# Transmetric Density Estimation

Sigve Hovda<sup>1</sup>

<sup>1</sup> Department of Petroleum Engineering and Applied Geophysics, Norwegian University of Science and Technology, Trondheim, Norway

Correspondence: Sigve Hovda, Department of Petroleum Engineering and Applied Geophysics Norwegian University of Science and Technology, 7491 Trondheim, Norway. E-mail: sigve.hovda@ntnu.no

Received: December 7, 2015	Accepted: January 6, 2016	Online Published: February 10, 2016
doi:10.5539/ijsp.v5n2p35	URL: http://dx.doi.org/10.5539/ijsp.v5n2p35	

# Abstract

A transmetric is a generalization of a metric that is tailored to properties needed in kernel density estimation. Using transmetrics in kernel density estimation is an intuitive way to make assumptions on the kernel of the distribution to improve convergence orders and to reduce the number of dimensions in the graphical display. This framework is required for discussing the estimators that are suggested by Hovda (2014).

Asymptotic arguments for the bias and the mean integrated squared error is difficult in the general case, but some results are given when the transmetric is of the type defined in Hovda (2014). An important contribution of this paper is that the convergence order can be as high as 4/5, regardless of the number of dimensions.

Keywords: kernel density estimation, semiparametric models, pseudometrics, nonparametric functional data analysis, convergence order

# 1. Introduction

The multivariate kernel density estimator with a fixed bandwidth is given by

$$\hat{f}_{\mathbf{H}}(\mathbf{x}) = n^{-1} \sum_{i=1}^{n} K_{\mathbf{H}}(\mathbf{x} - X_i),$$
(1)

where  $K_{\mathbf{H}}(\mathbf{x}) : \mathbb{R}^m \to \mathbb{R}$  is a single modal, symmetric, nonnegative and zero-mean function that integrates to 1. The bandwidth matrix **H** describes the level of smoothing. When using a kernel *K* that is independent of **H**, then

$$\hat{f}_{\mathbf{H}}(\mathbf{x}) = n^{-1} |\mathbf{H}|^{-\frac{1}{2}} \sum_{i=1}^{n} K(\mathbf{H}^{-\frac{1}{2}}(\mathbf{x} - X_{i})).$$
(2)

The first univariate version of (2) was proposed by Fix and Hodges (1951), while the general class was investigated by Rosenblatt (1956) and Parzen (1962). The multivariate extension was outlined by Cacoullos (1966) and Epanechnikov (1969), while the full bandwidth matrix was discussed by Deheuvels (1977).

This paper involves using generalizations of the metrics in kernel density estimation. To improve readability, we recollect the following definitions.

**Definition 1.1.** A metric on a set X is a function (distance function)  $d : X \times X \to \mathbb{R}$ , which for all x, y, z in X, satisfies the following conditions:

- 1.  $d(\mathbf{x}, \mathbf{y}) \ge 0$  (nonnegativity)
- $2. \quad d(\mathbf{x}, \mathbf{x}) = 0$
- 3.  $d(\mathbf{x}, \mathbf{y}) = 0$  iff  $\mathbf{x} = \mathbf{y}$  (identity)
- 4.  $d(\mathbf{x}, \mathbf{y}) = d(\mathbf{y}, \mathbf{x})$  (symmetry)
- 5.  $d(\mathbf{x}, \mathbf{z}) \le d(\mathbf{x}, \mathbf{y}) + d(\mathbf{y}, \mathbf{z})$  (triangle inequality)

**Definition 1.2.** A metric space *M* is an ordered pair (*X*, *d*). An open and a closed  $\epsilon$ -ball on *M*, centered at *a*, are defined respectively as

$$\begin{split} \mathbb{B}_d(\boldsymbol{a}, \boldsymbol{\epsilon}) &:= \{ \boldsymbol{x} \in X | d(\boldsymbol{x}, \boldsymbol{a}) < \boldsymbol{\epsilon} \} \quad and \\ \mathbb{B}_d[\boldsymbol{a}, \boldsymbol{\epsilon}] &:= \{ \boldsymbol{x} \in X | d(\boldsymbol{x}, \boldsymbol{a}) \le \boldsymbol{\epsilon} \}, \end{split}$$

where  $\epsilon$  is called the radius of the ball.

