Visualization of the spread of multivariate distributions

Jussi Klemelä
Department of Statistics, Economics Faculty
University of Mannheim
L 7 3-5 Verfügungsgebäude
68131 Mannheim, Germany
Email: klemela@rumms.uni-mannheim.de
Fax +49 621 1811931

January 16, 2006

Abstract

We visualize the spread of multivariate distributions with the help of univariate and 2-variate functions. We distinguish between density type visualizations and distribution function type visualizations. Density type visualizations visualize the functional relationship between the level and the volume of the level sets of a density whereas distribution function type visualizations apply general sequences of nested sets and visualize the probability content of the sets as function of the volume of the sets. We present methods which are able to visualize the anisotropic spread of a distribution: these visualization tools show how the spread varies in different directions. We define 2D functions whose each mode correspond to a specific tail of the distribution.

Mathematics Subject Classifications (AMS 2000): 62G07, 62-09
Key Words: Exploratory data analysis, Heavy tailedness, Nonparametric density estimation, Quantile function, Shape transforms
Short title: Visualization of the spread

1 Introduction

We discuss visualization of the shape of a multivariate distribution and we concentrate on the non-central regions of the distribution. We consider continuous distributions and assume that the distribution is unimodal or nearly unimodal. The methods are usually robust with respect to small deviations from the unimodality; the density may have several small local extremes. When the distribution is a mixture of distributions whose supports are nearly disjoint, then it makes sense to study the shape separately for each component.

We define transformations of a multivariate distribution to a univariate or a 2-variate function to visualize the multivariate distribution. In the case of 2-dimensional distributions we may apply perspective plots or contour plots to draw the graph of the density or the graph of the distribution function, but also in the 2-dimensional case it is useful to apply 1D curves to visualize the spread of the distribution, since perspective plots and contour plots do not give a clear visualization of the tail behavior of the distribution. We apply two types of visualizations: (1) density type visualizations and (2) distribution and quantile function type visualizations.

In density type visualizations we look how the volumes of the level sets of a density are changing as function of the level. This leads to one-dimensional curves which visualize the spread of a multivariate distribution. We define the transformations in such a way that univariate symmetric densities remain unchanged. In multivariate cases the volumes of the level sets are often exploding when we move to the lower levels, and they are shrinking fast to zero when we move to the higher levels. In addition, the volumes of some multivariate sets vary irregularly as function of the dimension. For example, the volumes of unit balls vary irregularly as function of the dimension. To solve these problems we apply a dimension normalization.

Univariate density type visualizations do not visualize an anisotropic tail behavior. We say that a distribution has anisotropic tails when the spread is different in various directions. We propose to visualize the anisotropicity of the spread of a multivariate distribution with 2-dimensional functions, whose modes correspond to the extensions of the tails of the multivariate density. We apply the *radius transform* to the level sets of the density and then clue these transforms together to get a 2D function. The radius transform is a shape isomorphic transform of a multivariate set to a univariate function and it is defined in Klemelä (2004*c*).

In distribution and quantile function type visualizations we may apply other sequences of sets than the sequence of the level sets of the density. For example, we may look at the depth regions defined by a depth function.

We visualize the functional relation between the probability content and the volume of the sets in the sequence. We consider a nested sequence of sets, centered at a center point. This is different from the univariate case where the distribution function visualizes the cumulation of the probability mass when one travels from left to right.

Again, it is useful to apply a dimension normalization to get more easily conceivable one dimensional spread functions and it is useful to clue together shape isomorphic transforms of the sets to get a 2D function which visualizes anisotropic spread of the multivariate distribution. For distribution and quantile function type visualizations we apply the *cumulative tail probability transform*, which is such shape isomorphic transform which visualizes the probability content of the tails of a set.

We may be interested in the spread of a distribution but we have available only a sample of observations from the distribution. To make density type visualizations we need to estimate the level sets. Applying distribution and quantile function type visualizations is sometimes easier, since we may need only to calculate the empirical probabilities over some sequence of sets. However, it may be useful to apply also in distribution and quantile function type visualizations a sequence of estimates of level sets.

A multivariate function is given (stored in a computer) typically either by giving its values on a multivariate grid, or by defining it as a linear combination of some simple functions, for example as an expansion in an orthonormal system, or as a mixture of Gaussians. Kernel estimates are mixtures of scaled kernel functions. Orthogonal series estimators are expansions with basis functions. Boosting and bootstrap aggregation give density estimates which have a mixture form, with a large number of mixture members. Even when there are only few terms in the linear combination it is typically difficult to grasp the shape of the function, when we are only given the coefficients of the expansion. That is why we need visualization tools.

In Section 2 we define the density type visualizations. In Section 2.1 we define a unimodal volume transform, in Section 2.2 we define a dimension normalized unimodal volume transform, and in Section 2.3 we define a 2D volume transform. In Section 3 we define the distribution and quantile function type visualizations. In Section 3.1 we define univariate distribution and quantile functions of a multivariate distribution, in Section 3.2 we define the dimension normalized versions, and in Section 3.3 we define a 2D probability content function. Section 4 contains a discussion.

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

2 Density type visualizations

Density type visualizations apply the sequence $(A_\lambda)_{\lambda\in[0,\infty)}$ of the level sets of a density. The level set of a density $f : \mathbf{R}^d \to [0,\infty)$ with level λ is defined by

$$A_\lambda = \{x \in \mathbf{R}^d : f(x) \geq \lambda\}, \qquad \lambda \in [0,\infty). \tag{1}$$

Density type visualizations visualize the functional relation between the level and the volume of the level sets in the sequence. When the density is unknown, we have to estimate the level sets with a sample of observations from the distribution of the density. We may estimate the level sets by estimating the density function or we may estimate the level sets directly, see Klemelä (2004*a*) and the references there.

2.1 Unimodal volume function

2.1.1 Definition

We call a *level-to-volume* function a function which maps levels to the volumes of the level sets in the sequence. A *volume-to-level* function is the generalized inverse of this function. These univariate functions characterize the spread of a multivariate function.

Definition 1 *The* level-to-volume *function* $V : [0,\infty) \to [0,\infty)$*, associated to a multivariate density* $f : \mathbf{R}^d \to \mathbf{R}$*, is defined by*

$$V(\lambda) = \mathrm{volume}(A_\lambda), \qquad \lambda \in [0,\infty). \tag{2}$$

The volume-to-level *function* $V^{-1} : [0,\infty) \to [0,\infty)$*, is defined by*

$$V^{-1}(v) = \sup\{\lambda \in [0,\infty) : V(\lambda) \geq v\}, \qquad v \in [0,\infty),$$

where we use the convention $\sup \emptyset = 0$*.*

For example, when $f(x) = (1/2)I_{[-1,1]}(x)$ is the uniform density on $[-1,1]$, then $V(\lambda) = 2$, when $\lambda \in [0,1/2]$ and $V(\lambda) = 0$, when $\lambda > 1/2$. Function V does not have inverse but the generalized inverse is $V^{-1}(v) = 1/2$, when $v \in [0,2]$ and $V^{-1}(v) = 0$, when $v > 2$.

