Visualization of Multivariate Density Estimates with Level Set Trees

Jussi Klemelä *

Abstract

We present a method for visualization of multivariate functions. The method is based on a tree structure, built from separated parts of level sets of a function, which we call level set tree. The method is applied for visualization of estimates of multivarate density functions. With different graphical representations of level set trees we may visualize the number and location of modes, excess masses associated with the modes, and certain shape characteristics of the estimate. We present simulation examples where projecting data to two dimension does not help to reveal the modes of the density, but with the help of level set trees one may detect the modes. We argue that level set trees provide a useful method for exploratory data analysis.

Keywords: Cluster analysis; Exploratory data analysis; Mixtures; Mode detection; Multivariate data.

1 Introduction

Nonparametric density estimators have been succesfully applied in exploratory data analysis for one and two dimensional data. For example, it is possible to detect modes by the inspection of one and two dimensional density estimates. For more than two dimensional data the difficulties with visualizing density estimates have often hindered the application of nonparametric density estimation. We construct a method for visualization of multivariate functions which can increase usefulness of multivariate density estimates in exploration and mining of multivariate data.

We present a method of visualization which is based on the level sets of the function. A level set is the set of those points at which the function exceeds a given value; level set of function $f : \mathbf{R}^d \to \mathbf{R}$ at level α is defined as

$$\Lambda_\alpha = \left\{ x \in \mathbf{R}^d : f(x) \geq \alpha \right\}. \tag{1}$$

*Researcher, Institut für Angewandte Mathematik, Universität Heidelberg, Im Neuenheimer Feld 294, 69120 Heidelberg, Germany (email: klemela@statlab.uni-heidelberg.de)

We assume that the function to be visualized has only a finite number of different level sets, in other words, it is piecewise constant. If the function is not piecewise constant, then we will approximate it with such a function, as will be done for kernel estimates with a continuous kernel.

A level set tree is a tree structure formed by taking as root nodes the separated regions of the lowest level set of the function. The child nodes of a given node correspond to the separated regions of a part of the level set whose level is one step higher than the level of this parent node. Thus the disjoint regions of a level set will be startpoints for the different branches of the tree.

In density estimation we are interested in the shape of the density function: the number and location of modes, relative size of the modes, skewness, kurtosis, tail behaviour of the function, and so on. The level set tree will reflect the mode structure of the function, because different branches of the tree will correspond to distinct modes (local extremes) of the function. By comparing level sets related to different levels one can find information on the shape of the density, also in high dimensional spaces. In particular, we may look how the volumes and barycenters of level sets are changing as a function of the level.

We present the *volume plot* and the *barycenter plot* for visualization of density functions. The *volume plot* visualizes the number and relative size of the modes of the density, and gives information on the kurtosis. With a mode we mean a local extreme of the density and with size of a mode we mean the probability mass associated with this local extreme, that is, the excess mass of the local extreme. The *barycenter plot* draws the "skeleton" of the function, visualizing locations of the modes and giving information on the skewness.

We claim that even in cases where projections may reveal the modes of the underlying density, level set trees provide an easy to use exploratory method which gives additional insight into the shape of the density. Furthermore, there exist examples where one or two dimensional marginal densities do not reveal the true number of modes of the density. We show this by constructing examples of mixtures of Gaussian densities, where the components of the mixture are so close to each other that all marginal densities show only few modes. For these examples level set trees however provide a method for finding the number and locations of the modes.

Level set trees provide a method for visualizing density estimates. Methods for making inference whether local maxima of a density estimate correspond to the modes of the underlying density has to be studied elsewhere. For visualization and inference concerning the mode structure of one and two dimensional densities, see for example Minnotte and Scott (1993), Marchette and Wegman (1997), Minnotte, Marchette and Wegman (1998), Chaudhuri and Marron (1999), Godtliebsen, Marron and Chaudhuri (2002). Various mode testing procedures are presented in Silverman (1981), Hartigan and Hartigan (1985), Müller and Sawitzki (1991), Hartigan and Mohanty (1992), Mammen, Marron and Fisher (1992), Fisher, Mammen and Marron (1994), Minnotte (1997), and Davies and

Kovac (2001). Since volume plots visualize excess masses associated with the local extremes of the density, they help to make a judgement whether the local extremes of an estimate correspond to the true modes of the underlying density function. Excess masses in mode detection has been applied for example by Müller and Sawitzki (1991).

In Section 2.1 we define level set trees for general multivariate functions. In Sections 2.2 and 2.3 we define volume plot and barycenter plot and discuss basic diagnostics with the help of these plots. In Section 2.4 we discuss the computational complexity of calculating a level set tree. In Section 3 we illustrate the level set trees used for visualizing histograms and kernel estimates. In Section 4 we give examples of the estimation of multimodal densities. Examples are 3 and 4 dimensional mixtures of standard Gaussian densities. Section 5 contains a summary and discusses further work.

Computations and graphics in this article have been made with an R-package called "denpro". This package may be downloaded from http://denstruct.net.

2 Definition of level set trees and level set plots

We will define the level set tree, define the volume plot and the barycenter plot, discuss basic diagnostics which can be made using these plots, and finally discuss the computational complexity of calculating the level set trees in some typical examples.

2.1 Definition of the level set tree

The function from which we form the level set tree is assumed to be piecewise constant. Thus the function has a finite number of distinct level sets. The level set tree is a tree whose nodes represent separated subsets of the level sets of the function. It is possible to define a corresponding structure for continuous functions but we consider only the discrete case because this case is relevant for the practical calculation of level set trees.

We say that the sets $A, B \subset \mathbf{R}^d$ are *separated* if $\inf\{\|x-y\| : x \in A,\ y \in B\} > 0$ where $\|\cdot\|$ denotes Euclidean distance. Thus, two sets are said to be separated if there is some space between them. We say that set $A \subset \mathbf{R}^d$ is *connected* if for each nonempty $B, C \subset \mathbf{R}^d$ such that $A = B \cup C$, sets B and C are not separated. Thus, a set is said to be connected if it cannot be written as a union of two separated sets.

The level set tree may have an arbitrary finite number of root nodes and every node may have an arbitrary finite number of child nodes. Root nodes of the level set tree correspond to separated regions of the lowest level set of the function. The child nodes of a given parent node correspond to certain separated regions of the level set whose level is one step higher than the level of the parent node. To

every node we associate a real value and a set. The set associated with the node is the corresponding separated subset of the level set and the value associated with the node is the minimum value of the function on the set associated with this node.

Let $f : S \to \mathbf{R}$, $S \subset \mathbf{R}^d$, be a function whose range is a finite set:

$$\{f(x) : x \in S\} = \{\lambda_1, \ldots, \lambda_N\} \tag{2}$$

where $\lambda_1 < \cdots < \lambda_N$.

Definition 1 A level set tree *is a multi-tree whose nodes are annotated with pairs* (a, A)*, where* $a \in \mathbf{R}$ *and* $A \subset \mathbf{R}^d$*. We call value* $a \in \mathbf{R}$ *the level of the corresponding node. We give a recursive definition of the level set tree of function* f *satisfying (2).*

1) Write the lowest level set of function f *as*

$$\Lambda_{\lambda_1} = A_1 \cup \cdots \cup A_M$$

where A_j*,* $j = 1, \ldots, M$*, are pairwise separated and each* A_j *is connected (no* A_j *can be further written as a union of two separated sets). Then the level set tree has* M *root nodes and to these nodes we associate sets* A_j *and values* $a_j = \min\{f(x) : x \in A_j\}$*,* $j = 1, \ldots, M$*.*

2) Assume that we have a node of the tree for which there is associated set $A \subset \mathbf{R}^d$ *and value* $a \in \mathbf{R}$*. If* $\{x \in A : f(x) > a\} = \emptyset$*, then this node does not have children. Otherwise, if* $\{x \in A : f(x) > a\} \neq \emptyset$*, write*

$$\{x \in A : f(x) > a\} = B_1 \cup \cdots \cup B_L$$

where B_j*,* $j = 1, \ldots, L$*, are pairwise separated and each* B_j *is connected. The given node has then* L *children with which we associate sets* B_j *and values* $b_j = \min\{f(x) : x \in B_j\}$*,* $j = 1, \ldots, L$*.*

From now on we will assume that function f is a density. Then we may without loss of generality assume that $\lambda_1 > 0$ (since for densities we may without loss of generality assume that $\Lambda_0 = \mathbf{R}^d$). This implies that level set Λ_{λ_1} of f has finite volume.