If the third criterion (identity) is relaxed, d is a pseudometric and equivalently the pseudometric space and the ball is defined in the same way as in definition 1.2. Similarly, the semimetric is defined by relaxing the fifth criterion of definition 1.1. The quasimetric is defined by relaxing the fourth criterion and the pseudoquasimetric is defined by relaxing both the third and the fourth criterion. The premetric is the most general of them all, in which it keeps only the two first criteria.

The above definition of a pseudometric is exactly the same as the definition of a *semimetric* in the field of *nonparametric functional data analysis* (Ferraty and Vieu, 2006). To avoid confusion, we will use the term pseudometric, even though most of the literature in nonparametric functional data analysis uses the term semimetric.

In nonparametric functional data analysis, the explanatory variables are functional. Most of the work has been done on regression and classification and an overview of the field is given in the book by Ferraty and Vieu (2006). In the case of density estimation, Delaigle and Hall (2010) proved that a probability function generally does not exist for functional data, but developed notions of density and mode, when functional data is considered in the space determined by the eigenfunctions of the principal components. In the paper by Ferraty, Kudraszow and Vieu (2012), an infinite dimensional analogue of a density estimator is discussed, when the small ball probability associated with the functional data can be approximated as a product of two independent functions; one depending on the center of the ball and one depending on the radius.

In Hovda (2014), some finite dimensional pseudometric spaces are considered. Here, the balls with finite radius have finite volumes. Such spaces exist as a specification of the theory of nonparametric functional data analysis (Ferraty et al., 2012). The theory in Hovda (2014) is distinguished from that of nonparametric functional data analysis, because the pseudometric spaces that are considered have unit ball volumes that are dependent on the locations of centers. This is mitigateted by defining *equivalent distances of pseudometrics*.

In this paper, we propose the definition of the transmetric which generalized both the equivalent distances of pseudometrics as of Hovda (2014), the finite dimensional pseudometrics that are common in nonparametric functional data analysis and also metrics that are used in common kernel density estimation. Moreover, the formulation in this paper allows using multiple transmetrics. This is a generalization of Hovda (2014), where only one transmetric is allowed. It is proven that the convergence rates can be as high as 4/5 regardless of the number of dimensions. This was only indicated numerically in Hovda (2014).

Using transmetrics in density estimation involve training parameters of the transmetrics to fit the kernel of the distribution. To avoid confusion, it is worth emphasizing that the kernel of the distribution is a different concept than the kernel function, such as  $K(\cdot)$  in equation (2). The kernel of the distribution, or more roughly the family of level sets of the distribution, is defined as follows:

**Definition 1.3.** The kernel of a function  $f : X \to Y$  is an equivalence relation  $=_f$  given by  $\mathbf{x} =_f \mathbf{y} \Leftrightarrow f(\mathbf{x}) = f(\mathbf{y})$ , where  $\mathbf{x}$  and  $\mathbf{y}$  are elements in X.

The choice of the family of transmetrics, describes the constraints on the kernel of the distributions that can be modeled. In case of fitting parameters, it is therefore important that the level sets of the actual distribution are not in violation with these constraints. This is somewhat analogous to parametric models, where serious misbehavior are observed when the distribution has a different shape than any of the members of the family of models. If the true distribution is not in violation of the constraints imposed by the choice of family of transmetrics, it is shown in this paper that convergence orders can be improved and the distribution can be graphically displayed in fewer dimensions.

