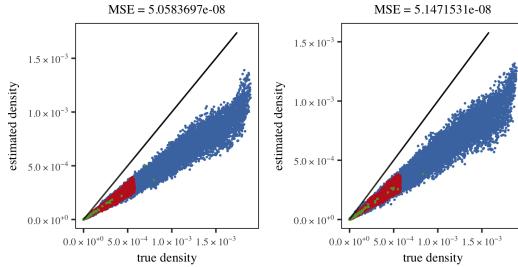
In this section we present the results of the two estimators on dataset  $M_1$ ,  $M_2$ ,  $M_3$ , and  $M_4$ . Based on the small differences between the mean squared errors of the estimators in Table 5 the estimators perform comparably on these datasets. Comparing the MSE between components within the data sets between estimators yields no differences. However within 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  $M_4$ .



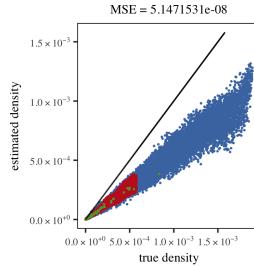
**Figure 3:** Plot of the density as estimated by (a)-(d) MBE and (e)-(h) saMBE as a function of the known density of the datasets with a single Gaussian.

	● Gaussian 1		● Gaussian 2		● Gaussian 3		● Gaussian 4		● Noise	
	$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$
$S_1$	1.50	0.531	1.32	0.175	1.30	0.143			1.89	0.759
$S_2$	1.61	0.570	1.41	0.278	1.49	0.345			1.95	0.783
$S_3$	1.46	0.551	1.29	0.189	1.27	0.130	1.29	0.210	1.28	0.165
$S_4$	1.53	0.571	1.31	0.219	1.49	0.339	1.29	0.210	1.40	0.285
									1.85	0.820

**Table 6:** The mean and standard deviation of the anisotropy of the kernels used for points from the datasets with multiple Gaussians, split per component and for the full dataset.

(a) Set  $M_1$ , MBE

MSE = 5.0497586e-08

(c) Set  $M_1$ , saMBE

MSE = 1.5097586e-08

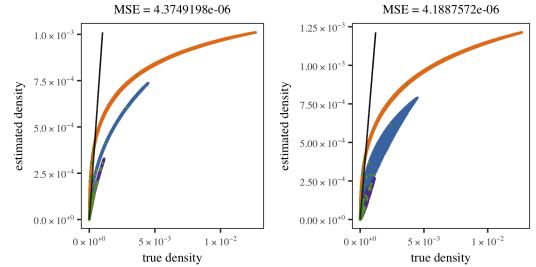
(b) Set  $M_2$ , MBE

MSE = 5.1678718e-08

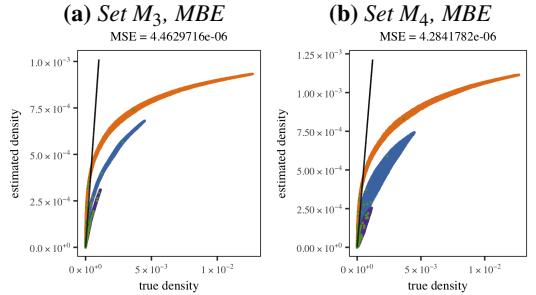
(d) Set  $M_2$ , saMBE

MSE = 5.1678718e-08

**Figure 4:** Plots of the true versus estimated density of datasets  $M_1$  and  $M_2$  for the shape-adaptive and the symmetric Modified Breiman Estimator.

(a) Set  $M_3$ , MBE

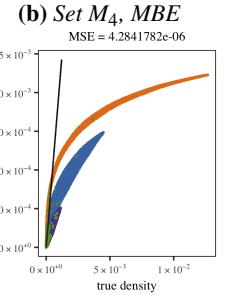
MSE = 4.4629716e-06

(c) Set  $M_3$ , saMBE

MSE = 4.4629716e-06

(b) Set  $M_4$ , MBE

MSE = 4.2841782e-06



**Figure 5:** The estimated density plotted as a function of the true density for datasets  $M_3$  and  $M_4$  for MBE and saMBE.