The volume-to-level function seems to be more natural to be used in visualizations. In addition, we prefer to modify this function so that the univariate symmetric densities remain unchanged through the transform. We reflect the volume-to-level function with respect to the origin and scale it with the factor 2 to get a symmetric density. We call this function a *unimodal volume function.*

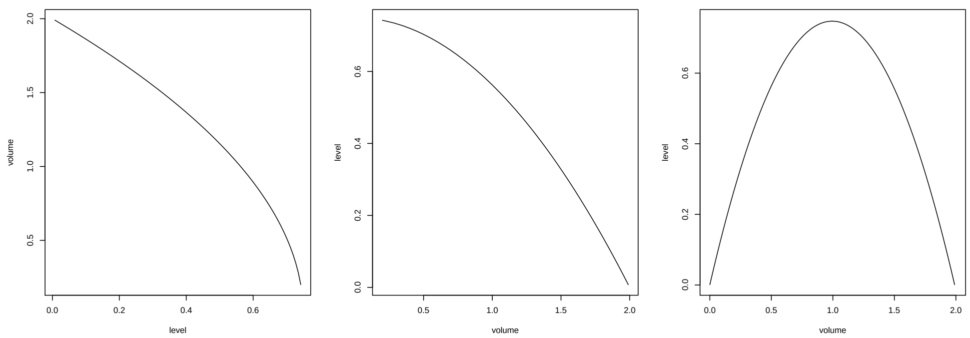

Figure 1: Illustration of Definition 1 and Definition 2; the level-to-volume function, the inverse of the level-to-volume function, and the unimodal volume function of the Bartlett density.

Definition 2 *A* unimodal volume function*, associated to a multivariate density* $f : \mathbf{R}^d \to \mathbf{R}$*, is any translation of function* $W : \mathbf{R} \to [0, \infty)$*,*

$$W(t) = \begin{cases} V^{-1}(2t), & t \geq 0 \\ V^{-1}(-2t), & t < 0, \end{cases}$$

where V^{-1} *is the volume-to-level function.*

We have that $\int_{-\infty}^{\infty} W = \int_0^{\infty} V^{-1} = 1$ and $W \geq 0$. We may use the term *unimodal volume transform*, since we have defined a transformation of a multivariate density to a univariate (unimodal symmetric) density. The unimodal volume transform of Definition 2 is related to the volume transform defined in Klemelä (2004*b*). When the density $f : \mathbf{R}^d \to \mathbf{R}$ is unimodal, then the definitions coincide. For multimodal densities the definition of Klemelä (2004*b*) is more informative since it visualizes the volumes of the separated regions of level sets, and thus it visualizes the relative largeness of the modes.

Illustration. Figure 1 illustrates Definition 1 and Definition 2. Frame a) shows a level-to-volume function of the univariate Bartlett density $t \mapsto (3/4)(1 - t^2)_+$, where $(t)_+ = \max\{0, t\}$. Frame b) shows the inverse of the level-to-volume function of the Bartlett density (volume-to-level function). Frame c) shows the unimodal volume function. We have positioned the unimodal volume function so that the left boundary of the support is at the origin.

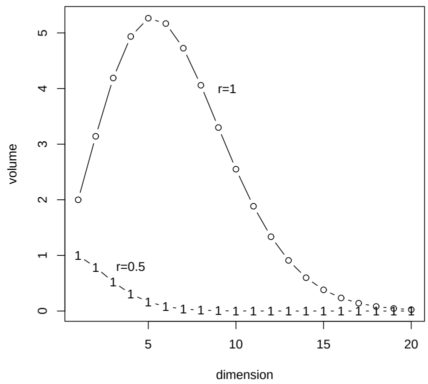

Figure 2: Volumes of balls with radius $r = 1/2$ and $r = 1$ as function of the dimension.

2.1.2 Illustrations with spherically symmetric densities

Let $f(x) = g(\|x\|^2)$, where $g : [0, \infty) \to \mathbf{R}$. We call g the density generator. We have

$$\{x : f(x) \geq \lambda\} = \{x : \|x\| \leq r_\lambda\}, \qquad r_\lambda = \sqrt{g^{-1}(\lambda)}, \tag{3}$$

when g is monotonically decreasing. The level-to-volume function is thus $V(\lambda) = \text{volume}(B_{r_\lambda})$, where $B_r = \{x \in \mathbf{R}^d : \|x\| \leq r\}$.

Volumes in high dimensional spaces. The volume of a ball with radius $r > 0$ is

$$\text{volume}(B_r) = C_d\, r^d, \tag{4}$$

where

$$C_d = \text{volume}(B_1) = \frac{\pi^{d/2}}{\Gamma(d/2+1)}\,. \tag{5}$$

Figure 2 shows the volumes of balls with radius $r = 1/2$ and $r = 1$ as function of the dimension. When $r = 1/2$, then the ball is inside a unit square and its volume is always less than one. As the dimension grows the dimension of the ball vanishes. When $r = 1$, then the volume is first increasing as the dimension increases but finally the volume starts decreasing. When $r > (\Gamma(3/2)\sqrt{\pi})^{-1} \approx 0.6366198$, then the volume is first increasing before it starts decreasing.

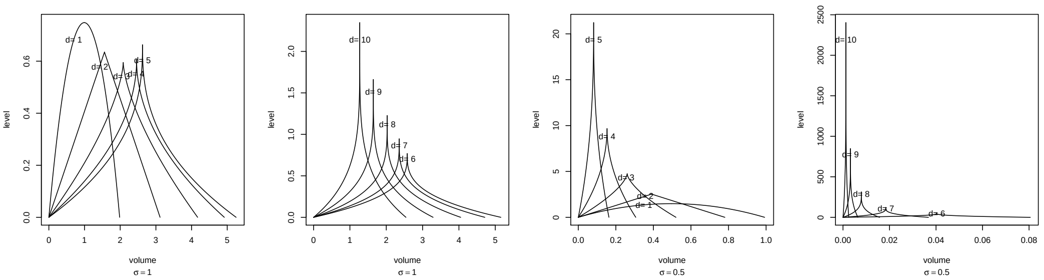

Figure 3: Frame a) shows unimodal volume functions of Bartlett densities with $\sigma = 1$ for dimensions $1-5$; in Frame b) $\sigma = 1$ and $d = 6-10$, in Frame c) $\sigma = 0.5$ and $d = 1-5$, in Frame d) $\sigma = 0.5$ and $d = 6-10$.