In section 2, *transmetric density estimation* is defined as a product of kernels that take transmetrics as arguments. The transmetric as of Hovda (2014) is outlined in section 3. In section 4, asymptotic arguments are discussed, while the proofs are given in section 5. The paper is concluded in section 6. Table 1 contains a list of symbols that are frequently used in this paper.

#### 2. Definition of Transmetric Density Estimators

In general there are two popular paths for constructing multi-dimensional kernels. The product kernels are products of single dimensional kernels, while the circular symmetric kernels are kernels that are dependent of the Euclidean distance from  $\mathbf{0}$ . We propose a different definition

$$\int_{\mathbb{R}^m} K(d(\boldsymbol{x}, \boldsymbol{y})) d\boldsymbol{y} = 1,$$
(3)

\_

Symbol	Explanation and properties	
n	Number of elements in data matrix.	
x	A point in the sample space.	
$x_i$	The <i>i</i> 'th component of $\boldsymbol{x}$ .	
$X_i$	The <i>i</i> 'th data point of a data set.	
$X_{i,j}$	The <i>j</i> 'th component of $X_i$	
m	Number of dimensions.	
q	Number of transmetrics.	
$\{P_1, P_2, P_q\}$	Partition of {1, 2, <i>m</i> }.	
$m_j$	Size of $P_j$ , i.e. $\sum_{j=1}^q m_j = m$ .	
Н	Symmetric bandwidth matrix of size $m \times m$ .	
$x_j$	The $P_j$ elements of $\mathbf{H}^{-1/2} \mathbf{x}$ .	
$y_j$	The $P_j$ elements of $\mathbf{H}^{-1/2}\mathbf{y}$ .	
$X_{i,j}$	The $P_j$ elements of $\mathbf{H}^{-1/2} X_i$ .	
$c_j$	Center point parameter in the <i>j</i> th transmetric (not the center of the balls).	
$p_j$	Vector of positive exponents in the <i>j</i> 'th transmetric.	
$ ilde{p}_j$	Harmonic average of the elements in $p_j$ .	
$V_{p_i}$	Volume of the generalized unit ball with parameters $p_j$ .	
$V_{t,j}$	Volume of the unit ball in the <i>j</i> th transmetric space.	
С	Concatenated vector of all the $c_j$ s. Size is $m$ .	
$c_j$	The <i>j</i> th component of $c$ , not any of the $c_j$ s.	
p	Concatenated vector of all the $p_j$ s. Size is $m$ .	
$p_j$	The <i>j</i> th component of $p$ , not any of the $p_j$ s.	
$d_{type,j,par}$	The <i>j</i> th premetric of type <i>type</i> with parameters <i>par</i> .	
$d_{type}$	Tuple of q premetrics of type <i>type</i> .	

Table 1. List of important symbols

where *m* is the number of dimensions and *d* is some metric. It is assumed that the kernel is non-negative and monotonically decreasing with finite first and second moments. This kernel generalizes all symmetric kernels when *d* is chosen to be Euclidean. It also generalizes all product kernels with convex level sets. The convex restriction is a consequence of the triangle inequality property of the metric. Moreover, it makes sense to also assume that the metrics need to be of at least differentiability class  $C^1$  and that they are translation invariant (i.e inducing a normed vector space).

One example is the normal kernel where the metric is Euclidean and

$$K(\epsilon) = \frac{1}{(2\pi)^{\frac{m}{2}}} \exp\left(-\frac{1}{2}\epsilon^2\right).$$
(4)

Another example is the uniform kernel where the metric is Chebyshev distance and

$$K(\epsilon) = \begin{cases} \frac{1}{3^{\frac{m}{2}}2^m} & \text{for } |\epsilon| \le \sqrt{3} \\ 0 & \text{for } |\epsilon| > \sqrt{3}. \end{cases}$$
(5)

Notice that the kernels are dependent on the metric being used. In this paper we propose a generalization of the metric as an input to the kernel.

