Analysis of the shapes of unimodal densities with nonparametric density estimation

Jussi Klemelä*

Abstract

Nonparametric density estimation has been applied successfully in mode detection. However, nonparametric analysis of the shapes of unimodal densities has attracted less interest. Level set tree based techniques can be applied to analyze the shapes of unimodal densities. A level set tree of a function is a tree structure of the separated components of level sets of the function. Level set trees can be used to describe not only the shapes of functions but also the shapes of multidimensional sets; we can define a distance function or a height function on a set and construct a level set tree of this function. Finally, level set trees can be used to describe the shapes of point clouds, by applying appropriate smoothing. This leads to a computationally efficient way of describing the shapes of unimodal densities.

Key Words: multivariate data, nonparametric density estimation, level set tree, shape analysis, visualization of multivariate functions

1. Introduction

We are interested in the visualization of the dependence of random variables $X_1, \dots, X_d \in \mathbf{R}$. We shall assume that the density function $f : \mathbf{R}^d \to \mathbf{R}$ of random vector $(X_1, \dots, X_d)$ is unimodal, and we shall visualize the dependency by visualizing the shapes of the level sets of the density f. Usually density function f is unknown and has to be estimated using a sample. We shall not discuss methods for density estimation but concentrate on the problem of visualization. On the other hand, we shall discuss methods to directly visualize dependency using a data set $X^1, \dots, X^n \in \mathbf{R}^d$ sampled from f, bypassing the problem of density estimation.

We are particularly interested in the cases when $d > 3$. In order to visualize multivariate objects we have to transform these objects to 2- or 3-dimensional objects, since humans cannot see higher than 3-dimensional objects. Projections and slices are often applied to derive lower dimensional objects from high dimensional objects. We consider an other possibility: the use of *shape isomorphic transforms*. This approach uses the fact that it is possible to visualize a multidimensional object with a low dimensional object if these objects have the same shape.

A *volume transform* is the basic shape isomorphic transform which we shall apply. A volume transform is a transform which is used to visualize local extremes of a function. A volume transformed function has the same shape as the original function in the sense that the number and the sizes of the local extremes (either minima or maxima) are equal.

A volume transform is used to visualize multimodality of functions, but we are interested in the visualization of the dependency structure of unimodal densities. A volume transform can be applied to this purpose when we define functions on the level sets of the density, and then use the volume transform to visualize the shapes of the functions defined on the level sets. The (upper) level set of f with level $\lambda \in \mathbf{R}$ is defined by

$$\Lambda(f, \lambda) = \{x \in \mathbf{R}^d : f(x) \geq \lambda\}. \tag{1}$$

Section 2 motivates the approach of dependency visualization via visualization of level sets. Section 3 contains the basic definitions for the visualization of multimodal functions: the definitions of a level set tree and a volume transform. Section 4 applies the concepts introduced in Section 3 to the visualization of sets. Section 5 applies the concepts introduced in Section 3 to the direct visualization of data.

2. Dependency and level sets

Distribution function $F : \mathbf{R}^d \to \mathbf{R}$ may be decomposed into a part which describes the dependency and into a part which describes the marginal distributions. We call a *copula* the part which describes the dependency. Copulas are multivariate distribution functions whose one-dimensional marginal distributions are uniform on the interval $[0, 1]$. The basic idea is that any distribution function $F : \mathbf{R}^d \to \mathbf{R}$ of a random vector $(X_1, \dots, X_d)$ may be written as $F(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d))$, where C is the copula and F_i, $i = 1 \dots, d$, are the marginal distribution functions: $F_i(x_i) = P(X_i \leq x_i)$. Let X_1 and X_2 be random variables with distribution functions F_1 and F_2. We

*University of Oulu, Department of Mathematical Sciences, P.O. Box 3000, FIN-90014, University of Oulu, Finland

have

$$\begin{aligned} F(x_1, x_2) &= P(X_1 \le x_1, X_2 \le x_2) \\ &= P(F_1(X_1) \le F_1(x_1), F_2(X_2) \le F_2(x_2)) \\ &= C(F_1(x_1), F_2(x_2)), \end{aligned} \tag{2}$$

where

$$C(u, v) = P(F_1(X_1) \le u, F_2(X_2) \le v), \qquad u, v \in [0, 1], \tag{3}$$

that is, copula C is the joint distribution function of uniformly distributed random vectors $F_1(X_1)$, $F_2(X_2)$. An exposition of copulas is given by Nelsen (1999).

