

# Shape-Adaptive Kernel Density Estimation

L.E.N. Baakman

September 5, 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

Estimating densities with kernels has been fairly popular of late; 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 these bumps adaptive to their local neighborhood. Before introducing the process used to determine the shape of the kernel we first review the different symmetric kernel density estimation methods that our approach is based on.

The Parzen approach [8] is one of the most simple kernel density estimation methods. It estimates the

density of  $\mathbf{x}$  according 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 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 fixed-width methods is that they cannot respond appropriately to variations in the magnitude of the density function, i.e. the peakedness 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. In areas with high density, the sample points are more densely packed together, which causes the Parzen estimate to spread out [2]. Adaptive-width methods address this disadvantage of the fixed-width methods by allowing the width of the kernel to vary per data point. For example the estimator introduced by Breiman, Meisel, and Purcell uses the distance between  $\mathbf{x}_i$  and the  $k$ -nearest neighbor of  $\mathbf{x}_i$ , denoted by  $D_{i,k}$ , to adapt the width

of the kernel:

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

In this equation  $K_{\mathcal{G}}$  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 Equation (1) with (2) one finds that the bandwidth  $h$  of the Parzen estimator is defined as  $\alpha \cdot D_{i,k}$  in Equation (2).  $D_{i,k}$  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  $\gamma_i$  as

$$\gamma_i = \left( \frac{\tilde{f}(\mathbf{x}_i)}{\text{GM}(\tilde{f}(\mathbf{x}_0), \dots, \tilde{f}(\mathbf{x}_N))} \right)^{-\beta}, \quad (3)$$

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 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 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. The Modified Breiman Estimator (MBE), introduced by Wilkinson and Meijer [10], reduces this computational complexity in two ways. Firstly it replaces the infinite base Gaussian kernel with a spherical Epanechnikov kernel in both the computation of the pilot densities and the final density estimate. This kernel is defined as:

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

where  $c_d$  denotes the volume of the  $d$ -dimensional unit sphere. 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  $K_{\mathcal{E}}$ , i.e.  $\sqrt{5}$ . There are two advantages to using this kernel, firstly it is computationally much simpler than the Gaussian kernel, in part due to its finite base, and secondly it is optimal in the sense of the Mean Integrated Square Error (MISE) [3]. One downside of this kernel is that it is not continuously differentiable. This is irrelevant when computing the pilot densities, as they are only used to determine the local bandwidths, however for the final densities it is a trade off between a continuously differentiable density estimate and a density estimator that has a low computational complexity.

The second change Wilkinson and Meijer introduce is the indirect computation of the pilot densities. They are first computed for the vertices of a grid that covers all data points, before the actual pilot densities are determined by multi-linear interpolation. The bandwidth of the kernel used in the computation of the pilot densities is defined as

$$h = \sigma \cdot N^{-1/(d+4)} \left( \frac{8(d+4) \cdot (2\sqrt{\pi})^d}{c_d} \right)^{\frac{1}{d+4}}, \quad (6)$$

where  $\sigma$  represents the square root of the average of the variances of the different dimensions. Wilkinson and Meijer estimate the final densities with Equation (4), using the general and local bandwidths estimated with Equation (6) and (3), respectively.

Ferdosi et al. [4] consider the application of density estimation on large datasets, i.e. sets with more than 50 000 points with the dimension of the data

points ranging 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 determined according to:

$$h_l = \frac{P_{80}(l) - P_{20}(l)}{\log N}, l = 1, \dots, d, \quad (7)$$

where  $P_{20}(l)$  and  $P_{80}(l)$  are the twentieth and eightieth percentile of the data in dimension  $l$ , respectively. To avoid oversmoothing they use this pilot window width:

$$h = \min_l h_l.$$

Although the widths of the kernels of the discussed adaptive-width methods are sensitive to the data, the shapes of the kernels depend only on the kernel itself. 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 experiments used to investigate the performance of these kernels are discussed in Section 3, their results are presented in Section 4. The discussion of these results can be found in Section 5, and the paper is concluded in Section 6.