Bartlett densities. The Bartlett density generator is

$$g(t) = c \cdot (1-t)_+, \qquad t \in \mathbf{R}, \tag{6}$$

where $(a)_+ = \max\{0, a\}$, and $c = d(d+2)/[2\mu(S_d)]$, where $\mu(S_d) = \frac{2\pi^{d/2}}{\Gamma(d/2)}$ is the volume of the unit sphere. The multivariate scaled Bartlett density is $x \mapsto \sigma^{-d} g(\|x/\sigma\|^2)$. Figure 3a-b shows unimodal volume functions for dimensions $1-10$ of the standard Bartlett density. Since $\sigma = 1$, the volume of the support is increasing until $d = 5$, but then starts decreasing, according to Figure 2. Figure 3c-d shows unimodal volume functions for dimensions $1-10$ of the Bartlett density with $\sigma = 1/2$. Since $\sigma = 1/2$, the volume of the support is decreasing as function of the dimension.

Gaussian densities. Figure 4a-b shows unimodal volume functions of the standard Gaussian densities for dimensions $1-2$ and $3-4$. The volume of the support, with the given discretization level, is increasing when the dimension is increasing.

Student densities. The Student density generator is

$$g(t) = c \cdot (1 + t/\nu)^{-(d+\nu)/2}, \qquad t \in \mathbf{R}, \tag{7}$$

where $\nu > 0$ is the parameter (degrees of freedom) and $c = \frac{\Gamma((\nu+d)/2)}{(\pi\nu)^{d/2}\Gamma(\nu/2)}$. Figure 4c-d shows unimodal volume functions of the Student densities with degrees of freedom $\nu = 1$ for dimensions $1-2$ and $3-4$.

2.1.3 Problems

The examples in Section 2.1.2 have brought up some problems with the unimodal volume function.

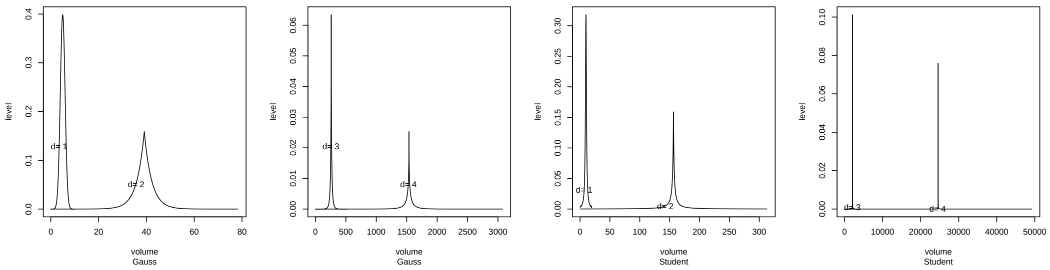

Figure 4: Frames a-b) show unimodal volume functions of the standard Gaussian densities for dimensions $1-2$ and $3-4$. Frames c-d) show unimodal volume functions of the Student densities with degrees of freedom $\nu = 1$ for dimensions $1-2$ and $3-4$.

(1) Concentration effect. The volume of the ball with radius r contains the term r^d. When r is small (smaller than 1), then the volume of the ball with radius r is very small, and when r is large (larger than 1), then the volume of the ball with radius r is very large. Thus a unimodal volume function has often a sharp peak at the center and its tails are spreading out. Note that in the case of visualizing multimodal densities with a volume transform, zooming may be used to see the details, as in Klemelä (2004*b*).

(2) Dimension non-invariance. The volume of a ball contains multiplier C_d which is not monotonic with respect to dimension d and thus the shapes of unimodal volume functions may vary irregularly when the dimension varies. For example, the unimodal volume functions of Bartlett densities have different shapes for various dimensions.

To address these problems we define a dimension normalized version of a unimodal volume function.

2.2 Dimension normalized unimodal volume function

We define a *dimension normalized unimodal volume function* which visualizes the shape of a multivariate density in a dimension insensitive way.

Definition 3 *The* dimension normalized level-to-volume function $V^* : [0,\infty) \to [0,\infty)$*, associated to a multivariate density* $f : \mathbf{R}^d \to \mathbf{R}$*, is defined by*

$$V^*(\lambda) = \left(\frac{1}{C_d} \, \text{volume}(A_\lambda) \right)^{1/d}, \qquad \lambda \in [0,\infty),$$

where C_d *is defined in (5). The* dimension normalized volume-to-level function $(V^*)^{-1} : [0,\infty) \to [0,\infty)$*, is defined by*

$$(V^*)^{-1}(v) = \sup\{\lambda \in [0,\infty) : V^*(\lambda) \geq v\}, \qquad v \in [0,\infty),$$

where we use convention $\sup \emptyset = 0$.

As in the case of non-dimension normalized functions we prefer the volume-to-level function and we symmetrize this function. In addition, we normalize the function to integrate to one which makes the function more dimension insensitive.

Definition 4 *A* dimension normalized unimodal volume function*, associated to a multivariate density* $f : \mathbf{R}^d \to \mathbf{R}$*, is any translation of function* $W^* : \mathbf{R} \to [0,\infty)$*,*

$$W^* = c\, W_0^*, \qquad W_0^*(t) = \begin{cases} (V^*)^{-1}(t), & t \geq 0 \\ (V^*)^{-1}(-t), & t < 0, \end{cases}$$

where $(V^*)^{-1}$ *is the dimension normalized volume-to-level function, and c is the normalization constant:* $c^{-1} = \int_{-\infty}^{\infty} W_0^*$.

As in the non-dimension normalized case, symmetric univariate densities remain unchanged through the transform, up to a translation.

Proposition 1 *When* $d = 1$ *and density* f *is symmetric unimodal, then* $W^* = f(\cdot - \mu)$*, for some* $\mu \in \mathbf{R}$.