We shall call a *standard copula* such a copula whose marginals are uniform distributions on $[0, 1]$. We can choose the marginal distributions of a copula to be some other continuous distribution than the uniform distribution on $[0, 1]$. It turns out that we get simpler copulas by choosing the marginal distributions of a copula to be the standard Gaussian distribution. The densities of standard copulas have typically several local maxima and minima, whereas copula densities with the standard Gaussian distribution as marginals are typically unimodal.

Similarly as in (2) we can write F as

$$F(x_1, x_2) = C\left(\Phi^{-1}(F_1(x_1)), \Phi^{-1}(F_2(x_2))\right),$$

where $\Phi : \mathbf{R} \to \mathbf{R}$ is the distribution function of the standard Gaussian distribution and

$$C(u, v) = P\left(\Phi^{-1}(F_1(X_1)) \le u, \Phi^{-1}(F_2(X_2)) \le v\right), \qquad u, v \in \mathbf{R}. \tag{4}$$

Now C is a distribution function whose marginals are standard Gaussian. Function C defined by (4) is a nonstandard copula with standard Gaussian marginals. When F_1 and F_2 are continuous, then

$$C(u, v) = F\left(F_1^{-1}(\Phi(u)), F_2^{-1}(\Phi(v))\right), \qquad u, v \in \mathbf{R}, \tag{5}$$

is a nonstandard copula with standard Gaussian marginals. When copula C is defined by (5), then the copula density is

$$c(u, v) = f\left(F_1^{-1}(\Phi(u)), F_2^{-1}(\Phi(v))\right) \frac{\phi(u)\phi(v)}{f_1(F_1^{-1}(\Phi(u))) \cdot f_2(F_2^{-1}(\Phi(v)))},$$

where f is the density of F, f_1, f_2 are the densities of F_1, F_2, and ϕ is the density of the standard Gaussian distribution.

Figure 1(a) shows a scatter plot of exchange rates of Brazilian Real and Mexican Peso between 1995-01-05 and 2007-09-26. The rates are with respect to one U.S. Dollar and transformed to returns ($r_i \mapsto (r_i - r_{i-1})/r_{i-i}$). There are 3197 observations. The data is provided by Federal Reserve Economic Data (http://research.stlouisfed.org). Frame b) shows copula transformed data where the marginals are approximately uniform, and frame c) shows copula transformed data where the marginals are approximately standard Gaussian. Frame b) shows that the standard copula transform leads to data whose distribution is multimodal: the density has a local maxima at the lower left corner and at the upper right corner, and an additional local maxima at the center of the plot. Also, the density has a local minima at the left upper corner and at the right lower corner. Frame c) shows that the copula transform with the standard Gaussian marginals leads to a unimodal density.

The situation in Figure 1 is typical: the standard copula leads to densities with many local extremes but the copula with Gaussian marginals leads to unimodal densities. Using the copulas with the standard Gaussian marginals makes it possible to visualize dependence with the level set tree based methods: unimodal densities have connected and often even star shaped level sets, and we can visualize these level sets by defining functions on the level sets.

3. Shapes of functions

We shall define a level set tree in Section 3.1 and a volume function in Section 3.2. These concepts are used to visualize multimodal functions but we shall apply these concepts later to visualize level sets of a unimodal density.

3.1 Level set tree

We use level sets to define a transform of a multivariate function to a univariate function. This *volume transform* can be used to visualize the local maxima of a function: we visualize the number, size, and the tree structure of