We will illustrate the definition by an example. In Figure 1 we display a density function which takes 5 different values and has two modes. In Figure 2 we display the corresponding level set tree. Two modes of the function are represented as two branches of the tree.

The separated parts of the level sets which we associate with the nodes of the level set tree usually have a complex structure. Indeed, for a function $f : \mathbf{R}^d \to \mathbf{R}$, the level sets are subsets of $\mathbf{R}^d$. Only in the three-dimensional case, when $f : \mathbf{R}^3 \to \mathbf{R}$, we may succeed in direct visualization of these sets. Our strategy

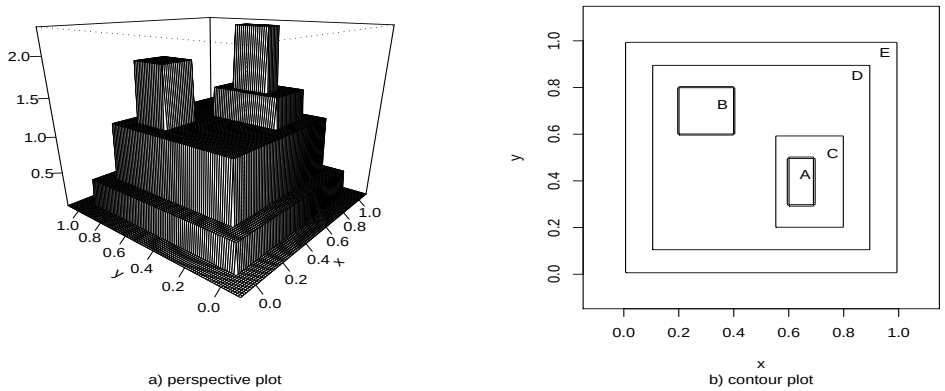

Figure 1: Piecewise constant function

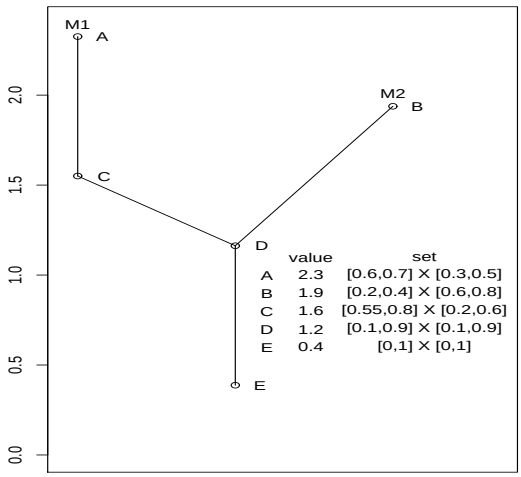

Figure 2: Level set tree of the function of Figure 1. For each node we associate a value and a set.

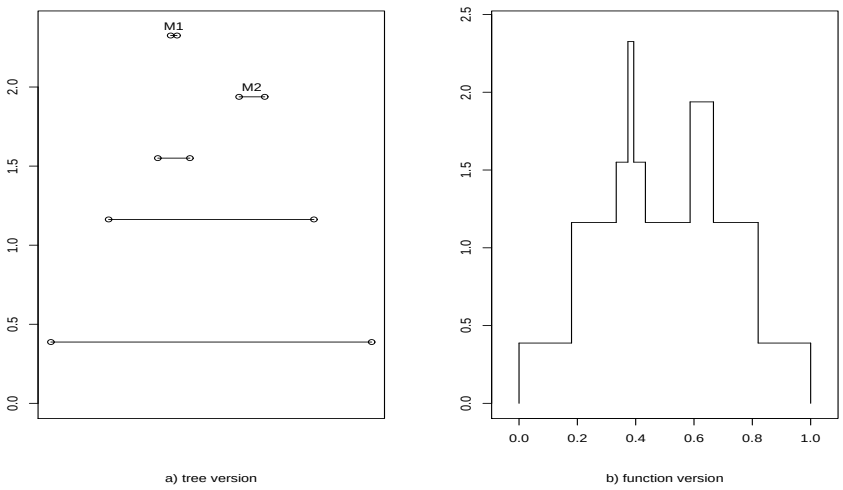

Figure 3: Volume plots of the function of Figure 1. Frame a) shows a tree version and frame b) shows a function version.

is to calculate certain characteristics of the sets and use these characteristics in visualization. In this article we utilize only volumes and barycenters of the sets associated with the nodes in visualizing the level set tree. The volume of a set $A \subset \mathbf{R}^d$ is volume$(A) = \int_A dx$ and the barycenter of A is

$$\text{barycenter}(A) = \frac{1}{\text{volume}(A)} \int_A x\, dx.$$

The barycenter is a d-dimensional vector giving the "center of mass" of a set. We call *volume plot* a plot which visualizes the volumes of separated parts of level sets. We call *barycenter plot* a plot which visualizes the barycenters.

2.2 Volume plot

The standard tree plot of a level set tree as in Figure 2 visualizes the number of local extremes and the levels of those extremes. With volume plots we visualize the importance of local extremes in terms of the excess mass. Figure 3 shows examples of volume plots. Figure 3 a) is a tree version of a volume plot and Figure 3 b) is a function version of a volume plot.

Tree version of the volume plot. In the tree version of the volume plot nodes of the level set tree are represented as horizontal lines. The height of a line representing a node is determined by the level of the node (the level of the corresponding level set). The length of the line representing a node is proportional to the volume of the corresponding set. The parent-child relations are expressed by the left-right positioning of the nodes, so that the horizontal space a child

node occupies is contained to the horizontal space occupied by the parent. This is possible, since the sum of the volumes of the child nodes is always less than the volume of the parent node.

The left-right ordering of siblings (root nodes and the child nodes of a given node) may be done in various ways. We have applied the following rule in this article.

1. Order first the root nodes. The leftmost root node is the one with the largest Euclidean distance of the barycenter from the origin. After that, the next node is the one with the closest Euclidean distance of the barycenter from the barycenter of the previous node.

2. The children of a node will be ordered by the same rule as the root nodes were ordered.

Above we used for simplicity the phrase "the barycenter of a node" when we meant the barycenter of the set associated with the node.

To show details in the upper levels of a volume plot we use a *zoomed volume plot*, which is a plot showing only the upper levels of the volume plot. Figure 11 and Figure 13 in Section 4 show examples of zoomed (function versions of) volume plots.

Function version of the volume plot. We may associate one dimensional density functions with each tree version of a volume plot. We call these one dimensional density functions *volume plot transformations*. Volume plot transformations give certain one dimensional representations of the multivariate density, which are not any slices, marginal densities, or conditional densities of the original density. We may define volume plot transformations in the following way: a one dimensional density is a volume plot transformation of the multivariate density f, if it belongs to the equivalence class of one dimensional functions whose tree version of the volume plot is identical with the tree version of volume plot of f. This equivalence class is closed with respect to translations (shiftings). We will always choose the representative g from the equivalence class which is such that $\inf\{t : g(t) \geq 0\} = 0$. In addition, we choose the representative which is not skewed; that is, when a node of the level set tree of g has only one child, then this child has the same barycenter as the parent.

Let us denote with $v_p(f) : \mathbf{R} \to \mathbf{R}$ a volume plot tranformation of a multivariate density $f : \mathbf{R}^d \to \mathbf{R}$. We may justify the volume plot transformation with the following 2 facts: (1) the level sets of $v_p(f)$ have as many pairwise separated and connected components as the level sets of f and (2) for all $\alpha \geq 0$, $\int_{(f \geq \alpha)} f = \int_{(v_p(f) \geq \alpha)} v_p(f)$. The second fact states that the *excess masses* of densities are equal for all levels. Facts (1) and (2) say together that $v_p(f)$ has *isomorphic mode structure* with f.