---

 *old* 

---

In this section we present the results of the two estimators on dataset  $M_1, M_2, M_3, M_4$ .

Comparing Figure 4(a) with Figure 4(c) we find that both estimators underestimate the density and that the densities estimated by the saMBE are spread out more than those estimated by MBE. In spite of this the difference in mean squared error between the two estimators is small enough to be insignificant. The same holds for the mean squared error of the individual components.

Figures 4(b) and 4(d) show the same general trend as Figures 4(a) and 4(c): both estimators underestimate, the shape-adaptive estimator less so than the symmetric estimator, but the differences between the two estimators are small. The differences in MSE within the different components between the estimators are negligible. Comparing the performance of the estimators between datasets  $M_1$  and  $M_2$  we find that the performance of both estimators hardly suffers from the elongation of the Gaussians.

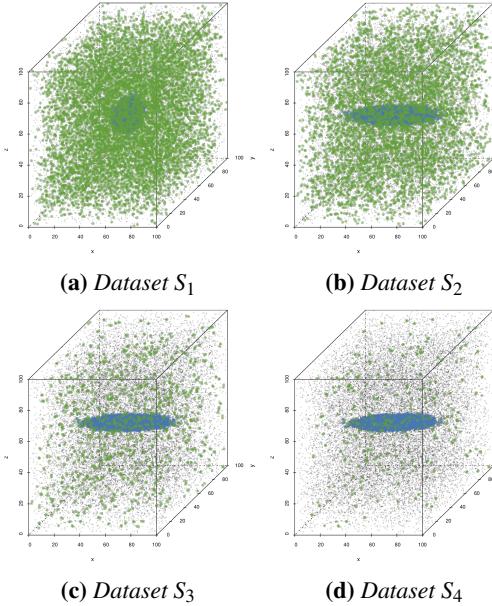
Figures 5(a) and 5(c) clearly show that both estimators significantly underestimate the true density, saMBE more so than MBE. Comparing the mean squared error of the different components we find that both estimators performed worst on the densest component, and best on the component with the highest value on the diagonal of its covariance.

There is no significant difference between the estimators within the different components.

Figures 5(b) and 5(d) shows the same underestimating of densities as the plot of the plots associated with datasets  $M_3$ . Compared to densities estimated for that dataset the range of densities estimated by both estimators for dataset  $M_4$  is greater. The difference in mean squared error within both the complete set and its components between the two estimators is negligible. Contrary to our expectations both estimators perform better on the elongated dataset, i.e.  $M_3$ , than on the spherical set.

In general we have found that the number of Gaussian distributions embed in the noise negatively influences the performance of both estimators. Furthermore the denser a Gaussian distribution is, the more difficulty the estimators have with correctly approximating the density of the points sampled from it.

By comparing the mean squared error of the different components of the datasets we have also found that both estimators are better at estimating the density of points sample from uniform random noise than points sampled from a Gaussian distribution.



**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 points with a higher opacity where the absolute error of saMBE is larger than or equal to the absolute error of MBE.

## 5 Discussion

This section is concerned with the difference in performance between the two estimators within datasets. Section 5.1 focuses on datasets containing a single Gaussian, whereas Section 5.2 discusses the four other datasets.

### 5.1 Datasets with a Single Gaussian

Why does the positive correlation between anisotropy of the Gaussian and the mean and sd of the anisotropy of the kernels make sense?

Why are the differences in anisotropy mostly caused by the Gaussian, what caused the differences between the anisotropy of the noise component: some of the noise points end up in the Gaussian.

The scatter plots of the datasets with a single Gaussian in Figure 6 emphasize the points where the absolute error of the symmetric estimator is smaller than that of the shape-adaptive estimator.