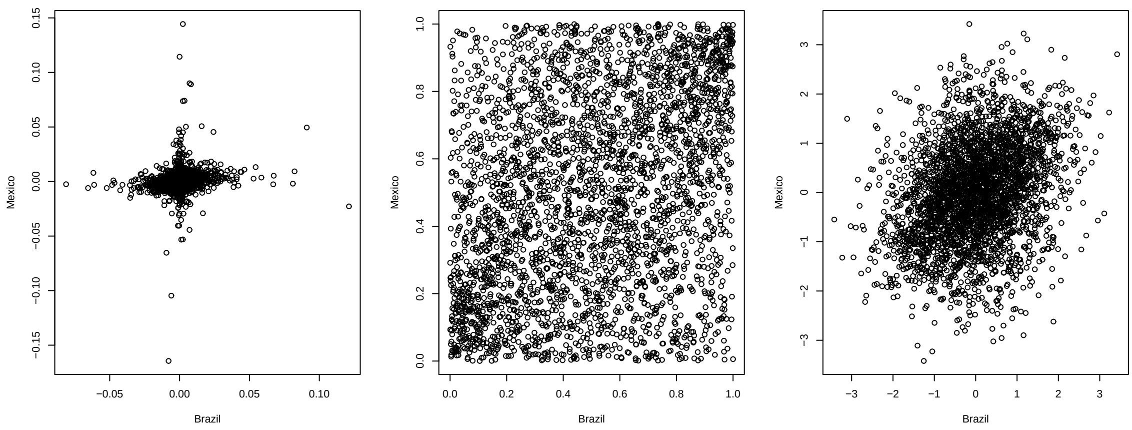

Figure 1: The figure illustrates data preprocessing with data of exchange rates of Brazilian Real and Mexican Peso (n=3197). Frame a) shows a scatter plot of the data, frame b) shows the copula transformed data with uniform marginals, and frame c) shows the copula transformed data with Gaussian marginals.

the local maxima. We may apply lower level sets to define an analogous transform to visualize the local minima of a function. Combined together, these transforms give a comprehensive visualization of the local extremes of a multivariate function. The concept of a level set tree is the basic concept underlying the definition of a volume function. A level set tree and a volume function were defined in Klemelä (2004) for piecewise constant functions. Here we use more general definitions.

Definition 1 (Level set tree.) *A level set tree of function* $f : \mathbf{R}^d \to \mathbf{R}$*, associated with set of levels* $\mathcal{L} = \{\lambda_1 < \cdots < \lambda_L\}$*, where* $\lambda_L \le \sup_{x\in\mathbf{R}^d} f(x)$*, is a tree whose nodes are associated with subsets of* $\mathbf{R}^d$ *and levels in* $\mathcal{L}$ *in the following way.*

1. *Write*
$$\Lambda(f, \lambda_1) = A_1 \cup \cdots \cup A_K,$$
where sets A_i *are pairwise separated, and each is connected. The level set tree has* K *root nodes which are associated with sets* A_i*,* $i = 1, \ldots, K$*, and each root node is associated with the same level* λ_1*.*
2. *Let node* m *be associated with set* $B \subset \mathbf{R}^d$ *and level* $\lambda_l \in \mathcal{L}$*,* $1 \le l < L$*.*
   (a) *If* $B \cap \Lambda(f, \lambda_{l+1}) = \emptyset$*, then node* m *is a leaf node.*
   (b) *Otherwise, write*
   $$B \cap \Lambda(f, \lambda_{l+1}) = C_1 \cup \cdots \cup C_M,$$
   where sets C_i *are pairwise separated, and each is connected. Then node* m *has* M *children, which are associated with sets* C_i*,* $i = 1, \ldots, M$*, and each child is associated with the same level* λ_{l+1}*.*

Above we say that sets $B, C \subset \mathbf{R}^d$ are *separated* if $\inf\{\|x - y\| : x \in B,\ y \in C\} > 0$. and 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, two sets are separated if there is some space between them and a set is connected if it cannot be written as a union of two separated sets.

3.2 Volume transform

Now we are ready to define a volume function. A volume transform is defined as the mapping which maps a function to its volume function.