The excess mass associated with a node of the level set tree. We may state the mode isomorphism of f and $v_p(f)$ still in other way with the help of excess masses. Excess mass may be associated with every node of a level set tree. To define the excess mass associated with a node we introduce the following notation. Assume that with node n of a level set tree are associated value a and set A. Then we write

$$n = (a, A), \qquad \text{set}(n) = A, \qquad \text{val}(n) = a.$$

Furthermore, with $\text{parent}(n)$ we mean the unique parent of node n. We say that a node is a decendant of node n if it is either a child of n or a child of an other decendant of n.

Definition 2 The excess mass associated with node n *of the level set tree of density function* f *is defined by*

$$\begin{aligned} \text{excmass}(n) &= \int_{\text{set}(n)} (f(x) - \text{val}(\text{parent}(n)))\, dx \\ &= \sum \{\text{volume}(\text{set}(n_0)) \cdot [\text{val}(n_0) - \text{val}(\text{parent}(n_0))] : \\ &\qquad n_0 = n \text{ or } n_0 \text{ is a decendant of } n\}, \end{aligned} \tag{3}$$

where density f *satisfies (2) with* $\lambda_1 > 0$. *For the case that* n *is a root node, we denote* $\text{val}(\text{parent}(n)) = 0$.

In words, the excess mass is the volume of the area which the function delineates over a given level, in a given branch of the level set tree. When a level set tree has only one root node, then the excess mass of this root node is equal to one. Excess masses of the other nodes are fractions of the total probability mass. We state the mode isomorphism of f and $v_p(f)$ in the following way: the level set trees of f and $v_p(f)$ are isomorphic and excess masses associated with the corresponding nodes of these trees are equal. Excess masses has been applied in cluster analysis and mode testing for example by Hartigan (1987), Müller and Sawitzki (1991), Minnotte (1997).

Diagnostics on kurtosis. By comparing volumes of level sets at different levels we may get information about the kurtosis. If the volumes of level sets are decreasing fast when we move to the level sets corresponding to higher levels, this may indicate that the density has sharp peaks.

2.3 Barycenter plots

The barycenter plot draws the "skeleton" of the function, visualizing locations of the modes and giving information on the skewness. Figure 4 shows an example of a barycenter plot.

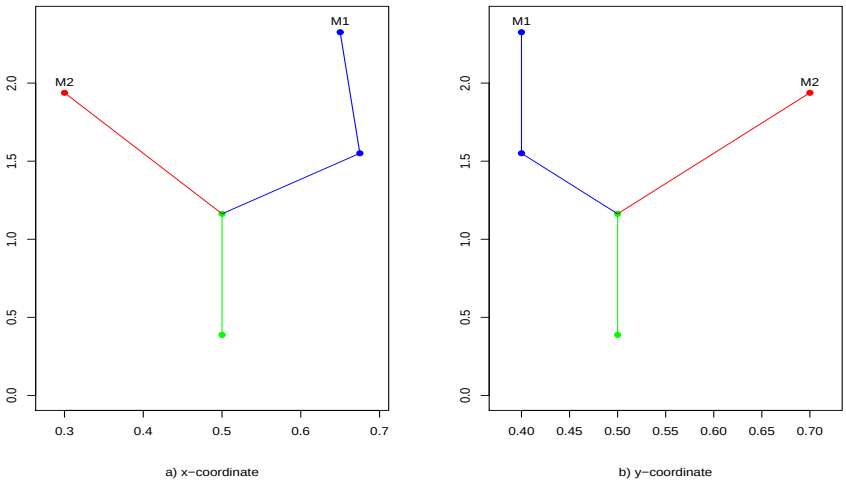

Figure 4: Barycenter plot of the function of Figure 1.

The barycenter plot consists of d windows when the function is d dimensional. We have a window for each coordinate. Each window shows the positions of one coordinate of barycenters for different levels. Barycenter plots are tree structured plots where the vertical position of a node is determined by the level of the corresponding separated component of a level set, and the horizontal positioning of a node in the ith window is determined by the ith coordinate of the barycenter of the corresponding separated component of a level set. A parent-child relation is expressed by the straight line joining the parent with the child. Note that these lines may have crossings.

When density $f : \mathbf{R}^d \to \mathbf{R}$ is unimodal, the barycenter plot visualizes the 1-dimensional curve $b_c : [0, M] \to \mathbf{R}^d$, $b_c(\alpha) = \text{barycenter}(\Lambda_\alpha)$, in d-dimensional space, where $M = \sup_{x \in \mathbf{R}^d} f(x) < \infty$ and Λ_α is the level set defined in (1). In the general multimodal case Λ_α may have many separated components and the barycenter plot visualizes the mapping $b_c : [0, M] \to (\mathbf{R}^d)^\infty$, $b_c(\alpha) = (b_1, \ldots, b_l)$ where $b_i \in \mathbf{R}^d$ is the barycenter of the ith separated component of Λ_α.

To identify the nodes between different windows of a barycenter plot (and between volume plots and a barycenter plot) we label the modes. The labeling of modes will uniquely determine the correspondence of all nodes in different windows. However, to ease the identification of nodes across different windows, we will also color the nodes. We will first choose distinct colors for the leaf nodes and then travel towards the root nodes, by changing the color always when two branches are merging. We will also color the lines joining a child and a parent. The color of a line will be the same as the color of the child node which is at the child end (upper end) of the line. This printed version of the article shows the colors only in gray scale.

Diagnostics on skewness. If the barycenters of the level sets are not the same for different levels, this may indicate skewness of the function. In Figure 4 one may note the slight skewness of the branch leading to the mode labelled as $M1$, by noting the change of x-coordinate of barycenters as the level grows.

2.4 Computational complexity

Algorithms for finding the level set tree depend on the underlying function. However, one may formulate some algorithms which apply for a number situations.

In this article we consider density estimators whose level sets may be written as unions of a finite number of connected sets which belong to a basic library of "atoms". Level sets of histogram estimates are unions of bins. A kernel estimate may be approximated by calculating its values at gridpoints, and forming a piecewise constant function which is constant over the rectangles whose centers are the gridpoints. Then these rectangles centered at the gridpoints form the library of atoms.

The main step in algorithms for forming the level set tree is to find the maximally separated regions of a given (part of a) level set. That is, we want to find a partition of the level set to pairwise separated sets, where each set is connected. A naive algorithm for decomposing a level set, which is a union of atoms, to the pairwise separated connected components is based on pairwise comparisons of atoms to find which atoms touch each other.

More precisely, for a given atom A we find all atoms which touch atom A and put these atoms to a stack. We pull atoms from the stack one at a time and find atoms which touch this atom pulled from the stack, and put also these atoms to the stack (unless they were already encountered). Continuing in this way as long as the stack is not empty we find the separated connected component whose one member is atom A. This procedure must be repeated for all atoms not yet associated to some component.

Let us analyze the complexity of finding the level set tree for the case in which we apply the naive algorithm for finding the separated connected components of level sets. Assume that the given estimate has Q different level sets and every level set is a union of at most N atoms. For each level set, naive algorithm makes at most $N(N-1)$ tests between atoms to find which atoms touch each other. In our examples testing whether two atoms touch each other takes d steps, when d is the dimension of the Euclidean space where the estimate is defined. Then, under these assumptions, construction of the level set tree from the estimate will take

$$O\left(QN^2d\right) \tag{4}$$

steps.

For the case of the histogram estimator the naive algorithm is feasible. For the case of a kernel estimate evaluated on a grid this algorithm is not feasible,

since the multivariate grid with the help of which we approximate the kernel estimate may contain a huge number of knots. One may develop more sophisticated algorithms which are based on a dynamic programming algorithm, which finds solutions for spatially local subsets of the support of the estimate, and builds the global solution from the previously found solutions to the local problems. This kind of algorithm is described in Klemelä (2005).