## 2 Method

We use shape adaptive kernels in combination with the Modified Breiman Estimator introduced by Wilkinson and Meijer [10], we refer to the resulting estimator as the shape-adaptive Modified Breiman Estimator (saMBE). The general bandwidths are computed according to the method introduced by Ferdosi et al. [4] for its lower computational complexity. The resulting bandwidth is used in the estimation of the pilot densities, which is done according to Equation (1), with an Epanechnikov kernel. We have empirically determined that using  $\beta = 1/2$  works better than  $\beta = 1/d$  in combination with shape-adaptive kernels. The final density esti-

mate is computed with an Epanechnikov kernel according to:

$$\hat{f}(\mathbf{x}) = \frac{1}{N} \sum_{i=1}^N \frac{1}{\det(\mathbf{H}_i)} K(\mathbf{H}_i^{-1} (\mathbf{x} - \mathbf{x}_i)). \quad (8)$$

The shape of the kernel  $K(\bullet)$  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 (8) 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

$$s = h \cdot \gamma_i \left( \prod_{l=1}^d \lambda_l \right)^{-\frac{1}{d}} \quad (9)$$

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, however since  $k$  is rather large the use of an approximation instead of the actual  $k$ -nearest neighbors should not impact the final kernel result 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.

Step (ii) determines the basic shape of the bandwidth matrix.

Step (iii) ensures that the kernels used in the density estimation of different patterns have a comparable domain. Equation (9) scales the bandwidth matrix in such a way that the volume of the ellipsoid defined by the eigenvectors and values of  $\mathbf{H}_i$ , modulo the local bandwidth, is equal to that of the eigenellipsoid of the bandwidth matrix that is implicitly used in Equation (4).

### 3 Experiment

We compare the performance of the shape-adaptive method with that of the Modified Breiman Estimator, on simulated datasets with known density fields. This allows us to test how well the proposed method can recover simple density distributions.

To quantify the performance of the estimators we use the Mean Squared Error (MSE):

$$\text{MSE}(\hat{f}(\bullet)) = \frac{1}{N} \sum_{j=1}^N (\hat{f}(\mathbf{x}_j) - f(\mathbf{x}_j))^2.$$

The simulated datasets are a superset of a selection of the sets used by Ferdosi et al. [4]. Figure 1 shows scatter plots of these sets, their definition is given in Table 1.

Dataset one, two, and three, shown in Figures 1a, 1e and 1g, respectively, are taken directly from Ferdosi et al. They consist of a number of spherical Gaussian distributions with random noise added. The means of the Gaussian distribution are chosen in such a way that it is unlikely that the distributions overlap.

Figures 1b, 1f and 1h present dataset four, five, and six. These datasets are created from dataset one, two and, three, respectively. This is done in such a way that the volumes of the eigenspheres of the covariance matrices of the components in the derived datasets have the same volume as the eigenellipsoids of the covariance matrices of the associated components in the original dataset. Furthermore if  $a$  is the eigenvalue of the original covariance matrix, the eigenvalues of the covariance matrices of the derived component are  $a^2$ ,  $\sqrt{a}$  and  $\sqrt{a}$ . Consequently the volumes the eigenspheres of the covariance matrices of the Gaussians in dataset four, five and, six are equal to those of dataset one, two and, three, respectively.

In dataset seven and eight, illustrated in Figures 1c and 1h the semi axes of the ellipsoids all have different lengths. The largest minor axis of the Trivariate Gaussian in dataset seven is a factor two larger than the smallest minor axis in that dataset. Whereas in dataset eight the largest minor axis is exponentially larger than the smallest minor axis.

We expect the MBE and shape-adaptive MBE to perform comparable on dataset one through three, as other than the randomly sampled noise these sets only contain data sampled from a Gaussian distribution with a diagonal covariance matrix. Which results in an equal spread of the data in all dimensions for the non-noise data. Given the elongated shape of the non-noise components in dataset four,

five, six, seven, and eight we hypothesize that the shape-adaptive estimator outperforms the estimator that is not shape adaptive.

The datasets with a single Gaussian which are increasingly more elongated, i.e. dataset one, four, seven and, eight, allow us to investigate the influence of the ellipticalness on the performance of the estimators. Whereas the datasets with multiple ellipsoids, i.e. dataset two, three, five, and six, make it possible to investigate the performance of the classifier on more complex density fields that better approximate real world data.