Proof. We have $C_1 = 2$ and thus when $d = 1$, $W_0^* = W^*$ is a density whose level sets have lengths equal to the lengths of the level sets of f. □

Elliptical densities. Let $f(x) = |\det\Sigma|^{-1/2} g(x^T \Sigma^{-1} x)$ be an elliptical density, where $g : [0,\infty) \to \mathbf{R}$ is a density generator and Σ is a symmetric positive semi-definite dispersion matrix. We have

$$\{x : f(x) \geq \lambda\} = \{x : x^T \Sigma^{-1} x \leq r_\lambda^2\}, \qquad r_\lambda = \sqrt{g^{-1}\left(|\det\Sigma|^{1/2}\lambda\right)}, \qquad (8)$$

when g is monotonically decreasing. The volume of the ellipsoid in (8) is equal to

$$\text{volume}\left(\{x : x^T \Sigma^{-1} x \leq r_\lambda^2\}\right) = |\det\Sigma|\, C_d\, r_\lambda^d,$$

where C_d is defined in (5). The dimension normalized level-to-volume function is $\lambda \mapsto |\det\Sigma|^{1/d} r_\lambda$, where r_λ is defined in (8), and the dimension normalized volume-to-level function is $r \mapsto |\det\Sigma|^{-1/2} g(|\det\Sigma|^{-2/d} r^2)$, $r \geq 0$. A dimension normalized unimodal volume function is $t \mapsto c\, g(|\det\Sigma|^{-2/d} t^2)$, $t \in \mathbf{R}$, where $c^{-1} = |\det\Sigma|^{1/d} 2 \int_0^\infty g(u^2) du$. Thus a dimension normalized unimodal volume function depends on the dimension only through g and $|\det\Sigma|^{1/d}$. We have proved the following proposition.

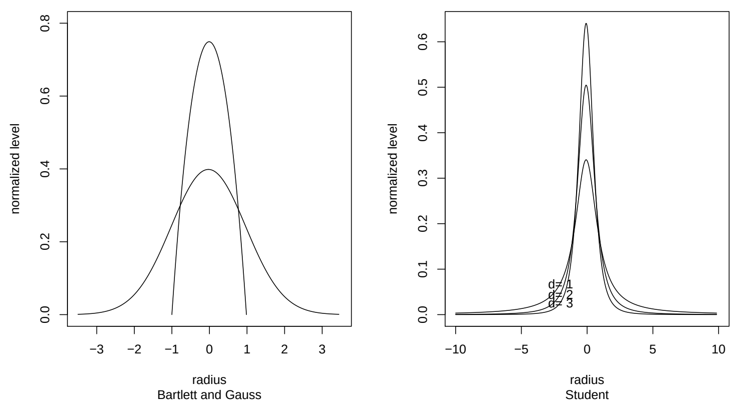

Figure 5: Dimension normalized unimodal volume functions for the Bartlett, standard Gaussian, and Student densities with degrees of freedom $\nu = 1$ and $d = 1 - 3$.

Proposition 2 *Let* $f_i : \mathbf{R}^{d_i} \to \mathbf{R}$, $i = 1, 2$, *be elliptical densities,* $f_i(x) = |\mathrm{det}\Sigma_i|^{-1/2} g_i(x^T \Sigma_i^{-1} x)$. *If* $g_i = c_i g$ *for some* $g : [0,\infty) \to [0,\infty)$, $i = 1, 2$, *and* $|\mathrm{det}\Sigma_1|^{1/d_1} = |\mathrm{det}\Sigma_2|^{1/d_2}$, *then the dimension normalized unimodal volume functions of the corresponding elliptical densities are equal, up to translation.*

In particular, for spherically symmetric Gaussian and Bartlett densities the dimension normalized unimodal functions are equal for all dimensions, up to translation. Note that the shape of the Student density generator depends on the dimension, in order the density to be integrable in all dimensions when $\nu > 0$. Figure 5 shows dimension normalized unimodal volume functions for the Bartlett, standard Gaussian, and Student densities with $\nu = 1$ and $d = 1 - 3$.

2.3 A 2D volume function

There are 2 ways to modify unimodal volume functions. (1) Klemelä (2004*b*) visualizes multimodality with 1D volume functions. (2) In this section we visualize anisotropic tail behaviour. We define a plot which visualizes the spread of a multivariate distribution so that one sees how the spread varies in different directions. We define a transformation of a multivariate density to a 2D function, which is called a *2D volume function*.

The definition of a 2D volume function is based on the concept of the *radius transform*, defined in Klemelä (2004*c*). The radius transform maps a multivariate connected set to a univariate radius function so that the modes of the 1D radius function correspond to the extensions of the set. The radius transform has 2 quantitative properties. (1) The lengths of the disconnected parts of the level sets of a radius function are equal to the volumes of the corresponding tail regions of the multivariate set. (2) The levels of the level sets of a radius function are equal to the distance of the corresponding tail region of the multivariate set from a center point of the set.

We define a normalized radius function so that its support is interval $[0,1]$ and it integrates to the volume of the set, or to the dimension normalized volume of the set. If $g_\lambda : [0,v] \to [0,\infty)$ is a radius function of level set A_λ, where $v = \mathrm{volume}(A_\lambda)$, then the normalized radius function is $h_\lambda : [0,1] \to [0,\infty)$,

$$h_\lambda(t) = c\, g_\lambda(v\,t), \qquad t \in [0,1],$$

where the normalization constant c may be chosen in 2 ways,

$$c = \begin{cases} v / \int_0^1 g_\lambda(vt)\, dt \\ v^* / \int_0^1 g_\lambda(vt)\, dt, \end{cases}$$

where $v^* = (v/C_d)^{1/d}$ is the normalized volume with C_d defined in (5).

There are at least 3 ways to clue normalized radius transforms $(h_\lambda)_{\lambda\in[0,\infty)}$ together to make a 2D function. (1) We let the normalized radius transforms to be the slices of a 2D function. (2) We let the graphs of the normalized radius transforms to be the level sets of a 2D function. (3) We let the normalized radius tranforms to be the boundary functions of 2D sets and define these sets to be the level sets of a 2D function. We choose the first of these ways, since it seems to lead to smooth 2D functions and this way of defining a 2D function does not require the nestedness of the graphs of the normalized radius transforms. The nestedness holds in typical cases but we cannot guarantee the nestedness in all cases.

Definition 5 *The* 2D volume function $\mathcal{V} : (0,\infty) \times [0,1] \to [0,\infty)$*, corresponding to density* $f : \mathbf{R}^d \to [0,\infty)$*, is the function whose slices are given by the normalized radius functions:*

$$\mathcal{V}(\lambda, t) = h_\lambda(t), \qquad t \in [0,1],$$

for each level $\lambda \in (0,\infty)$*.*

Figure 6 illustrates Definition 5. Frame a) shows a contour plot of a density with Gumbel copula with parameter $\theta = 2$ and the standard Gaussian marginals. Frame b) shows a radius function of the 10% level set. The radius function visualizes the egg-shape of the level set: the radius function has two

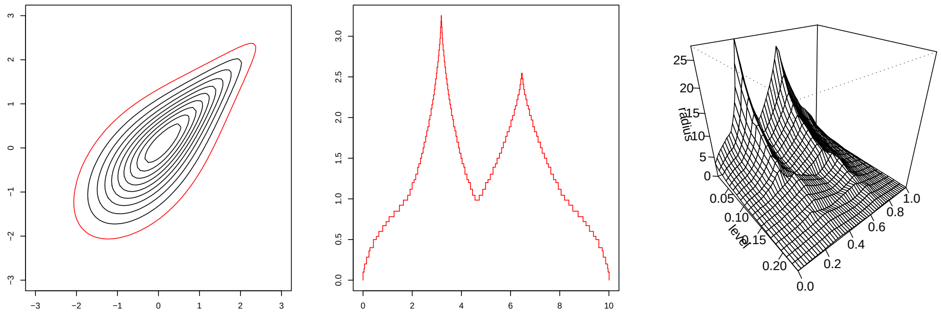

Figure 6: A contour plot of a density with Gumbel copula and standard Gaussian marginals, the radius function of the 10% level set, and the 2D volume function of the 2D Gumbel density.

modes of unequal size and shape. The shapes of the level sets do not change much when we move to the higher levels. Frame c) shows the 2D volume function of the density. We do not apply the dimension normalization in this example and in the examples below.