**Definition 2.1.** A transmetric on a set  $\mathbb{R}^m$  is a function (distance function)  $d_t : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}^+$ , which for all  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^m$  and  $\epsilon \in \mathbb{R}^+$ , satisfies the following conditions:

1.	$d_t(\boldsymbol{x}, \boldsymbol{y}) \geq 0$	(nonnegativity)
2.	$d_t(\boldsymbol{x}, \boldsymbol{x}) = 0$	(mild identity)
3.	$(\mathbb{R}^m, \{\mathbb{B}_{d_t}(\boldsymbol{x}, \boldsymbol{\epsilon}) \mid \boldsymbol{\epsilon} > 0 \& \boldsymbol{x} \in \mathbb{R}^m\})$	
	is a topological space	(topology)
4.	$V_{\mathbb{B}_{d_t}(\boldsymbol{x},\epsilon)} = V_{\mathbb{B}_{d_t}(\boldsymbol{y},\epsilon)}$	(translation invariance
		of ball volumes)

This is a premetric with the additional criteria of topology and translation invariance of ball volumes. This is the reason for the *trans* prefix in the name. For comparison, a metric that induces a norm is a premetric with the additional criteria of translation invariance and the last three criteria of 1.1. Clearly, all metrics that induces a norm are transmetrics, which is practically all metrics that are used in common kernel density estimation. This is why definition 2.3 below, is a generalization of common kernel density estimation.

The key idea of this definition is to generalize the metric that induces a norm as much as possible, but still make sense for kernel density estimation. The first two criteria ensures that *close* points (in the sense of the transmetric) are weighted higher than points that are *further* away. The third criterion implies that if the kernel is continuous, then the estimator is also. The last criterion in definition 2.1, is important to achieve the result in theorem 4.1. It is worth emphasizing that relaxing the identity criterion yield some opportunities in smoothing along level sets of the distribution.

As commented on with regards to equation (4) and (5), there is a relationship between the kernel being used and the metric. This motivates the following definition.

**Definition 2.2.** A kernel K is said to be associated to a transmetric space  $(\mathbb{R}^m, d_t)$ , or associated to a transmetric  $d_t$ , if K is nonnegative, monotonically decreasing with compact support and satisfy

$$\int_{\mathbb{R}^m} K(d_t(\boldsymbol{x},\boldsymbol{y})) d\boldsymbol{y} = 1,$$

in the case when the unit ball  $V_{\mathbb{B}_{d_t}(\mathbf{x},1)}$  is finite and in the infinite case

$$\int_{\mathbb{R}^m} K(d_t(\boldsymbol{x},\boldsymbol{y})) d\boldsymbol{y} \propto V_{\mathbb{B}_{d_t}(\boldsymbol{x},1)}.$$

In the case when  $V_{\mathbb{B}_{d_t}(\mathbf{x},1)}$  is infinite, it is not meaningful to set the integral equal to one, since the kernel would clearly approach zero everywhere. To elaborate on this, we define the equivalence relation  $\sim_{d_t}$  as for all  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^m$ ,  $\mathbf{x} \sim_{d_t} \mathbf{y}$  iff  $d_t(\mathbf{x}, \mathbf{y}) = 0$ . It is clear that all equivalence classes in  $\mathbb{R}^m$  with respect to  $\sim_{d_t}$  are unbounded. To compensate for this unboundedness, we have decided to let the integral be proportional to  $V_{\mathbb{B}_{d_t}(\mathbf{x},1)}$ . Another way to handle this, would be to impose boundaries on the transmetric space (i.e let it be a subset of  $\mathbb{R}^m$ ), such that all  $\mathbb{B}_{d_t}(\mathbf{x}, 1)$ s would be bounded. This is not discussed further in this paper.

We are now ready to define transmetric density estimation.