Figure 6(a) shows that the shape-adaptive estimator results outperforms the symmetric estimators on most points near the boundary of the dataset. It also seems to illustrate that that MBE results in a

lower error than saMBE on the other points, however the raw data shows that on  $5.272 \times 10^1$  % of the full dataset, and on  $5.566 \times 10^1$  % of the Gaussian component the symmetric estimator results in a lower absolute error. Reviewing the shape of the kernels used for the points in dataset  $S_1$  we find that the used kernels are all near spherical. The kernels with the largest differences between their eigenvalues are associated with points near the boundary of the dataset. The largest differences in error between the two estimators can be found near the mean of the Gaussian component, where the shape-adaptive kernels are relatively spherical. We think it likely that the difference between the two estimators at these points is caused by the difference in physical density at that location in the dataset, which ensures that a kernel that is slightly elliptical has a result that differs strongly from that of a spherical kernel. This is confirmed by the large differences in the number of patterns that are used in the density estimate of the points near the mean of the Gaussian component between the two estimators.

The results in Figure 6(b) are comparable to those in Figure 1(a), however fewer points of the noise component seem to emphasized in the former. This would mean that the absolute error for most points of the noise component is lower when using symmetric kernels instead of shape-adaptive kernels. Reviewing the raw data shows that this difference is primarily caused by the  $7.803 \times 10^3$  points for which both estimators estimate the same density. We expect that this is due to the elongated shape of the distribution, as its shape influences the kernel of fewer points of the noise component, resulting in more spherical kernels, which causes the estimators to give the same estimate. As in dataset  $S_1$  the points with the most ellipsoidal kernels are positioned near the boundaries of the dataset, where saMBE outperforms MBE. The points whose differences in estimated densities are largest are, as in dataset  $S_1$ , found near the mean of the Gaussian distribution. We hypothesize that this occurs in this dataset for the same reason as it occurs in dataset  $S_1$ .

In Figure 6(c) we observe that the MBE outperforms saMBE on very few points in dataset  $S_3$ , to be exact on 92.71 percent of the points the absolute error of the shape-adaptive estimator was at least as low as that of the symmetric estimator. We expect that the effect of the Gaussian component is stronger than in dataset  $S_2$  since it is more elongated. Reviewing the shape of the kernels we find kernels with large differences between their eigenvalues both near the boundary of the dataset, where saMBE outperforms MBE, and near the Gaussian

component. We expect that the last occurs in this dataset and not in set  $S_2$  due to how much more elongated the Gaussian component is. Near the mean of this component we also find the biggest differences in estimated densities between the two estimators.

The effect of how elongated the Gaussian component is on how well the density of the noise is estimated is stronger in dataset  $S_4$ , as illustrated in Figure 6(d). The density estimate of the two estimates is the same for  $9.540 \times 10^1\%$  of the points drawn from the uniform distribution, contrastingly this is only the case for  $5.575\%$  of the points from the Gaussian component. As in dataset  $S_3$  the shape of the kernels is most strongly influenced by the data near the boundaries of the dataset and the Gaussian component. The largest differences between the results of the two estimators are, comparable to what we observed in dataset  $S_3$ , found near the mean of the Gaussian component.

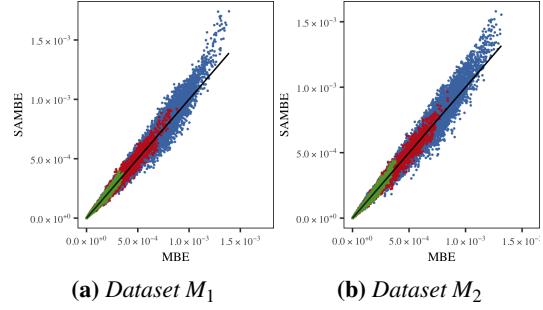
In general we have found that if the Gaussian component is strongly elongated, as in dataset  $S_3$  and  $S_4$ , the shape of the kernels near the mean of the Gaussian component is influenced, whereas the spherical and ellipsoidal Gaussian component in dataset  $S_1$  and  $S_2$ , respectively, hardly influences the shape of the kernels of the data points near its mean. We expect that this is caused by a lower physical density of points near the means of the more spherical Gaussians. In spite of this difference between the four datasets, in all datasets the difference in estimated density between the estimators is largest near the mean of the Gaussian component. We expect that this is due to the relative high density of points at this location, which causes small differences in the shapes of the kernels to have a large effect on the final density.