Figure 7a shows the 2D volume function of the product of two 1D Student densities, the other has degrees of freedom $\nu = 1$ and the other $\nu = 3$. The density has 4 extensions: it has heavy tails in each coordinate direction. Figure 7b shows the 2D volume function of the equal mixture of two 2D Gaussian densities, centered at the origin and whose marginal standard deviations are $(0.5, 1.5)$ and $(3.5, 0.5)$. Also this density has 4 extensions. The tails of the mixture of the Gaussians are less heavy than the tails of the product of the Student densities.

Figure 8a shows the 2D volume function of the product of three 1D Student densities, with degrees of freedom $\nu = 1$. This 3D density has 6 extensions, which show up as 6 modes in the 2D volume function. Figure 8b shows the 2D volume function of a 3D density with Gumbel copula with parameter $\theta = 2$ and the standard Gaussian marginals. The shape of the function is similar to the 2D Gumbel density shown in Figure 6c, but the 2D volume function of the 3D density has higher modes.

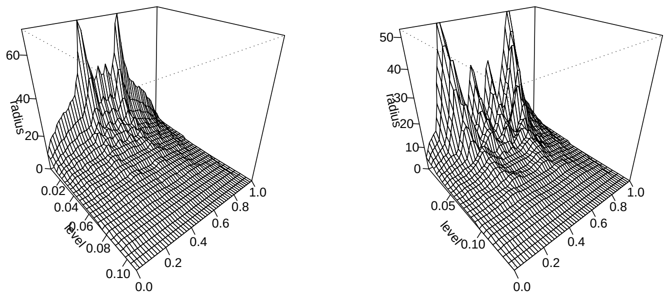

Figure 7: 2D volume functions of 2D densities: the product of two univariate Student densities and a mixture of two 2D Gaussians.

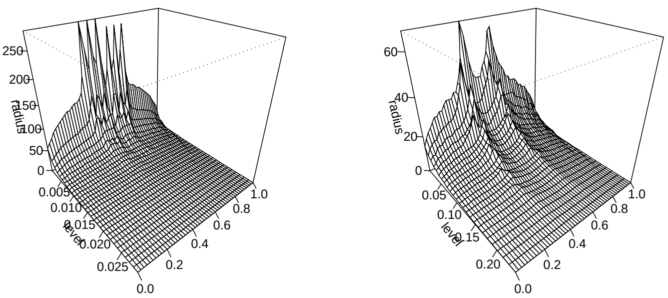

Figure 8: 2D volume function of 3D densities: the product of three univariate Student densities and a Gumbel density.

3 Quantile and distribution function type visualizations

In Section 2 we considered only sequencies of level sets of densities. When we define quantile and distribution function type visualizations, then it makes sense to consider also other sequencies of sets. We visualize the relationship between the probability content and the volume of the sets in the sequence.

Let $A_\lambda \subset \mathbf{R}^d$, $\lambda \in [0, \infty)$, be a collection of sets indexed with parameter λ. We assume that the collection of sets is nested and decreasing:

$$A_{\lambda_1} \supset A_{\lambda_2}, \text{ when } 0 \le \lambda_1 \le \lambda_2 < \infty. \tag{9}$$

We index the sequence of sets $(A_\lambda)_{\lambda \in [0,\infty)}$ with the probabilities:

$$C_p = A_{\lambda_p}, \qquad p \in [0, 1], \tag{10}$$

where

$$\lambda_p = \sup\{\lambda \in [0, \infty) : P(A_\lambda) \ge p\}. \tag{11}$$

The sequence A_λ is decreasing as λ is increasing and thus the level λ_p corresponding to probability p is well defined. We may consider the following cases.

1. The sequence of sets depends on the underlying distribution.
   - (a) The sequence is constructed assuming knowledge of the underlying distribution.
     - i. Level sets as defined in (1)
     - ii. Depth regions
     - iii. Minimum volume sets
     - iv. Central regions of a quantile function
   - (b) The sequence is estimated based on a sequence of identically distributed random vectors.
2. The sequence of sets does not depend on the underlying distribution.

Depth regions. We may choose the collection of sets to be the depth regions associated with various depth functions $D : \mathbf{R}^d \to [0, \infty)$: $A_\lambda = \{x \in \mathbf{R}^d : D(x) \ge \lambda\}$. A depth function D corresponding to a distribution function $F : \mathbf{R}^d \to \mathbf{R}$ is such that when $D(x)$ is large, then x is close to the center of the distribution (it is deep inside the distribution), and when

$D(x)$ is small (close to 0), then x is distant from the center. Examples of depth functions include Mahalabonis depth, the half-space depth defined by Hodges (1955) and Tukey (1975), and data depths based on simplices like the Oja depth defined by Oja (1983) and the simplicial depth defined by Liu (1990). The convex hull peeling depth function is defined only at the data points and it was considered by Barnett (1976).

We may relate many of the notions of data depth regions to level sets. Indeed, one way to define a depth function is to take the depth equal to the density: $D(x) = f(x)$. We may call this the density-depth, or likelihood-depth. When the density is elliptical, then the depth regions of a Mahalabonis depth with a suitable center vector and dispersion matrix are equal to the level sets of the density. The simplicial depths measure the local concentration of the probability mass and are thus related to the density-depth. The depth regions of the convex hull peeling depth estimate the level sets of densities whose level sets are convex.