### 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 2, 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 Table 1 and Figure 1.

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.

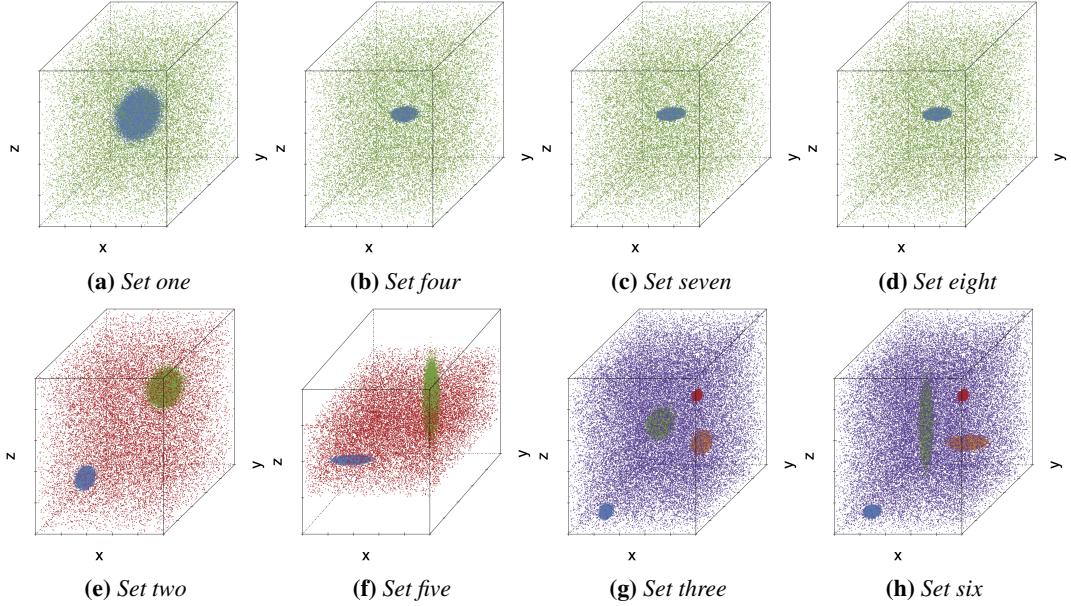
Only on the dataset with a spherical Gaussian is the performance of the two estimators comparable, on all dataset with a single elliptical Gaussian the non-shape adaptive estimator performs significantly better than its shape-adaptive cousin.

Figure 2 presents the results of using the Modified Breiman Estimator and its shape adaptive variant to estimate the densities of the datasets in Table 1 that contain a single Gaussian distribution.