**Definition 2.3.** Let  $\{P_1, P_2, ..., P_q\}$  be a partition of  $\{1, 2, ..., m\}$ , where  $m_j = |P_j|$ . Let  $\mathbf{x}_j$ ,  $\mathbf{y}_j$  and  $\mathbf{X}_{i,j}$  be vectors that, respectively, consist of the  $P_j$  elements of  $\mathbf{H}^{-1/2}\mathbf{x}$ ,  $\mathbf{H}^{-1/2}\mathbf{y}$  and  $\mathbf{H}^{-1/2}\mathbf{X}_i$ . If we choose a tuple of transmetrics  $\mathbf{d}_t$ , such that for every  $P_j$ , there is a  $K_j$  associated with the transmetric space  $(\mathbb{R}^{m_j}, d_{t,j}(\mathbf{x}_j, \mathbf{y}_j))$ , then the transmetric density estimator is defined as

$$\begin{split} \hat{f}_{\mathbf{H},d_{t}}(\mathbf{x}) &= |\mathbf{H}|^{-\frac{1}{2}} \int_{\mathbb{R}^{m}} \prod_{j=1}^{q} K_{j}(d_{t,j}(\mathbf{x}_{j}, \mathbf{y}_{j})) \hat{f}(\mathbf{y}) d\mathbf{y} \\ &= n^{-1} |\mathbf{H}|^{-\frac{1}{2}} \sum_{i=1}^{n} \prod_{j=1}^{q} K_{j}(d_{t,j}(\mathbf{x}_{j}, \mathbf{X}_{i,j})), \end{split}$$

where  $\hat{f}(\mathbf{y})$  is equal to 1/n at the  $X_i$ s and 0 elsewhere.

The idea is to weight the contributions of the  $X_{i,j}$ s according to the distances in the transmetric spaces. It is very important to note that in the case when any of the transmetrics has an unbounded unit ball,  $\hat{f}_{\mathbf{H},d_t}(\mathbf{x})$  is only proportional to  $f(\mathbf{x})$ . As said before, this is because the whole  $\mathbb{R}^m$  is considered, rather than a subset where all transmetric spaces have bounded unit balls.

In the trivial case, the estimator  $\hat{f}_{\mathbf{H},d_t}(\mathbf{x}) = \hat{f}_{\mathbf{H},d}(\mathbf{x})$ , where all  $P_j$  are equal to  $\{j\}$  and for all  $j, d_{t,j}(\mathbf{x}_j, \mathbf{y}_j) = |\mathbf{x}_j - \mathbf{y}_j|$ . If we let  $\prod_{i=1}^m K_i(u_i) = K(u)$ , then we arrive at the estimator defined in equation (1).

It is easy to see how any metric that defines a normed vector space can be used, but the interesting question is what other options do we have. What transmetrics make sense to use and which kernels can be associated to them? It is of particular interest to investigate the opportunities of relaxing the identity criterion. This motivates the definition of *associated distributions*, which was first defined in ?. This term is adapted to a tuple of transmetrics here.

Vol. 5, No. 2; 2016

**Definition 2.4.** A probability density function is said to be an associated distribution of  $\hat{f}_{\mathbf{H},d_t}$ , denoted as  $f_{\sim_{\mathbf{H},d_t}} : \mathbb{R}^m \to \mathbb{R}^+$ , when all  $\mathbf{x}_j \sim_{d_{t,j}} \mathbf{y}_j$ , imply that  $f_{\sim_{\mathbf{H},d_t}}(\mathbf{x}) = f_{\sim_{\mathbf{H},d_t}}(\mathbf{y})$  for all  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^m$ . Here, the equivalence relation  $\sim_{d_{t,j}}$  is defined as for all  $\mathbf{x}_j, \mathbf{y}_j \in \mathbb{R}^{m_j}, \mathbf{x}_j \sim_{d_{t,j}} \mathbf{y}_j$  iff  $d_{t,j}(\mathbf{x}_j, \mathbf{y}_j) = 0$ .

While discussing the usefulness of a family of transmetrics, it makes sense to ask which density functions are invariant to the equivalence relations that are implied by the transmetrics.