Minimum volume sets. We may choose the sets C_p to be the minimum volume sets with a given probability content, when we perform the minimization over a given collection of sets. Let $\mathcal{S}$ be a class of measurable sets and define

$$C_p = \mathrm{argmin}_{C \in \mathcal{S}} \{\mathrm{volume}(C) : P(C) \geq p\}. \tag{12}$$

These sets where considered for example by Einmahl and Mason (1992) and Polonik (1999). Again, we may find a connection to level sets. When P has density $f : \mathbf{R}^d \to \mathbf{R}$, and class $\mathcal{S}$ is the class of Borel sets of $\mathbf{R}^d$, then the minimum in the definition of C_p is achieved by a level set of f, if f has no flat parts: $\mathrm{volume}(\{x : f(x) = \lambda\}) = 0$ for $\lambda > 0$. Indeed, for all measurable $C \subset \mathbf{R}^d$ with $P(C) \geq p$, $0 < p < 1$, we have that

$$\mathrm{volume}(C) \geq \mathrm{volume}(A_{\lambda_p}),$$

where $A_\lambda = \{x : f(x) \geq \lambda\}$ and $\lambda_p = \inf\{\lambda : P_f(A_\lambda) \leq p\}$.

Central regions of a quantile function. A multivariate quantile function may be defined to be a function $Q : S_d \times [0,1] \to \mathbf{R}^d$, where $S_d = \{x \in \mathbf{R}^d : \|x\| = 1\}$ is the unit sphere. In the univariate case we may define $S_1 = \{-1, 1\}$ and $Q(-1,p) = F^{-1}((1-p)/2)$ as the left quantile and $Q(1,p) = F^{-1}(1-(1-p)/2)$ as the right quantile, where F is the distribution function. Define the pth central region as

$$C_p = \{Q(u,q) : u \in S_d,\ \ 0 \leq q \leq p\}, \qquad p \in [0,1].$$

These sets were proposed by Serfling (2002a). Multivariate quantile functions may be defined as solutions to a minimization problem as in Chaudhuri (1996) and Koltchinskii (1997), but they may be defined also through boundaries of star shaped level sets, or star shaped depth regions, indexed with the probability content.

Sequences independent of the distribution. We may take the sequence to be independent of the underlying distribution. A simple and useful example is provided by the complements of balls, or the balls itself: $A_\lambda = \mathbf{R}^d \setminus \{x : \|x - \mu\| \le \lambda\}$ or $A_\lambda = \{x : \|x - \mu\| \le 1/\lambda\}$, where μ is a center point. Note that the center point μ has typically to be chosen based on the data.

3.1 Quantile and distribution function

A quantile function maps probabilities to the volumes and a distribution function is the generalized inverse of a quantile function.

Definition 6 *A* probability-to-volume function, *or a* quantile function, *corresponding to distribution P, is defined by*

$$Q(p) = \text{volume}\left(A_{\lambda_p}\right), \qquad p \in [0,1],$$

where λ_p is defined in (11). A volume-to-probability function, *or a* distribution function, *is defined by*

$$F(v) = \inf\{p \in [0,1] : Q(p) \ge v\}, \qquad v \in [0,\infty).$$

Alternatively, we may index the sequence of sets with volumes:

$$\lambda_v = \sup\{\lambda \in [0,\infty) : \text{volume}(A_\lambda) \ge v\}, \qquad v \in [0,\infty), \tag{13}$$

and define the distribution function as $v \mapsto P(A_{\lambda_v})$. The quantile function is then defined as the generalized inverse of the distribution function.

Liu, Parelius and Singh (1999) call a probability-to-volume-function a "scale curve" or a "scalar form of scale/dispersion", in the case sets A_λ are depth regions and Serfling (2002b) considers the case where the sets A_λ are the central regions determined by a multivariate quantile function.

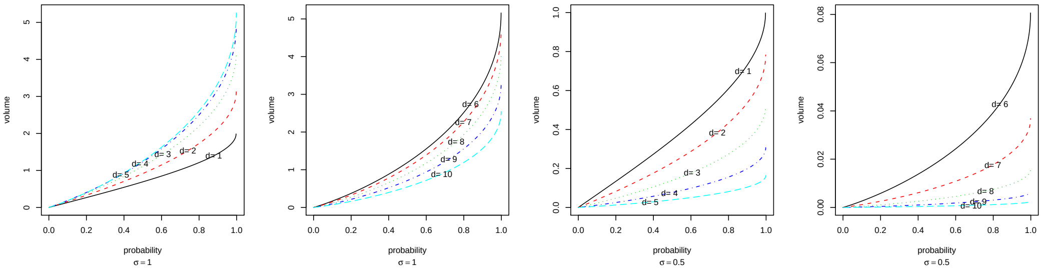

Figure 9: Probability-to-volume functions for Bartlett densities; in frame a) $\sigma = 1$ and $d = 1 - 5$, in frame b) $\sigma = 1$ and $d = 6 - 10$, in frame c) $\sigma = 0.5$ and $d = 1 - 5$, in frame d) $\sigma = 0.5$ and $d = 5 - 10$.

Bartlett distribution. Figure 9 shows probability-to-volume functions for the Bartlett density. Frame a) shows dimensions $1-5$ for $\sigma = 1$ and Frame b) shows dimensions $6-10$ for $\sigma = 1$. In the range $d = 1-5$ a quantile function corresponding to a higher dimension dominates a quantile function corresponding to a lower dimension, but in the range $d = 6 - 10$ this relation reverses. Frame c) shows dimensions $1 - 5$ for $\sigma = 0.5$ and Frame d) shows dimensions $6 - 10$ for $\sigma = 0.5$. Because $\sigma = 0.5$, a quantile function corresponding to a lower dimension dominates a quantile function corresponding to a higher dimension. The dominating relations are explained by Figure 2.

Gaussian distribution. Figure 10a-b shows probability-to-volume functions for the standard Gaussian density. Frame a) shows dimensions $1 - 2$ and Frame b) shows dimensions $3 - 4$. A quantile function corresponding to a higher dimension dominates a quantile function corresponding to a lower dimension.

Student distribution. Figure 10c-d shows probability-to-volume functions for the Student density with degrees of freedom $\nu = 1$. Frame a) shows dimensions $1 - 2$ and Frame b) shows dimensions $3 - 4$. As in the Gaussian case a quantile function corresponding to a higher dimension dominates a quantile function corresponding to a lower dimension.