3 Level set trees from histograms and kernel estimates

We will consider visualization of histogram and kernel estimates with level set trees.

3.1 Histogram estimator

We construct histogram estimates by first finding minimum and maximum values of every variable, constructing the rectangle whose vertices are these minima and maxima, constructing bins by dividing every side of the rectangle into an equal number of intervals, and finally counting the number of observations in each bin.

Figure 5 shows examples of two dimensional histograms. These histograms are constructed from a sample of size 200 from a density, which is an equal mixture of three standard Gaussian densities. The three components of the mixture are $N(\mu_i, I)$, $i = 1, 2, 3$, where $\mu_1 = (0, 0)$, $\mu_2 = (0, 4)$, $\mu_3 = (4, 0)$, and I is 2×2 unit matrix. Figure 5 shows two histograms with 8^2 and 13^2 bins. On the left hand side we show perspective plots and on the right hand side volume plots. In Figure 6 we show barycenter plots for the two histograms.

Note that although the main application of level set trees is for the cases where dimension is higher than 2, we may also apply level set trees to highlight the number of local maxima of the estimate, which is often not easy to see from a perspective plot of a histogram. The histogram of Figure 5 b) has 7 local maxima, which can be seen from the volume plot and from the barycenter plot.

It is feasible to construct level set trees from histograms with the algorithm of Section 2.4. The number of bins gives an upper bound for the number of "atoms" of the level sets. The normal reference rule for the number of bins suggests that the number of bins should be chosen as $N \asymp n^{d/(2+d)}$, see Scott (1992), page 82.

3.2 Kernel estimator

Define the kernel estimator based on data $X_1, \ldots, X_n \in \mathbf{R}^d$ by

$$f_n(x) = (nh^d)^{-1} \sum_{i=1}^{n} K((x - X_i)/h), \qquad x \in \mathbf{R}^d,$$

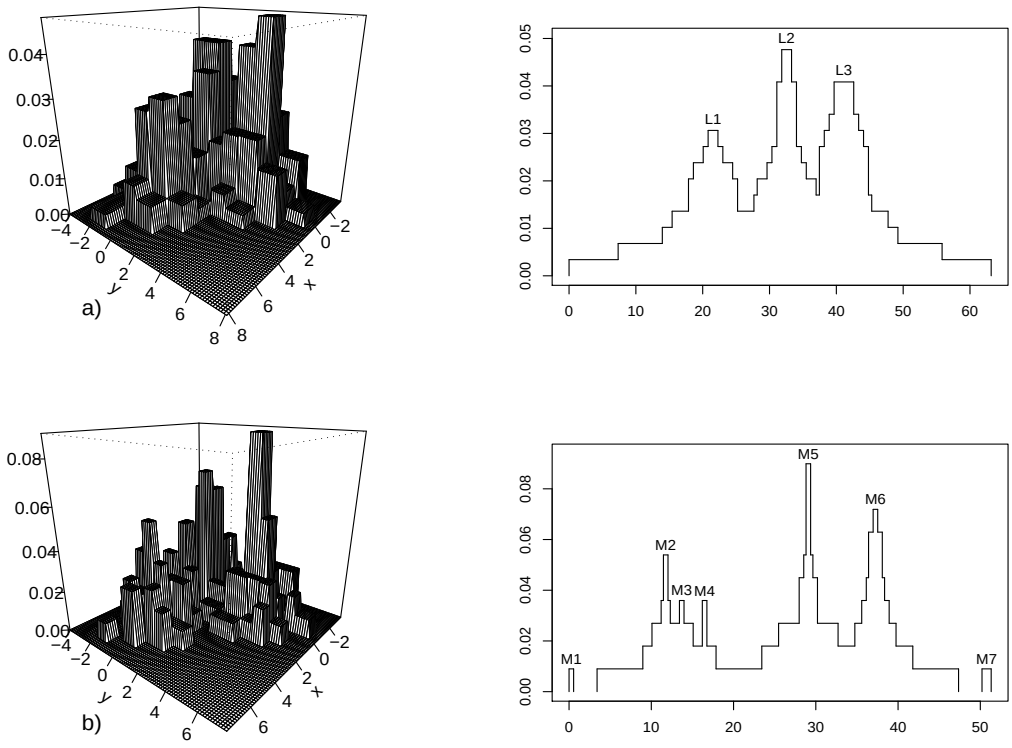

Figure 5: Histogram estimates with samplesize 200; a) 8^2 bins; b) 13^2 bins. On the left we have perspective plots and on the right volume plots.

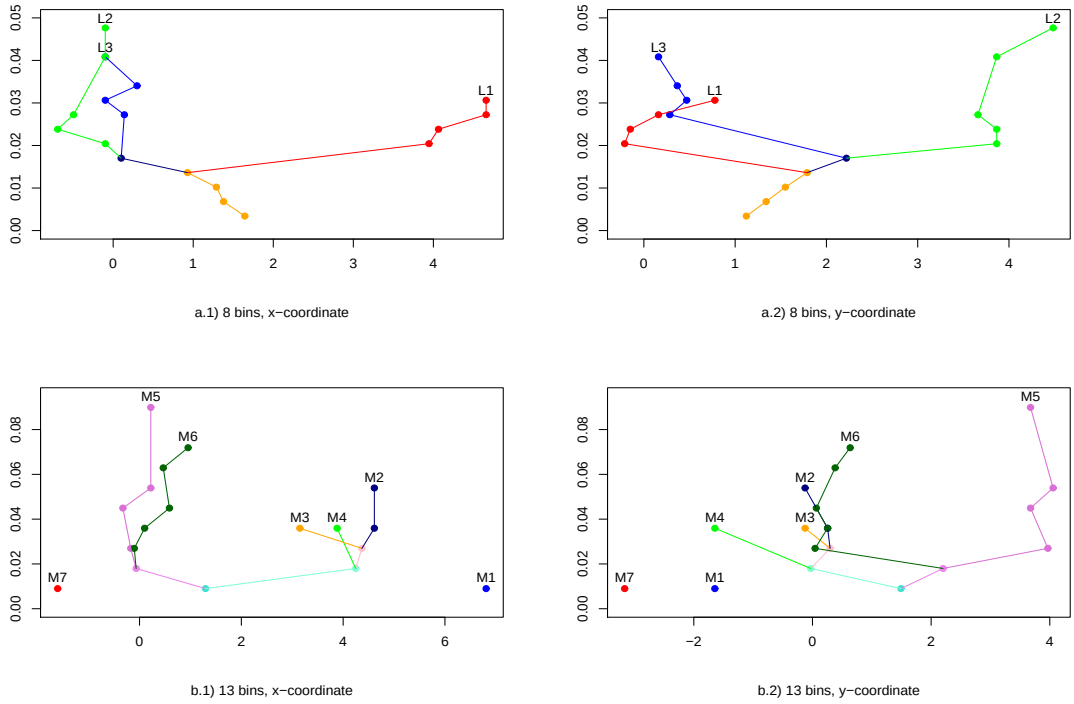

Figure 6: Barycenter plots from the histograms in Figure 5

where $h > 0$ is the smoothing parameter and $K : \mathbf{R}^d \to \mathbf{R}$ is the kernel function. In order to apply level set trees we evaluate the kernel estimate at the knots of a grid and then consider the function which is constant over the rectangles whose centers are the knots. We also quantize the kernel estimate.

The right number of quantization levels depends on the smoothness of the estimate. We may choose the number of quantization levels by increasing the number of levels until the number of local extremes of the quantized estimate is not increasing. Note that the problem of the choice of the quantization levels is related to the problem of choosing the levels for a contour plot of a two dimensional kernel estimate.

In Figure 7 we visualize kernel estimates. The estimates are based on the same sample of 200 which was considered in Section 3.1. On the left hand side we show perspective plots of the kernel estimates and on the right hand side we show the corresponding volume plots, based on 80 levels and 64^2 gridpoints. Bartlett-Epanechnikov product kernel is used in the estimates. Figure 8 shows barycenter plots for the three estimates.