## 5.2 Datasets with Multiple Gaussians

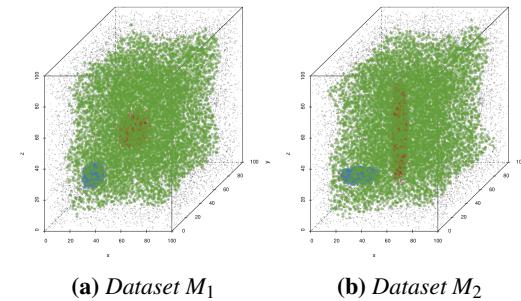
**Why better on more sparse components?**

In Section 4.2 we observed that the differences in performance between the two estimators are small.

Plotting the MBE density as a function of the saMBE density for dataset  $M_1$  and  $M_2$ , see Figure 7, we find that saMBE generally estimates densities to be higher, and nearer to the true density, than MBE. Figure 8 shows that the shape-adaptive estimator outperforms the symmetric estimator on the boundary of both datasets, however the symmetric estimator seems to perform better on the center of the dataset where a lot of points are located. Reviewing the raw data reveals that the latter is not that case, and that the density of approximately half



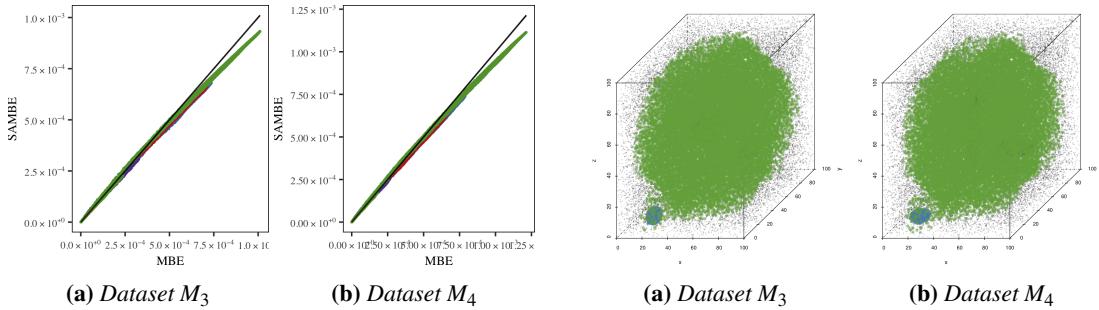
**Figure 7:** Plots of the density estimated by saMBE as a function of those estimated by MBE for dataset (a)  $M_1$  and (b)  $M_2$ .



**Figure 8:** Low opacity scatter plot of dataset (a)  $M_1$  and (b)  $M_2$  with an overlay of high opacity larger points where the absolute error of MBE is smaller than that of saMBE.

of the points are better estimated by the symmetric kernel. Neither dataset shows a correlation between the distance to the mean of a Gaussian component, the error and the used estimator.

Plotting the densities estimated by saMBE versus the densities estimated by MBE for dataset  $S_4$  in Figure 9 shows that the differences between the estimators are as small for nearly all points, as indicated by the mean squared error. The shape-adaptive estimator slightly underestimates points drawn from the noise that have a high known density. As the noise component itself has an uniform density these points are quite likely positioned near the center of one of the Gaussian components. Figure 10 shows that the shape-adaptive estimator outperforms the symmetric estimator consistently on points near the boundary of the dataset. Although it is not visible due to occlusion in Figures 10(a) and 10(b) approximately half of the points in away from the boundary of the dataset have a lower absolute error for saMBE than for MBE. Taking a closer look at the Gaussians we find a correlation between which estimator has the lowest absolute error on a point and the distance of that point to the mean. For dataset  $M_3$  we can say



**Figure 9:** 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 10:** Low opacity scatter plot of dataset (a)  $M_1$  and (b)  $M_2$  with the points where the absolute error of MBE is smaller than the absolute error of saMBE emphasized.

that saMBE performs better on points further away from the mean, whereas MBE is better in approximating the density of points nearer to the mean of the distribution, this effect is especially strong in component ‘Trivariate Gaussian 1’ and ‘Trivariate Gaussian 3’. In  $M_4$  this effect is even stronger. Since the physical density of the points nearer to the mean of the Gaussian is higher we expect that using  $k$  points are insufficient to capture the shape of the Gaussian component, i.e. the local neighborhood is too small.