3.2 Dimension normalized versions

Quantile and distribution functions have similar problems as unimodal volume functions: the concentration effect and the dimension non-invariance. The concentration effect for quantile functions means that the function is

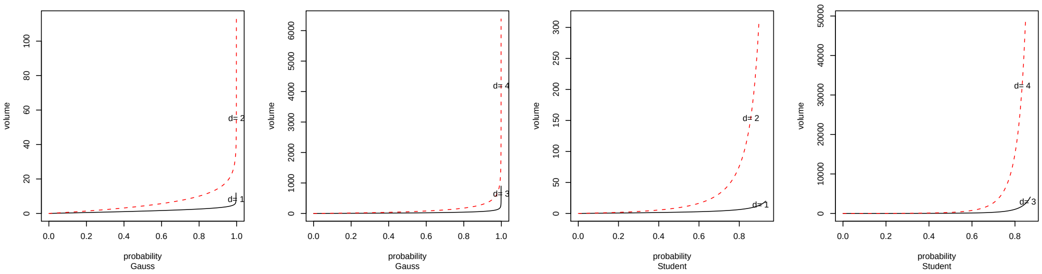

Figure 10: Frames a-b) show probability-to-volume functions for the standard Gaussian densities for dimensions $1-2$ and $3-4$. Frames c-d) show probability-to-volume functions for the Student density with degrees of freedom $\nu = 1$, for dimensions $1-2$ and $3-4$.

near zero at $[0, 1-\epsilon]$ but then explodes to take large values at $[1-\epsilon, 1]$. Thus we define a dimension normalized quantile and distribution function. A dimension normalized quantile function maps probabilities to the normalized volumes.

Definition 7 *A* dimension normalized probability-to-volume function, *or a* dimension normalized quantile function, *corresponding to distribution P, is defined by*

$$Q^*(p) = \left[\frac{1}{C_d}\,\text{volume}\left(A_{\lambda_p}\right)\right]^{1/d}, \qquad p \in [0,1],$$

where C_d is defined in (5). A dimension-normalized volume-to-probability function, *or a* dimension normalized distribution function, *is defined by*

$$F^*(v) = \inf\{p \in [0,1] : Q^*(p) \geq v\}, \qquad v \in [0,\infty).$$

As an example consider the case where the sequence of sets is a sequence of balls. The dimension normalized volume of a ball is equal to its radius. Thus the dimension normalized quantile function is equal to $p \mapsto r_p$, where r_p is the radius of a ball with probability content p.

Bartlett distribution. Figure 11a shows the dimension normalized probability-to-volume functions for the Bartlett density, for dimensions $1-20$. Figure 12a shows the dimension normalized volume-to-probability functions.

Gaussian distribution. Figure 11b shows the dimension normalized probability-to-volume functions for the standard Gaussian density, for dimensions $1-20$. Figure 12b shows the dimension normalized volume-to-probability functions.

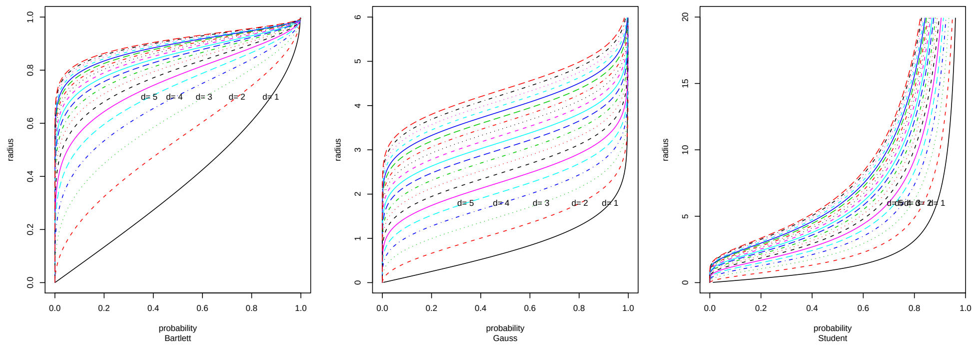

Figure 11: Dimension normalized quantile functions for dimensions 1 − 20; Frame a) shows the Bartlett distribution, Frame b) shows the standard Gaussian distribution, and Frame c) shows the Student distribution with degrees of freedom 1.

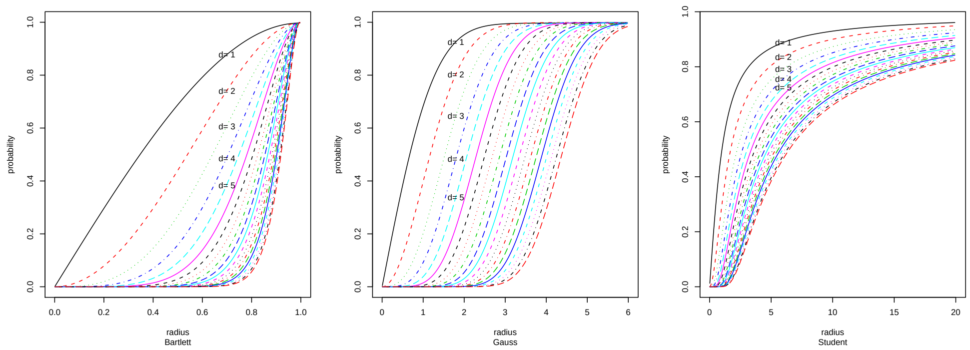

Figure 12: Dimension normalized distribution functions for dimensions 1 − 20; Frame a) shows the Bartlett distribution, Frame b) shows the standard Gaussian distribution, and Frame c) shows the Student distribution with degrees of freedom 1.

Student distribution. Figure 11c shows the dimension normalized probability-to-volume functions for the Student density with degrees of freedom $\nu = 1$, for dimensions $1 - 20$. Figure 12c shows the dimension normalized volume-to-probability functions for the same parameters. We plot the functions only up to radius 20, and thus the probability content of the sets does not reach 1.

3.3 A 2D probability content function

Quantile and distribution functions as defined in Section 3.1 and Section 3.2 do not visualize anisotropic tail behaviour. We may use similar techiques as in Section 2.3 to define a 2D function which visualizes the differences of the spread at various directions. In Section 2.3 we defined a 2D volume function by cluing together normalized radius functions. Now we clue together normalized *cumulative tail probability functions.* Similar to radius functions these are shape isomorphic transforms of a multivariate connected set to a univariate function. However, the length of a disconnected part of a level set of a cumulative tail probability function is equal to the probability content of the corresponding tail, and not equal to its volume, as in the case of a radius function. Figure 13a shows a cumulative tail probability function of the lowest level set in Figure 6a.

We define a normalized cumulative tail probability function so that a cumulative tail probability function is scaled to interval $[0, 1]$ and normalized to integrate to the probability content of the set. If $g_\lambda : [0, p] \to [0, \infty)$ is a cumulative tail probability function of set A_λ, where $p = P(A_\lambda)$, then the normalized cumulative tail probability function is $h_\lambda : [0, 1] \to [0, \infty)$,

$$h_\lambda(t) = c\, g_\lambda(p\, t), \qquad t \in [0, 1],$$

where $c = p / \int_0^1 g_\lambda(pt)\, dt = p^2 / \int_0^p g_\lambda$. We clue normalized cumulative tail probability functions together indexing the sets A_λ with their volumes, or with the dimension normalized volumes. Thus a 2D probability content function is an extension of a volume-to-probability function (distribution function).