The volume plot in Figure 7 c) does not reveal all 27 local extremes, which are visible in the barycenter plot of Figure 8 c). On the other hand, from the volume plot one sees the importance of the modes in terms of the excess mass.

The algorithm described in Section 2.4 is not feasible for kernel estimates since

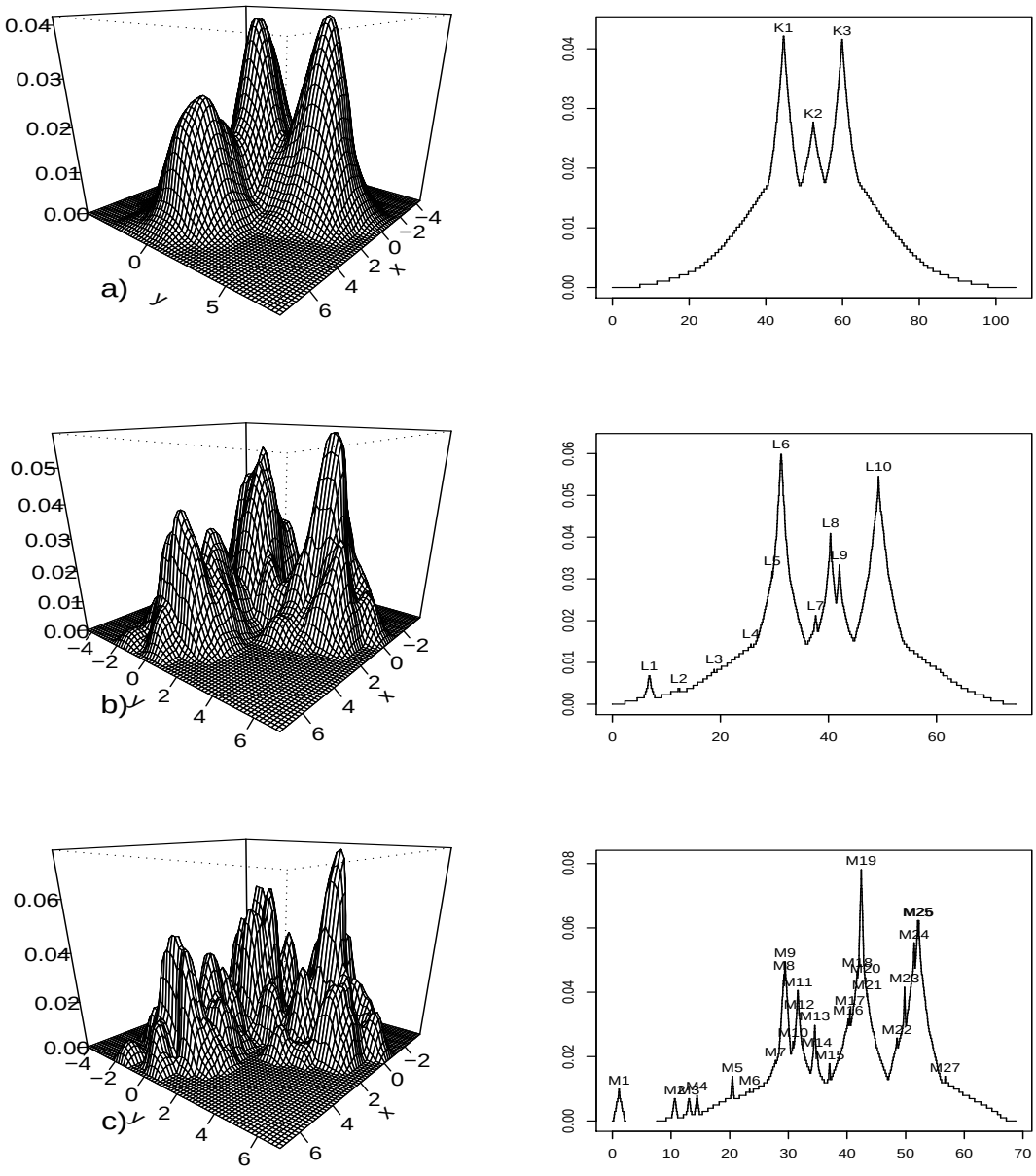

Figure 7: Kernel estimates with smoothing parameters a) $h = 1.4$, b) 0.8, and c) 0.6. On the left perspective plots and on the right volume plots. We have 80 levels and 64 gridpoints for both directions.

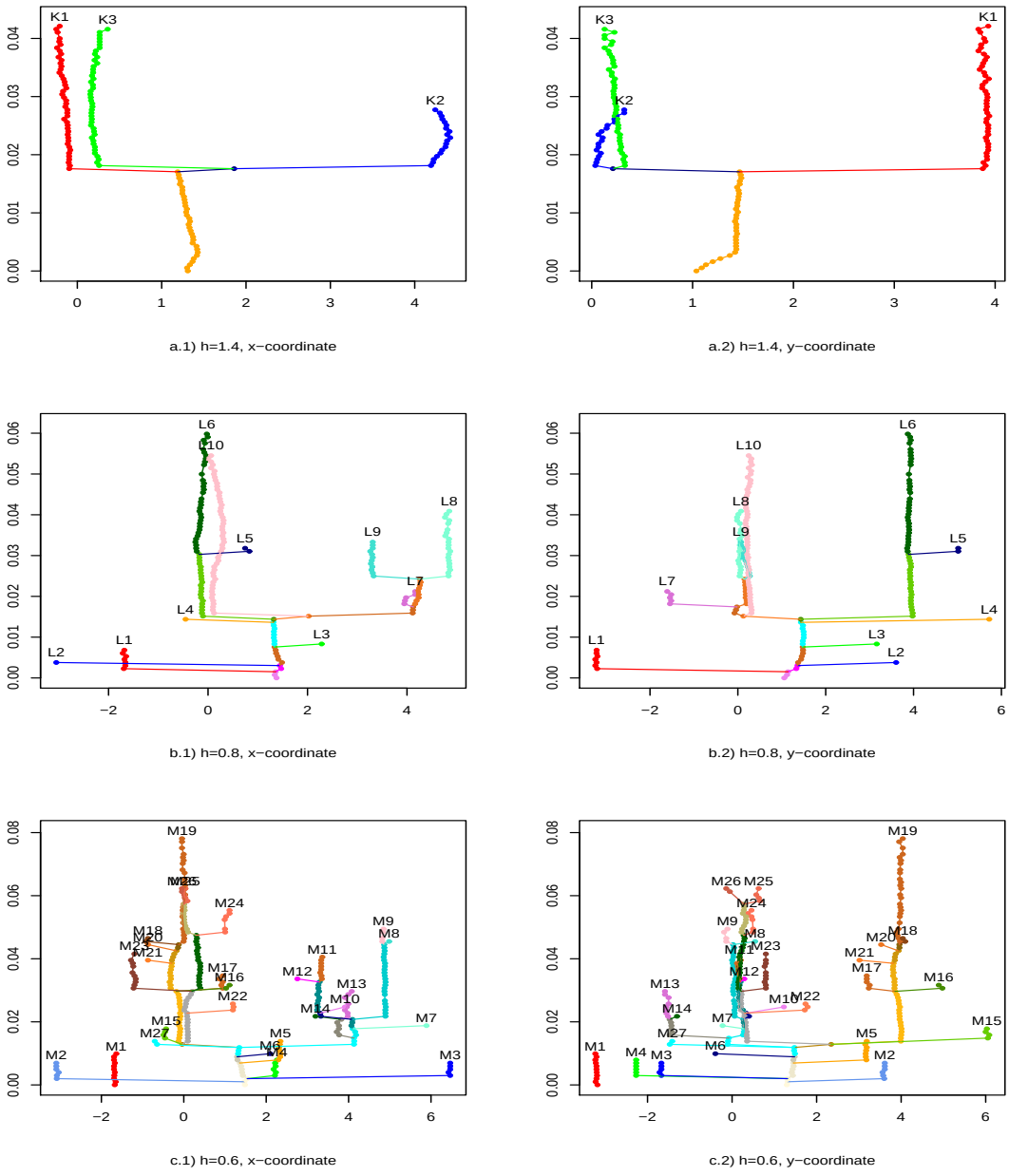

Figure 8: Barycenter plots corresponding volume plots of Figure 7.