Definition 2 (Volume function.) *Let* $f : \mathbf{R}^d \to \mathbf{R}$ *be a function, let* μ *be a Borel measure on* $\mathbf{R}^d$*, and let* T *be a level set tree of* f*.*

- *Annotate each node* m *of the level set tree* T *with an interval* $\mathcal{I}_m \subset \mathbf{R}$*. Let the length of an interval be equal to the* μ*-volume of the set annotated with the node. Let the intervals be nested according to the tree structure of the level set tree. We comment later on the exact definition of the intervals.*

- *volume function* $v(f;T) : \mathbf{R} \to \mathbf{R}$ *is such that for each level* $\lambda \in \mathbf{R}$*,*

$$\{x \in \mathbf{R} : v(f)(x) \geq \lambda\} = \bigcup \{\mathcal{I}_m : m \text{ is such node of } T \text{ that } \lambda_m \geq \lambda\},$$

where λ_m *is the level and* $\mathcal{I}_m$ *is the interval associated to node* m*.*

Definition 2 does not specify the locations of the intervals associated to the nodes of a level set tree. We could use rather arbitrary rules, but the following rule is quite natural. Choose first an interval $[0, L]$, where L is greater than the sum of the volumes of the sets associated to the root nodes. Then the intervals associated to the root nodes are positioned inside $[0, L]$ in a symmetric way. After that, one positions the intervals recursively, making a nested collection of intervals according to the tree structure, and positioning the intervals symmetrically. Note that we have note excluded the case where some of the level sets of the function have infinite volume. Note also that we have not defined a level set tree as an ordered tree, so that the positioning of the sibling intervals may be done in an arbitrary order.

4. Shapes of sets

When function $f : \mathbf{R}^d \to \mathbf{R}$ is unimodal, then its volume function does not contain much interesting information about the function. However, we can analyze the shapes of the level sets of f using similar tools as the volume function. We shall discuss the analysis of the shape of a set $A \subset \mathbf{R}^d$, where A could be a level set of a density. We define in Section 4.1 a shape tree and in Section 4.2 we shall define 3 ways to define a shape function based on a shape tree.

4.1 Shape tree

We can analyze and visualize shapes of sets by first defining functions on the sets, and secondly using the level set tree based methods of Section 3 for analyzing and visualizing these functions. The two basic ways to define functions on a set use either a *distance function* or a *height function*. Let $A \subset \mathbf{R}^d$.

1. A *distance function* $f_A : A \to \mathbf{R}$ with reference point $\mu \in \mathbf{R}^d$ is defined by

$$f_A(x) = \|\mu - x\| I_A(x). \tag{6}$$

2. A *height function* $f_A : A \to \mathbf{R}$ is the orthogonal projection of A onto a fixed line on $\mathbf{R}^d$, when the line is identified with $\mathbf{R}$.

A distance function seems more natural to be used in the analysis of the shapes of the level sets of a unimodal density. In many cases these level sets are star shaped, and there is a natural center point in the set (center of mass, for example). Then the distance function characterizes how the set is evolving in different directions around the center point. A shape tree was defined in Klemelä (2006).

Definition 3 (Shape tree.) A shape tree *of a set* $A \subset \mathbf{R}^d$*, associated with reference point* $\mu \in A$*, and set of radii* $\mathcal{R} = \{0 = r_0 < r_1 < \cdots < r_L\}$*, is the level set tree of distance function* f_A *of* A *with the reference point* μ*. The grid of levels of the level set tree is* $\mathcal{R}$*.*

4.2 Shape transforms

A *shape transform* is a transform of a multidimensional set to a univariate function. Shape transforms are defined similarly as a volume transform of a multivariate function in Section 3.2. In fact, we can define a shape transform as a volume transform of a distance function of a set. This particular shape transform shall be called a *radius transform*. However, we can define several different kinds of shape transforms. Other shape transforms include the *tail probability transform*, and the *probability content transform*. A radius transform and a probability content transform were defined in Klemelä (2006).

1. A radius transform is defined for connected sets. To put it shortly, a radius transform is a volume transform of a distance function.

2. A tail probability transform is defined for a connected set, when there is a probability measure defined on the set. For example, when the set is a level set of a density function, then the density function defines a probability measure on the set. A tail probability transform is otherwise similar to a radius transform, but now the length of the nodes is equal to the probability content of the sets, and not to the Lebesgue measure of the sets as in the case of a radius transform.