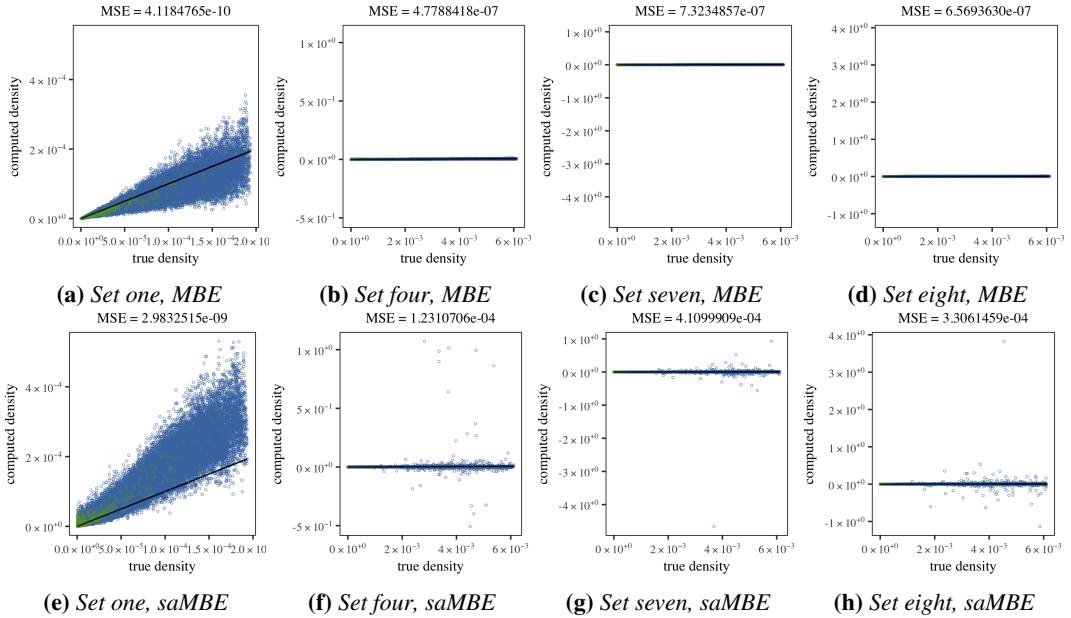
Set	Component	Number	Distribution
one	• Trivariate Gaussian	$4.0 \times 10^4$	$(x, y, z) \sim \mathcal{N}([50, 50, 50], \text{diag}(30))$
	• Uniform random background	$2.0 \times 10^4$	$(x, y, z) \sim \mathcal{U}([0, 0, 0], [100, 100, 100])$
two	• Trivariate Gaussian 1	$2.0 \times 10^4$	$(x, y, z) \sim \mathcal{N}([25, 25, 25], \text{diag}(5))$
	• Trivariate Gaussian 2	$2.0 \times 10^4$	$(x, y, z) \sim \mathcal{N}([65, 65, 65], \text{diag}(20))$
	• Uniform random background	$2.0 \times 10^4$	$(x, y, z) \sim \mathcal{U}([0, 0, 0], [100, 100, 100])$
three	• Trivariate Gaussian 1	$2.0 \times 10^4$	$(x, y, z) \sim \mathcal{N}([24, 10, 10], \text{diag}(2))$
	• Trivariate Gaussian 2	$2.0 \times 10^4$	$(x, y, z) \sim \mathcal{N}([33, 70, 40], \text{diag}(10))$
	• Trivariate Gaussian 3	$2.0 \times 10^4$	$(x, y, z) \sim \mathcal{N}([90, 20, 80], \text{diag}(1))$
	• Trivariate Gaussian 4	$2.0 \times 10^4$	$(x, y, z) \sim \mathcal{N}([60, 80, 23], \text{diag}(5))$
	• Uniform random background	$4.0 \times 10^4$	$(x, y, z) \sim \mathcal{U}([0, 0, 0], [100, 100, 100])$
four	• Trivariate Gaussian	$4.0 \times 10^4$	$(x, y, z) \sim \mathcal{N}([50, 50, 50], \text{diag}([9, \sqrt{3}, \sqrt{3}]))$
	• Uniform random background	$2.0 \times 10^4$	$(x, y, z) \sim \mathcal{U}([0, 0, 0], [100, 100, 100])$
five	• Trivariate Gaussian 1	$2.0 \times 10^4$	$(x, y, z) \sim \mathcal{N}([25, 25, 25], \text{diag}([25, \sqrt{5}, \sqrt{5}]))$
	• Trivariate Gaussian 2	$2.0 \times 10^4$	$(x, y, z) \sim \mathcal{N}([65, 65, 65], \text{diag}([\sqrt{20}, \sqrt{20}, 400]))$
	• Uniform random background	$2.0 \times 10^4$	$(x, y, z) \sim \mathcal{U}([0, 0, 0], [150, 150, 150])$
six	• Trivariate Gaussian 1	$2.0 \times 10^4$	$(x, y, z) \sim \mathcal{N}([24, 10, 10], \text{diag}([4, \sqrt{2}, \sqrt{2}]))$
	• Trivariate Gaussian 2	$2.0 \times 10^4$	$(x, y, z) \sim \mathcal{N}([33, 70, 40], \text{diag}([\sqrt{10}, \sqrt{10}, 100]))$
	• Trivariate Gaussian 3	$2.0 \times 10^4$	$(x, y, z) \sim \mathcal{N}([90, 20, 80], \text{diag}(1))$
	• Trivariate Gaussian 4	$2.0 \times 10^4$	$(x, y, z) \sim \mathcal{N}([60, 80, 23], \text{diag}([25, \sqrt{5}, \sqrt{5}]))$
	• Uniform random background	$4.0 \times 10^4$	$(x, y, z) \sim \mathcal{U}([0, 0, 0], [100, 100, 100])$
seven	• Trivariate Gaussian	$4.0 \times 10^4$	$(x, y, z) \sim \mathcal{N}([50, 50, 50], \text{diag}([9, 2 * \sqrt{3}, 1/2 * \sqrt{3}]))$
	• Uniform random background	$2.0 \times 10^4$	$(x, y, z) \sim \mathcal{U}([0, 0, 0], [100, 100, 100])$
eight	• Trivariate Gaussian	$4.0 \times 10^4$	$(x, y, z) \sim \mathcal{N}([50, 50, 50], \text{diag}([9, 3, 1]))$
	• Uniform random background	$2.0 \times 10^4$	$(x, y, z) \sim \mathcal{U}([0, 0, 0], [100, 100, 100])$