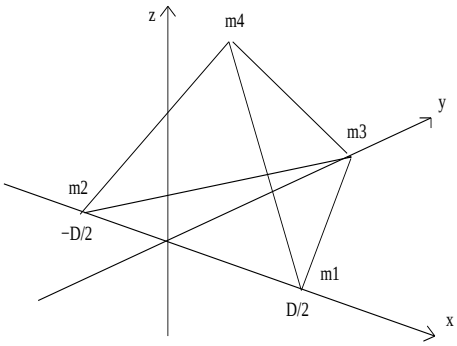

Figure 9: Tetrahedron with vertices m_1, m_2, m_3, and m_4. Distance between vertices is D.

the number of gridpoints needed to approximate accurately a kernel estimate is huge in multivariate cases. Klemelä (2005) describes a dynamic programming algorithm for finding the separated components of level sets of a kernel estimate.

4 Examples

We will give examples from estimation and visualization of mixtures of Gaussian densities. We consider 3 and 4 dimensional mixtures whose components are close to each other.

4.1 Three dimensional example

We consider an equal mixture of standard Gaussian densities. The means of the components of the mixture lie on the vertices of a tetrahedron. We choose as means

$$\begin{aligned} m_1 &= D \times (1/2, 0, 0), \\ m_2 &= D \times (-1/2, 0, 0), \\ m_3 &= D \times (0, \sqrt{3}/2, 0) \\ m_4 &= D \times (0, 1/(2\sqrt{3}), \sqrt{2/3}). \end{aligned}$$

Points m_i lie on the vertices of a tetrahedron and the distance between vertices is D. See Figure 9.

When we choose D sufficiently small then two dimensional projections do not reveal the modes of the density. In Figure 10 a)-c) we show marginal densities on the three coordinate planes of $\mathbf{R}^3$ when $D = 3$. One can see that there are no better projections than the coordinate planes. When the distance between the components of the mixture is large, then the projection to x-y-plane reveals all 4 modes: in Figure 10 d) we show the marginal density on x-y-plane when $D = 4$.

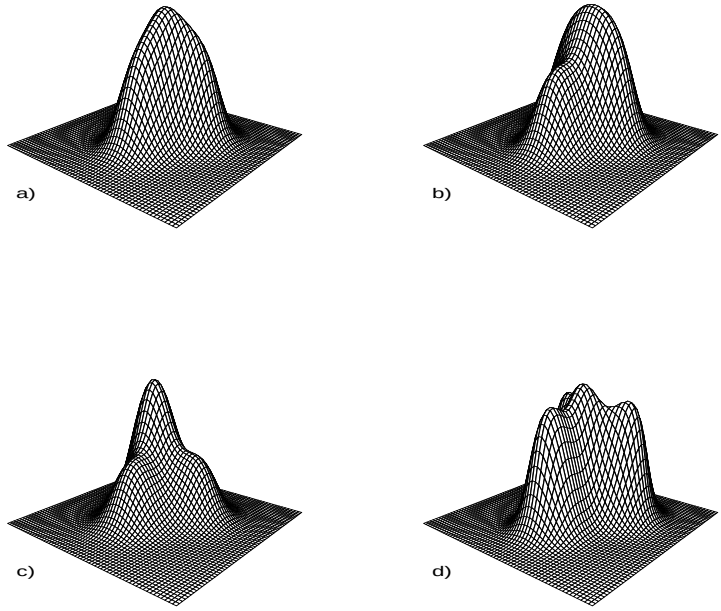

Figure 10: Marginal densities of a mixture of 4 standard Gaussian densities in $\mathbf{R}^3$; a) x-y plane, D=3; b) x-z plane, D=3; c) y-z plane, D=3; d) x-y plane, D=4.

For the simulation example we choose distance between modes $D = 3$. When the sample size is sufficiently large, then 3-dimensional histogram finds all 4 modes of the underlying density, even when marginal densities do not reveal the modes.

In Figure 11 we present a histogram, based on a sample of size 5000. We chose 9 bins for each direction. The mode labelled M1 has the largest excess mass.

The barycenter plots in Figure 11 reveal that the locations of the modes of the estimate are not too far from the locations of the modes of the true density. In fact, the locations of the modes are

$$\begin{aligned} M1 &= (-0.02, 1.1, 2.1) \approx m_4 = (0, 0.9, 2.4), \\ M2 &= (-0.02, 3.2, 0.2) \approx m_3 = (0, 2.6, 0), \\ M3 &= (-1.0, 0.04, 0.2) \approx m_2 = (-1.5, 0, 0), \\ M4 &= (1.0, 0.04, 0.2) \approx m_1 = (1.5, 0, 0). \end{aligned}$$

By M1-M4 we denote the barycenters of the sets where the histogram estimate achieves a local maximum.

4.2 Four dimensional example

We consider an equal mixture of standard Gaussian densities. The means of the components of the mixture lie on the vertices of a pentahedron. We choose as means

$$\begin{aligned} m_1 &= D \times (1/2, 0, 0, 0), \\ m_2 &= D \times (-1/2, 0, 0, 0), \\ m_3 &= D \times (0, \sqrt{3}/2, 0, 0), \\ m_4 &= D \times (0, 1/(2\sqrt{3}), \sqrt{2/3}, 0), \\ m_5 &= D \times (0, 1/(2\sqrt{3}), 1/(2\sqrt{6}), \sqrt{15/24}). \end{aligned}$$

Points m_i lie on the vertices of a pentahedron and the distance between vertices is D. When we choose m_i sufficiently close to each other, then two dimensional projections do not reveal the modes of the density. In Figures 12 a)-f) we show marginal densities on the six coordinate planes of $\mathbf{R}^4$ of the equal mixture of standard Gaussian densities, with $D = 4$. We denote coodinate directions by (x, y, z, u). Even when the distance between components of the mixture is large, there exists no projection to the coordinate planes that would reveal all 5 modes at once.

For the simulation example we choose distance between modes $D = 4$ and sample size 2000. We present in Figure 13 a kernel estimate with $h = 1$ and Bartlett-Epanechnikov product kernel. We quantized to 40 levels and evaluated