Interestingly the shape defined by the points where the absolute error of MBE is lower than that of saMBE defines a square in the dataset with two Gaussian components and approximately a sphere in the dataset with four Gaussian components. We expect that this difference is caused by the Gaussian components that are nearer to the boundaries in dataset  $M_3$  and  $M_4$ . Another fascinating difference between the datasets with respectively two and four Gaussian components is the influence of the distance to the mean. We expect that this difference is caused by the lower values on the diagonal of the covariance matrices of the Gaussian components where this effect occurred.

In all datasets we have found a large difference in the shape of the kernels near the boundaries of the dataset. Kernels of a point near an edge of the dataset cube have their shortest minor axis perpendicular to the direction of the edge, to account for the lack of data points in that direction. This boundary effect improves the density estimate in our case, as it allowed a sufficient number of patterns to contribute to the density estimate of points near the boundaries of the dataset. In contrast to the symmetric estimator, which underestimated densities near the boundary due to a lack of contributing points.

## 6 Conclusion

Review hypothesis!

Introduce separate section for further research

Summarize the conclusions.

In conclusion we have found that the shape-adaptive Modified Breiman Estimator gives results comparable to those of the symmetric Modified Breiman Estimator. The former is better in estimating the density of points near the boundary of the dataset, especially if the dataset has multiple Gaussian components. Further research is required to determine if this boundary effect also occurs if the data points near the boundaries are not sampled from a uniform random distribution. Furthermore shape-adaptive kernels are outperformed by symmetric kernels in areas where the physical density of points is high. To counteract this one could consider making the value of  $k$  adaptive. A quick survey showed that increasing  $k$  with a factor ten significantly improved the performance of saMBE on some datasets, however it also caused a drop in performance for the remaining dataset. To counteract this one could consider a  $k$  that is adaptive based on the physical density of its neighborhood.

## References

- [1] Jon Louis Bentley. “Multidimensional Binary Search Trees Used for Associative Searching”. In: *Commun. ACM* 18.9 (1975), pp. 509–517. URL: <http://doi.acm.org/10.1145/361002.361007>.

- [2] L. Breiman, W. Meisel, and E. Purcell. “Variable Kernel Estimates of Multivariate Densities”. In: *Technometrics* 19.2 (1977), pp. 135–144.
- [3] V.A. Epanechnikov. “Non-Parametric Estimation of a Multivariate Probability Density”. In: *Theory of Probability & Its Applications* 14.1 (1969), pp. 153–158.
- [4] B.J. Ferdosi et al. “Comparison of Density Estimation Methods for Astronomical Datasets”. In: *Astronomy & Astrophysics* 531 (2011).
- [5] Wolfgang Härdle et al. *Nonparametric and semiparametric models*. Springer Series in Statistics. Springer Science & Business Media, 2012.
- [6] Kirsty J Lees, Andrew J Guerin, and Elizabeth A Masden. “Using kernel density estimation to explore habitat use by seabirds at a marine renewable wave energy test facility”. In: *Marine Policy* 63 (2016), pp. 35–44.
- [7] Johanna Skarpman Munter and Jens Sjölund. “Dose-volume histogram prediction using density estimation”. In: *Physics in Medicine and Biology* 60.17 (2015), p. 6923.
- [8] E. Parzen. “On Estimation of a Probability Density Function and Mode”. In: *The Annals of Mathematical Statistics* 33.3 (1962), pp. 1065–1076.
- [9] B.W. Silverman. *Density Estimation for Statistics and Data Analysis*. Monographs on Statistics and Applied Probablity. Springer-Science+Business Media, B.V., 1986.
- [10] M.H.F. Wilkinson and B.C. Meijer. “DATAPLOT: A Graphical Display Package for Bacterial Morphometry and Fluorimetry Data”. In: *Computer Methods and Programs in Biomedicine* 47.1 (1995), pp. 35–49.