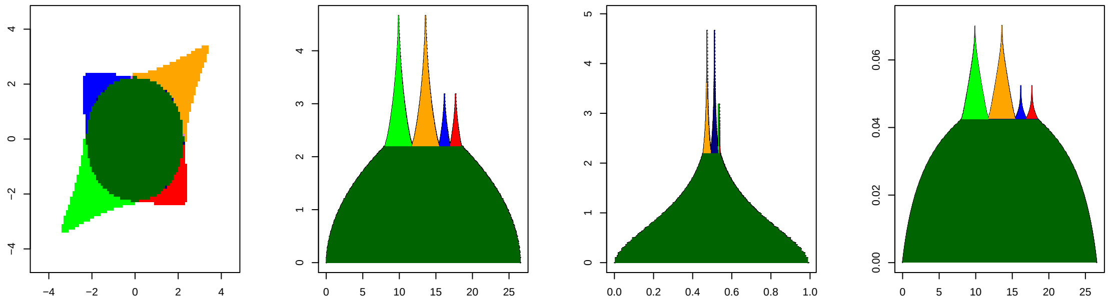

Figure 2: *(Shape functions.)* Frame a) shows the 0.5% level set of a density which has a Student copula with the standard Gaussian marginals. Frame b) shows a corresponding radius plot, frame c) shows a tail probability plot, and frame d) shows a probability content plot.

3. A probability content transform is defined for a connected set, when there is a probability measure defined on the set. A probability content transform is otherwise similar to a radius transform but now the heights of the nodes are determined so that a probability content function visualizes the probability content inside the set. The length of a node is taken to be the volume of the associated set, like in a radius function.

Figure 2 shows shape functions of the 0.5% level set of a density which has the Student copula with the standard Gaussian marginals. The correlation parameter of the copula is $\rho = 0.6$ and the degrees of freedom are $\nu = 2$. Frame a) shows the level set, frame b) shows a radius plot, frame c) shows a tail probability plot, and frame d) shows a probability content plot.

5. Shapes of point clouds

The concept of a *tail tree* replaces the concept of a shape tree for the sets of finite cardinality. With tail trees one may visualize the shape, the location, and the orientation of a multivariate point cloud $x_1, \ldots, x_n \in \mathbf{R}^d$. The point cloud is interpreted as realizations of n random vectors having a common density function. The visualizations with tail trees are tailored to the case where the point cloud does not have clusters (it is connected in the sense of a single linkage hierarchical clustering). The assumption of the connectedness of the point cloud may be interpreted as arising from the connectedness of the support, or a low-level level set, of the density which has generated the observations.

The method is designed for visualizing dependency among several variables, whereas many of the other visualization tools are more directed to finding and visualizing clusters in the data. The basic idea is to define a tail tree among the observations. A tail tree can be visualized with a tail frequency plot, similarly as a shape tree can be visualized with a tail probability function.

A tail tree is like a shape tree, defined in Section 4.1: we shall define a tail tree as a level set tree of a distance function. However, unlike in the case of a shape tree, we restrict ourselves to the case where the set has finite cardinality. Definition 1 of a level set tree was based on the concepts of separated sets and a connected set. Since the set which we consider has finite cardinality, we cannot use the same concepts of separated sets and a connected set. We shall first generalize the concepts of separated sets and a connected set and define a tail tree in Section 5.1. then we define a tail frequency plot in Section 5.2. A tail tree and a tail frequency plot were defined in Klemelä (2007).

5.1 Tail tree

We define ρ-separated sets and a ρ-connected set. The definition may be used for sets of finite cardinality, but it applies for general sets in the multivariate Euclidean space, and not only for sets of finite cardinality.

Definition 4 (ρ-separated sets, ρ-connected set.)

1. *Sets* $A, B \subset \mathbf{R}^d$ *are* separated *for the resolution threshold* $\rho \geq 0$ *(ρ-separated), if for each* $x \in A$ *and* $y \in B$, $\|x - y\| > 2\rho$, *where* $\|\cdot\|$ *is the Euclidean norm.*