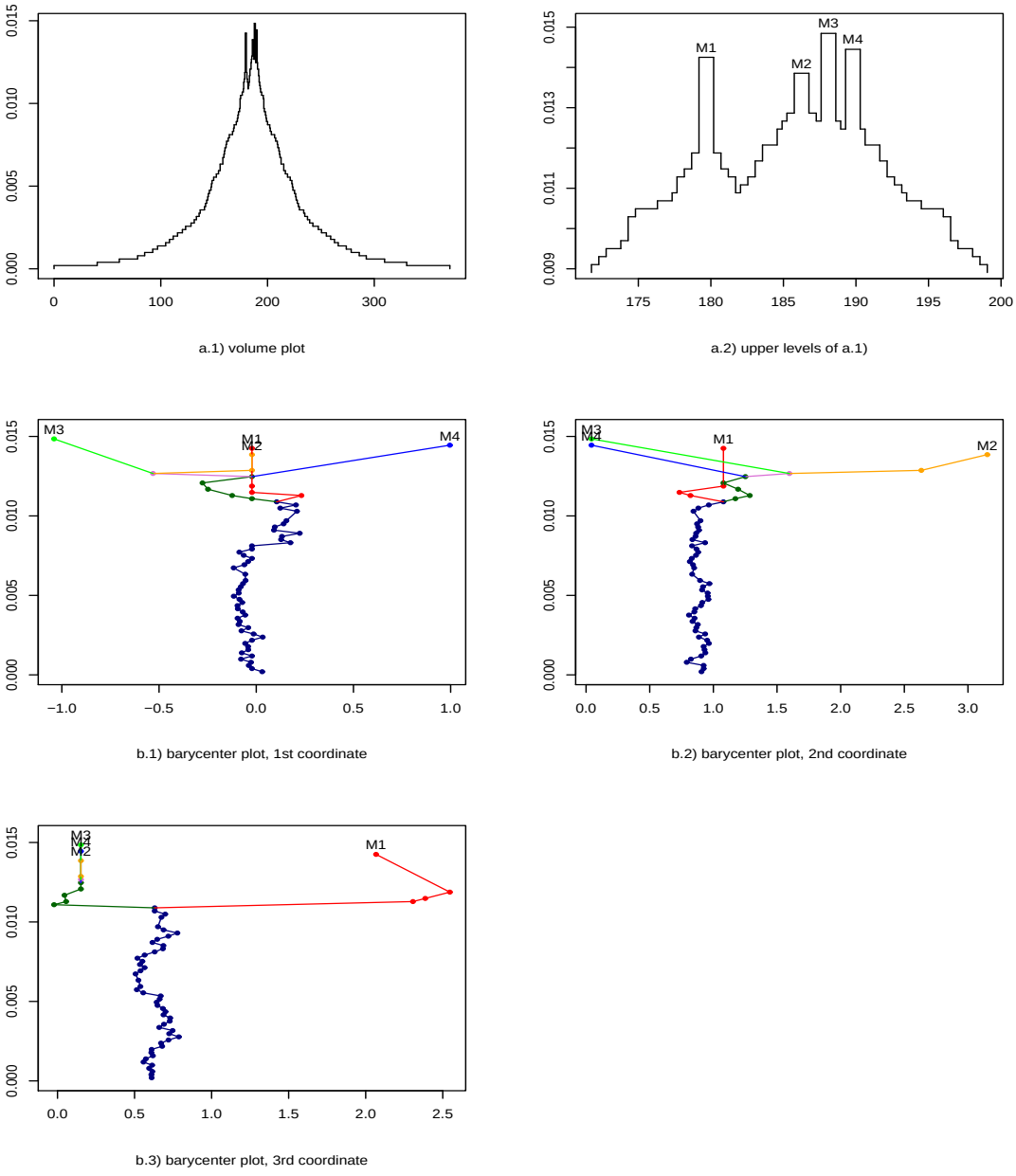

Figure 11: Histogram with 9^3 bins based on a sample of 5000 from a mixture of 4 Gaussian densities in $\mathbf{R}^3$. Window a.1) shows volume plot, a.2) shows zoomed volume plot, and b.1-3) show barycenter plots.

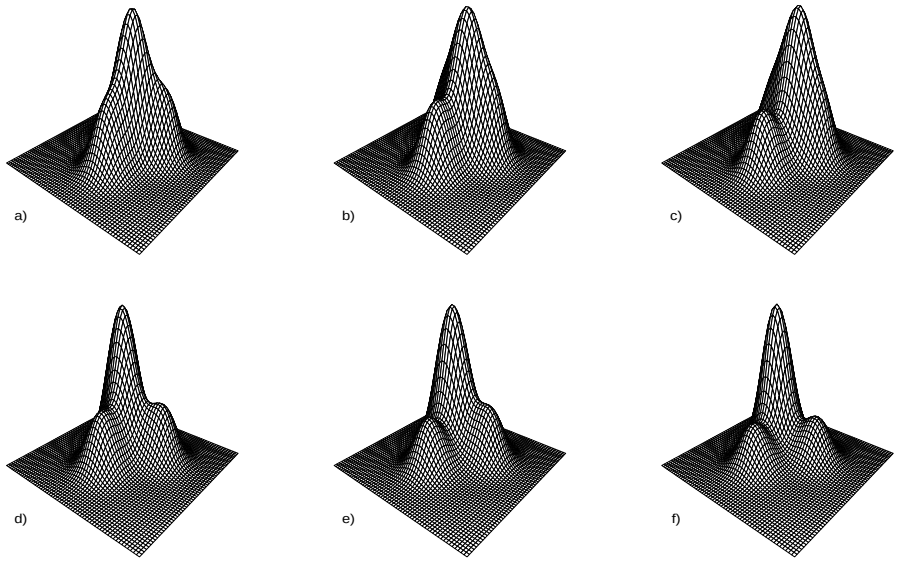

Figure 12: Marginal densities of a mixture of 5 standard Gaussian densities in $\mathbf{R}^4$, when $D = 4$; a) x-y plane; b) x-z plane; c) x-u plane; d) y-z plane; e) y-u plane; f) z-u plane.

the estimate at 16^4 gridpoints. With only 16 gridpoints in each direction the kernel estimator is still more accurate than the histogram estimator.

The locations of the modes of the estimate in Figure 13 are

$$\begin{aligned} M1 &= (0.2, 0.7, 0.9, 3.1) \approx m_5 = (0, 1.2, 0.8, 3.2), \\ M2 &= (0.2, 1.4, 3.2, -0.3) \approx m_4 = (0, 1.2, 3.3, 0), \\ M3 &= (2.2, 0.4, 0.2, -0.3) \approx m_1 = (2, 0, 0, 0), \\ M4 &= (0.2, 3.4, -0.6, -0.3) \approx m_3 = (0, 3.5, 0, 0), \\ M5 &= (-1.8, 0.02, 0.2, 0.4) \approx m_2 = (-2, 0, 0, 0). \end{aligned}$$

One can see that the locations of the modes of the estimate are not too far from the locations of the modes of the true density.

5 Summary and further work

5.1 Summary

Studying level sets for a series of levels provides information on the shape of a multivariate function. Level sets has been applied in density estimation and mode detection in 3 and 4 dimensional cases by Scott (1992) and Härdle and Scott (1992), who present a sliding technique for visualizing 4 dimensional functions.