**Table 1:** The datasets used to test the estimators. The column ‘Number’ indicates for each component of the dataset how many data points are sampled from that component.  $\mathcal{N}(\mu, \Sigma)$  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 colors shown in the second column correspond with the colors used for these components of the data set throughout the paper.

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



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



**Figure 2:** Comparative plots for dataset 1, 4, 7, and 8.

Figure 2a 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 2a, 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 2b and 2f 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 Fig-

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

**Table 2:** 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 Table 1.

ures 2b and 2f 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 2d and 2h 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

In this section we present the results of the two estimators on dataset two, five, three, six, i.e. the datasets that contain more than one Gaussian.

The plots in Figure 3 suggest some differences in how the two estimators handle the different components of the dataset, therefore Table 3 presents the mean square error of the different components of the datasets that contain multiple Gaussians.

Table 3 shows some difference in MSE between the different components of dataset two, and the elongated version of this dataset, i.e. dataset five: both estimators perform best on data points that were drawn from the uniform distribution, and worst on Gaussian 1, the Gaussian component with the smallest radius. Furthermore saMBE outperforms MBE on the component named ‘Gaussian 2’.

Comparing Figures 3a, 3b, 3e and 3f we find that on both datasets both estimators underestimate the densities. Although the ranges of the densities estimated by saMBE are slightly higher than those estimated by the Modified Breiman Estimator they do not differ as extremely as in dataset one, four, seven, and eight. The difference in range in dataset two seems to be caused by the pattern in the upper right corner of Figure 3e.

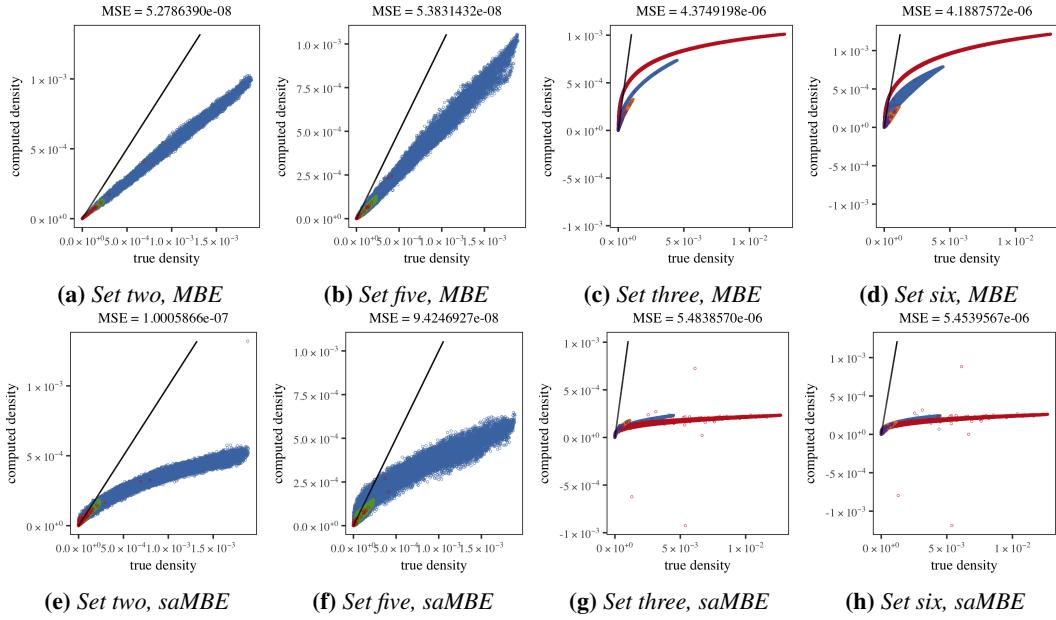
Comparing the MSE of the components of dataset three, and its elongated variant, dataset six, we find that on the Gaussian components in both datasets the estimators performed comparably, with the saMBE always resulting in the highest mean square error. The difference in performance between the two estimators is higher on the noise component than on the Gaussian component. As for dataset two and five we also find that for both estimators the MSE on the noise component is significantly lower than on the Gaussian component.