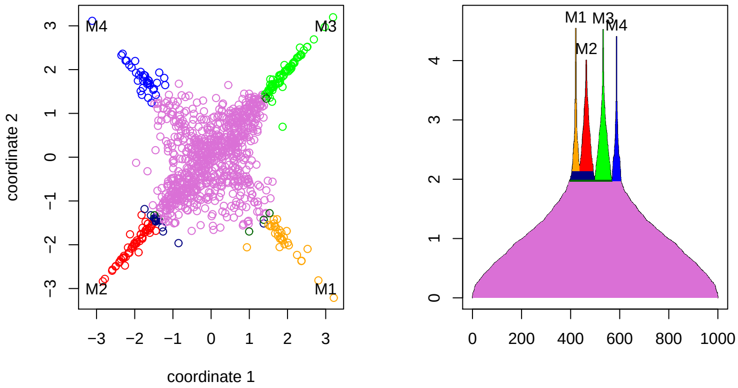

Figure 3: *(Tail frequency function.)* Frame a) shows a scatter plot of data of size $n = 1000$, generated from a distribution with a Student copula and standard Gaussian marginals. Frame b) shows a corresponding tail frequency plot with the resolution threshold 1.

2. *Set $A \subset \mathbf{R}^d$ is* connected *for the resolution threshold $\rho \geq 0$ (ρ-connected), if for every nonempty B, C such that $A = B \cup C$, B and C are not separated for the resolution threshold ρ.*

We shall define a tail tree for a finite set $A = \{x_1, \ldots, x_n\}$. A tail tree shall be defined as a level set tree of a distance function of A. A distance function $f_A : A \to \mathbf{R}$ with reference point $\mu \in \mathbf{R}^d$, of a set $A \subset \mathbf{R}^d$, is defined as $f_A(x) = \|\mu - x\| I_A(x)$. A level set tree depends on a grid of levels. In the case of a tail tree we shall define the grid of levels to be the distances of the points $x_1, \ldots, x_n$ from the reference point $\mu \in \mathbf{R}^d$.

Definition 5 (Tail tree.) *A* tail tree *of set $A = \{x_1, \ldots, x_n\} \subset \mathbf{R}^d$, associated with a resolution threshold $\rho \geq 0$, and with center point $\mu \in \mathbf{R}^d$, is a level set tree of a distance function f_A of A with the reference point μ. The level set tree is defined as in Definition 1, but with the concepts of ρ-separated sets and a ρ-connected set as in Definition 4. The grid of levels of the level set tree is $\mathcal{R} = \{r_1 < \cdots < r_n\}$, where*

$$r_i = \min\{\|x - \mu\| : x \in \mathcal{X}_i\},$$

with $\mathcal{X}_1 = A$ and for $i = 1, \ldots, n-1$,

$$\mathcal{X}_{i+1} = \{x \in A : \|x - \mu\| > r_i\}.$$

5.2 Tail frequency plot

A tail frequency plot visualizes the heaviness of the tails of the underlying distribution. In the multivariate case the tails of the distribution may have anisotropic heaviness: the tails may decrease at different rates in different directions. For elliptical distributions the density has isotropic tails, determined by the 1D generator function, but in the general case the tails are anisotropic.

The nodes of a tail tree are associated with subsets of the data. A tail frequency plot visualizes a tail tree so that each node of the tree is drawn as a line whose length is proportional to the number of observations in the set associated with the node. We identify the lines as level sets of a 1D function, and a tail frequency plot is similar to a plot of a volume function, as defined in Definition 2. A tail frequency plot is a plot of a 1D piecewise constant function, which is defined by associating each node of the tail tree to a separated component of a level set: (1) the length of the separated component of a level set is equal to the number of observations in the node, (2) the height of the separated component of a level set is equal to the distance of the closest observation from the center point, among all observations associated with the node, and (3) the separated components of level sets are nested according to the parent-child relations.

Figure 3 shows a data of size $n = 1000$ generated from a distribution which has the Student copula with the correlation parameter $\rho = 0.6$ and degrees of freedom $\nu = 1$. The marginals are the standard Gaussian distributions. Frame a) shows a scatter plot of the data and frame b) shows a corresponding tail frequency plot with the resolution threshold $\rho = 1$.

References

Klemelä, J. (2004), 'Visualization of multivariate density estimates with level set trees', *J. Comput. Graph. Statist.* **13**(3), 599–620.

Klemelä, J. (2006), 'Visualization of multivariate density estimates with shape trees', *J. Comput. Graph. Statist.* **15**(2), 372–397.

Klemelä, J. (2007), 'Visualization of multivariate data with tail trees', *Information Visualization* **6**, 109–122.

Nelsen, R. B. (1999), *An Introduction to Copulas*, Springer.