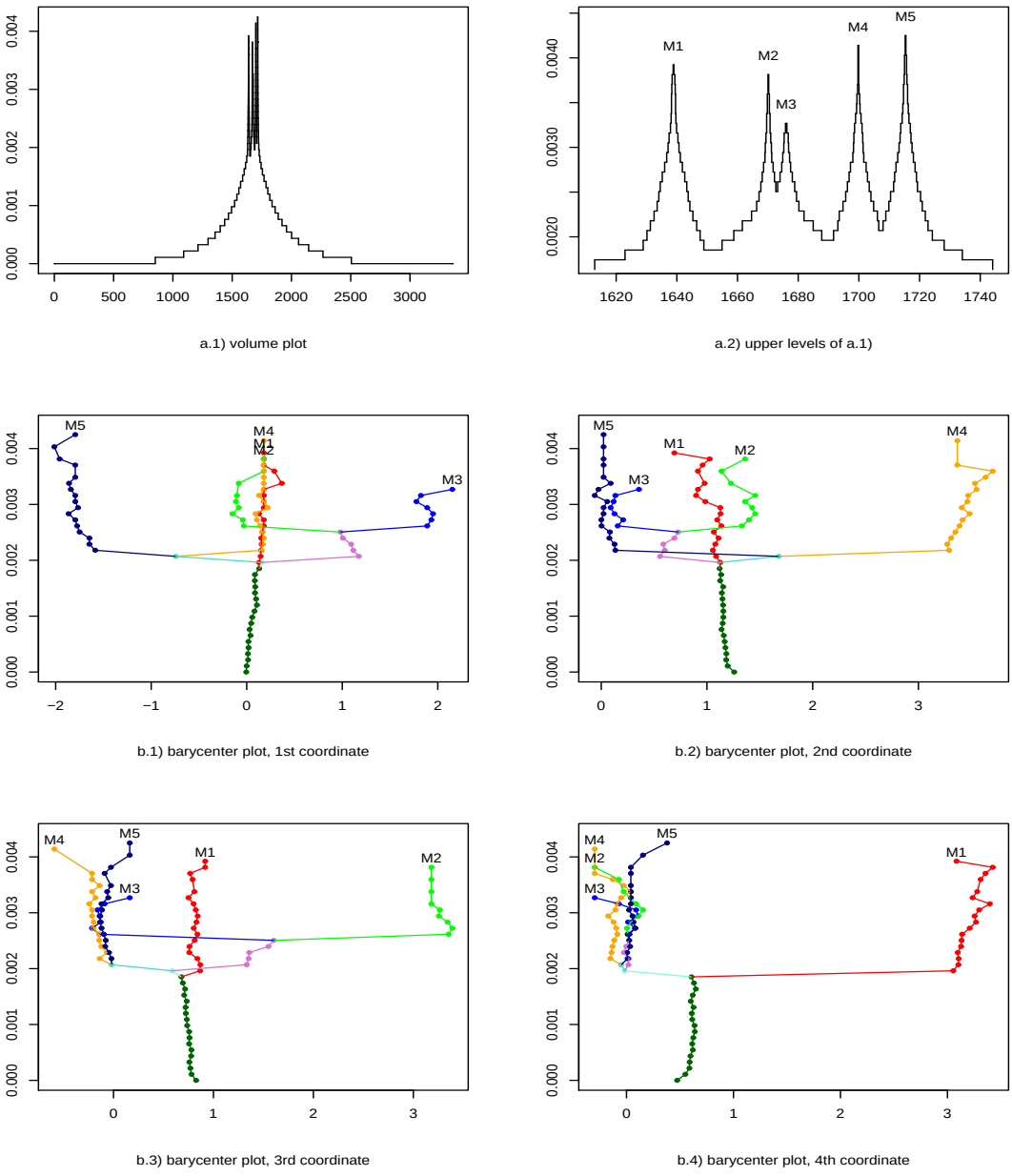

Figure 13: Kernel estimate with 2000 observations from a mixture of 5 Gaussian densities in $\mathbf{R}^4$. Window a.1) shows the volume plot, a.2) zooms to the upper levels of the estimate, and b.1-4) show barycenter plots.

They visualize 3D density contours as the fourth variable is changed over its range.

Our aim is to apply level sets in arbitrary dimension with the help of level set trees. Level set trees are complex objects which give an alternative way to represent functions. This representation is however fruitful for visualizing many features of functions. We are not able to make a single visualization which would without loss of information visualize a multivariate function. Instead, we make a number of visualizations and each plot visualizes one aspect or feature of the original function. Our approach is an alternative to the approach of projecting the data to a low dimensional space. Studying projections and marginal densities is in many cases sufficient but projections may hide some high dimensional features.

In density estimation we are interested how the probability mass is distributed over the d-dimensional Euclidean space. Local extremes of the density function express concentration of the probability mass. We are interested both in visualizing the locations of these local extremes, and also in visualizing the size of the probability mass associated with each local extreme. To achieve these aims we have introduced the volume plot, which visualizes the amount of probability mass associated with each local extreme, and the barycenter plot, which visualizes the locations of these probability masses. The volume plot defines a one dimensional transformation for multivariate densities, which is not any slice, conditional density, or marginal density of the original density.

5.2 Further work

In one and two dimensional cases the art of smoothing consists often in the inspection of the changes of the estimate as the smoothing parameter of the estimate changes. Formal tools to help this process has been given for example by Minnotte and Scott (1993), Minnotte (1997), Chaudhuri and Marron (1999). The graphical representations of level set trees may be utilized to extend versions of these tools to higher than 2 dimensional cases.

We have applied only volumes and barycenters in this article. Other potentially useful characteristics of sets include diameter, perimeter, radius (minimum, maximum, average), and compactness (ratio of the perimeter to the volume).

It is also interesting to apply level sets to visualize other multivariate functions than density functions. For example, it is important to detect modes from regression functions. Barycenter plots may be applied to detect monotone behaviour of a function with respect to some variables, which is often of interest in regression function estimation.

Cluster analysis. According to Hartigan (1975), page 205, clusters are regions of high density separated from other such regions by regions of low density. Thus clusters are separated subsets of some level set $\{x \in \mathbf{R}^d : f(x) \geq \alpha\}$ of the underlying density function f.

Typical clustering algorithms include algorithms based on hierarchical trees and the k-means algorithm. Note that the output of these algorithms is a partition of the sample: they will not provide clusters defined as subsets of the sample space. For some applications this is not sufficient and one needs additional analysis to identify the locations of the clusters of sample points. Another problem with the k-means algorithm and hierarchical trees is that they require prior knowledge of the number of clusters.

We may try to estimate clusters by constructing a nonparametric density estimate and finding the separated subsets of a level set of the estimate corresponding to some low level. Graphical representations of level set trees provide tools for realizing such an approach to clustering. An advantage of cluster analysis based on density estimation is that with this approach we may acquire lots of useful information on the shape of the density function, and not just the clusters.

Acknowledgements

I wish to thank Lasse Holmström, 3 referees, Associate Editor, and Günther Sawitzki for the comments that helped to improve the article, and Steve Marron for his suggestions that helped to construct the simulation examples. Writing of this article was financed by Rolf Nevanlinna Institute, Jenni and Antti Wihuri Foundation, and Deutsche Forschungsgemeinschaft under project MA1026/6-2. The R programming environement was used to prepare this article.

References

Chaudhuri, P. and Marron, J. S. (1999), 'Sizer for exploration of structures in curves', *J. Amer. Statist. Assoc.* **94**, 807–823.

Davies, P. L. and Kovac, A. (2001), 'Local extremes, runs, strings and multiresolution', *Ann. Statist.* **29**, 1–65.

Fisher, N. I., Mammen, E. and Marron, J. S. (1994), 'Testing for multimodality', *Comput. Statist. Data Anal.* **18**, 499–512.

Godtliebsen, F., Marron, J. S. and Chaudhuri, P. (2002), 'Significance in scale space for bivariate density estimation', *J. Comput. Graph. Statist.* **11**, 1–22.

Härdle, W. and Scott, D. (1992), 'Smoothing by weighted averaging of rounded points', *Comput. Statist.* **7**, 97–128.

Hartigan, J. A. (1975), *Clustering Algorithms*, Wiley, New York.

Hartigan, J. A. (1987), 'Estimation of a convex density cluster in two dimensions', *J. Amer. Statist. Assoc.* **82**, 267–270.

Hartigan, J. A. and Hartigan, P. (1985), 'The DIP test of unimodality', *Ann. Statist.* **13**, 70–84.

Hartigan, J. A. and Mohanty, S. (1992), 'The RUNT test for multimodality', *J. Classification* **9**, 63–70.

Klemelä, J. (2005), 'Algorithms for the manipulation of level sets of nonparametric density estimates', *Comput. Statist.* **20**, 349–368.

Mammen, E., Marron, J. S. and Fisher, N. I. (1992), 'Some asymptotics for multimodality tests based on kernel density estimates', *Probab. Theory Relat. Fields* **91**, 115–132.

Marchette, D. J. and Wegman, E. J. (1997), 'The iterated mode tree', *J. Comput. Graph. Statist.* **6**, 143–159.

Minnotte, M. C. (1997), 'Nonparametric testing of the existence of modes', *Ann. Statist.* **25**, 1646–1660.

Minnotte, M. C., Marchette, D. J. and Wegman, E. J. (1998), 'The bumpy road to the mode forest', *J. Comput. Graph. Statist.* **7**, 239–251.

Minnotte, M. C. and Scott, D. W. (1993), 'The mode tree: a tool for visualization of nonparametric density features', *J. Comput. Graph. Statist.* **2**, 51–68.

Müller, D. W. and Sawitzki, G. (1991), 'Excess mass estimates and tests of multimodality', *J. Amer. Statist. Assoc.* **86**, 738–746.

Scott, D. W. (1992), *Multivariate Density Estimation*, Wiley.

Silverman, B. W. (1981), 'Using kernel density estimates to investigate multimodality', *J. Roy. Statist. Soc. Ser. A* **43**, 97–99.