Definition 8 *The* 2D probability content function $\mathcal{F} : (0, \infty) \times [0, 1] \to [0, \infty)$*, corresponding to distribution P, is a function whose slices are given by the normalized cumulative tail probability functions:*

$$\mathcal{F}(v, t) = h_{\lambda_v}(t), \qquad t \in [0, 1],$$

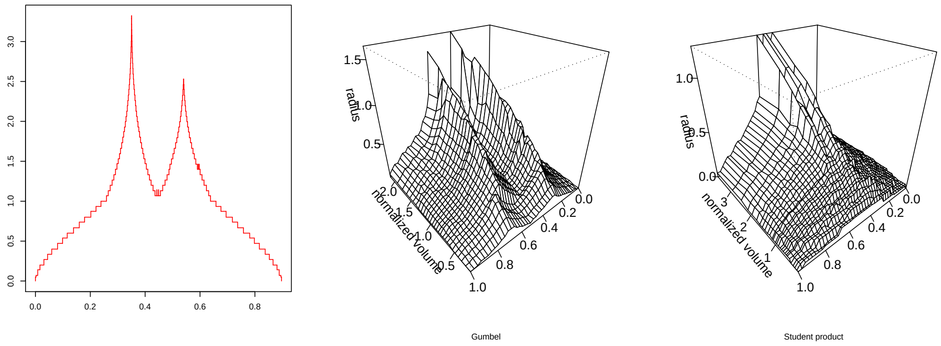

Figure 13: Frame a) shows a cumulative tail probability function of a Gumbel level set. Frames b) and c) show 2D probability content functions of 2D densities. Frame b) shows a distribution with a Gumbel copula and the standard Gaussian marginals and frame c) shows a product of two Student densities.

for $v \in (0,\infty)$, when λ_v is either defined by (13), or we index the sets with the dimension normalized volumes

$$\lambda_v = \sup\{\lambda \in [0,\infty) : \left(C_d^{-1}\text{volume}(A_\lambda)\right)^{1/d} \geq v\}, \qquad v \in [0,\infty),$$

where C_d is defined in (5).

Figure 13b shows the 2D probability content function of the distribution with Gumbel copula with parameter $\theta = 2$ and the standard Gaussian marginals. Note that Figure 6c shows the 2D volume function of the same distribution. Figure 13c shows the product of two univariate Student densities whose degrees of freedom are $\nu = 1$. Now the 2D probability content function has 4 modes. Sets A_λ are the level sets and we index the sets with the dimension normalized volumes.

4 Discussion

In the one dimensional case densities are superior in visualizing multimodality, but to visualize the spread and heavy tailedness we may use also distribution and quantile functions. The advantage of distribution and quantile functions is that they are easier to estimate than density functions. We have studied the multivariate case and we have defined one and two dimensional functions which visualize interesting features of the underlying multivariate distribution. We transform a multivariate function in such a way that certain shape characteristics become visible. This is an alternative for using projections and slices to visualize a function. We have emphasized the usefulness of level sets of densities in constructing the transforms. Many of the other proposals can be seen as related to the level set based approach.

Further proposals. When we visualize a function it is useful to compare the function to some well known reference case, for example to the standard Gaussian distribution. We may plot the two functions to the same window, or we may use PP-plot and QQ-plot type visualizations, where we plot the points $(g(t), \phi(t))$, $t \in \mathbf{R}$, to visualize the differences of function g to the reference function ϕ.

A further possibility to visualize the spread of a density f is to apply the real valued random variable $f(X)$, where $X \sim f$. Note that the density of random variable $f(X)$ is $g : [0, \infty) \to [0, \infty)$, $g(\lambda) = -\lambda V'(\lambda)$, where $V(\lambda)$ is the volume of the level set with level λ, as defined in (2). This was proved in Troutt (1991).

Liu et al. (1999) propose 4 univariate curves to visualize kurtosis, relative-spread, and heavy tailedness. (1) They propose to plot the Lorenz curve of $f(X)$, (2) they propose a data based Lorenz curve type plot using Mahalabonis distance, (3) a shrinkage plot visualizes the empirical frequencies of shrinked central hulls, and (4) a fan plot visualizes the volumes of the convex hulls of the central data points inside a central hull, in relation to the total volume of the central hull.

Acknowledgments

Writing of this article was financed by Deutsche Forschungsgemeinschaft under project MA1026/8-2.

References

Barnett, V. (1976), 'The ordering of multivariate data', *J. Roy. Statist. Soc. Ser. A* **139**, 319–354.

Chaudhuri, P. (1996), 'On a geometric notion of quantiles for multivariate data', *J. Amer. Statist. Assoc.* **91**, 862–872.

Einmahl, J. H. and Mason, D. M. (1992), 'Generalized quantile processes', *Ann. Statist.* **20**, 1062–1078.

Hodges, J. (1955), 'A bivariate sign test', *Ann. Math. Statist.* **26**, 523–527.

Klemelä, J. (2004*a*), 'Complexity penalized support estimation', *J. Multivariate Anal.* **88**, 274–297.

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

Klemelä, J. (2004*c*), 'Visualization of multivariate density estimates with shape trees'. Accepted to J. Comput. Graph. Statist., http://www.rni.helsinki.fi/~jsk/ps/unimod.ps.gz.

Koltchinskii, V. I. (1997), 'M-estimation, convexity and quantiles', *Ann. Stat.* **25**, 435–477.

Liu, R. Y. (1990), 'On a notion of data depth based on random simplices', *Ann. Statist.* **18**, 405–414.

Liu, R. Y., Parelius, J. M. and Singh, K. (1999), 'Multivariate analysis by data depth: descriptive statistics, graphics and inference', *Ann. Statist.* **27**(3), 783–858.

Oja, H. (1983), 'Descripitive statistics for multivariate distributions', *Statist. Probab. Lett.* **1**, 327–332.

Polonik, W. (1999), 'Concentration and goodness-of-fit in higher dimensions: (asymptotically) distribution-free methods', *Ann. Statist.* **27**(4), 1210–1229.

Serfling, R. (2002*a*), A depth function and a scale curve based on spatial quantiles, *in* Y. Dodge, ed., 'Statistical Data Analysis Based on the L1-Norm and Related Methods', Birkhäuser, pp. 25–38.

Serfling, R. (2002*b*), 'Quantile functions for multivariate analysis: approaches and applications', *Stat. Neerl.* **56**(2), 214–232.

Troutt, M. D. (1991), 'A theorem on the density of the density ordinate and an alternative interpretation of the Box-Muller method', *Statistics* **3**, 463–466.

Tukey, J. (1975), Mathematics and picturing data, *in* 'Proceedings of the 1975 International Congress of Mathematics', Vol. 2, pp. 523–531.