Figures 3c and 3g show that both estimators underestimate the densities. Figure 3g confirms what the mean square errors indicated, namely that the difference between the density estimate of saMBE and the known density is greater than the difference between the densities estimate by MBE and the true density. This figure also shows that the shape-adaptive estimator has resulted in extreme densities, with values as high as  $8.826 \times 10^{-4}$  and as low as  $-1.189 \times 10^{-3}$ .

We can conclude that on all datasets with multiple Gaussian components the non-shape adaptive estimator outperformed the shape-adaptive variant. Furthermore both estimators are better at estimating the density of the noise in the dataset than the density of points that were sampled from the Gaussians.

Set	Component	Estimator	
		MBE	saMBE
two	Gaussian 1	$1.561 \times 10^{-7}$	$2.993 \times 10^{-7}$
	Gaussian 2	$2.285 \times 10^{-9}$	$8.345 \times 10^{-10}$
	Noise	$2.021 \times 10^{-11}$	$2.363 \times 10^{-11}$
five	Gaussian 1	$1.587 \times 10^{-7}$	$2.810 \times 10^{-7}$
	Gaussian 2	$2.771 \times 10^{-9}$	$1.769 \times 10^{-9}$
	Noise	$4.829 \times 10^{-12}$	$4.865 \times 10^{-12}$
three	Gaussian 1	$2.386 \times 10^{-6}$	$3.377 \times 10^{-6}$
	Gaussian 2	$1.189 \times 10^{-8}$	$1.276 \times 10^{-8}$
	Gaussian 3	$2.374 \times 10^{-5}$	$2.936 \times 10^{-5}$
	Gaussian 4	$1.114 \times 10^{-7}$	$1.570 \times 10^{-7}$
	Noise	$4.287 \times 10^{-10}$	$6.279 \times 10^{-10}$
six	Gaussian 1	$2.328 \times 10^{-6}$	$3.353 \times 10^{-6}$
	Gaussian 2	$1.748 \times 10^{-8}$	$1.899 \times 10^{-8}$
	Gaussian 3	$2.266 \times 10^{-5}$	$2.918 \times 10^{-5}$
	Gaussian 4	$1.312 \times 10^{-7}$	$1.680 \times 10^{-7}$
	Noise	$3.886 \times 10^{-10}$	$5.887 \times 10^{-10}$

**Table 3:** 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 different components of the datasets with multiple Gaussians.



**Figure 3:** 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 2 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 2 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 Equation (8) 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 datasets.

### 5.2 Datasets with Multiple Gaussians

What does this section do?

Waarom zijn de resultaten voor beide estimators op noise component beter dan op de gaussians?

Waarom doe SAMBE het consistent slechter dan MBE?

Performs best on Gaussian with largest radius, reason: Most like uniform noise on which both estimators perform best?

Both have difficulty with Gaussian with smallest radius

Why are the densities underestimated?

Waarom hier weer negatieve resultaten, waarom niet in baakman2?

General Discussion

## 6 Conclusion

### 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 Probability. 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.