Another point to note is that the graphical display of the estimated density can be reduced to a *q*-dimensional display. This is because the estimated density at a certain point x is constant on  $\{y \in \mathbb{R}^m | y_i \in [x_i]_{\sim_{d_i}}\}$ .

All the transmetric spaces that will be discussed in this paper, have the following property regarding how the volumes of the balls vary as a function of the radius.

**Definition 2.5.** A transmetric  $d_t$  or a transmetric space  $(\mathbb{R}^m, d_t)$ , is said to be of order u, if the volume of the ball is on the form  $V_{\mathbb{B}_{d_t}(\mathbf{x},\epsilon)} = V_{\mathbb{B}_{d_t}(\mathbf{x},1)}\epsilon^u = V_t\epsilon^u$ , where  $\mathbf{x} \in \mathbb{R}^m$  and u is a positive integer. Such transmetrics are denoted  $d_{t,V_t,u}$ .

This property is useful as it makes it easy to find associated kernels.

**Theorem 2.1.** Given a transmetric space  $(\mathbb{R}^m, d_{t,V_{t,u}})$  of order u, then for any function  $g : \mathbb{R}^+ \to X$ , where  $X \subset \mathbb{R}$ , that is bounded and continuous almost everywhere, the following property is valid

$$\int_{\mathbb{B}_{d_{t,V_{t},u}}(\boldsymbol{x},U)} g(d_{t,V_{t},u}(\boldsymbol{x},\boldsymbol{y})) d\boldsymbol{y} = uV_{t} \int_{0}^{U} g(\epsilon) \epsilon^{u-1} d\epsilon,$$

where U is finite valued. If g also has compact support then

$$\int_{\mathbb{R}^m} g(d_{t,V_t,u}(\boldsymbol{x},\boldsymbol{y})) d\boldsymbol{y} = uV_t \int_0^\infty g(\boldsymbol{\epsilon}) \boldsymbol{\epsilon}^{u-1} d\boldsymbol{\epsilon}.$$

A direct consequence of this is that the normal and the uniform associated kernels of  $d_{t,V_{t,u}}$  are:

$$K(\epsilon) = \frac{1}{2^{\frac{u}{2}} \Gamma(\frac{u}{2}) V_t} \exp\left(-\frac{1}{2} \epsilon^2\right)$$
(6)

and

$$K(\epsilon) = \begin{cases} \frac{1}{3^{\frac{u}{2}} V_t} & \text{for } |\epsilon| \le \sqrt{3} \\ 0 & \text{for } |\epsilon| > \sqrt{3}, \end{cases}$$
(7)

when  $V_t$  is finite and in the infinite case, equation (4) and (5) can be used.

An important source for finding transmetrics is using pseudometrics. In general the volumes of the balls in pseudometric spaces with constant radius are dependent on the locations of the centers. This means that they can not be used directly. However, the following theorem shows how transmetrics can be designed from pseudometrics that are inducing a topology on  $\mathbb{R}^m$ .

**Theorem 2.2.** If  $(\mathbb{R}^m, d_p)$  is a pseudometric space and  $(\mathbb{R}^m, \{\mathbb{B}_{d_p}(\mathbf{x}, \epsilon) | \epsilon > 0 \& \mathbf{x} \in \mathbb{R}^m\})$  is a topological space, then  $d_{t,V_t,u}$  is a transmetric of order u, if

$$d_{t,V_t,u}(\boldsymbol{x},\boldsymbol{y}) = \left(\frac{V_{\mathbb{B}_{d_p}(\boldsymbol{x},d_p(\boldsymbol{x},\boldsymbol{y}))}}{V_t}\right)^{\frac{1}{u}},$$

where  $V_t$  is chosen to be the volume of the unit ball of the transmetric space  $(\mathbb{R}^m, d_{t,V_t,u})$ .

To fix ideas, the framework described in this section is required to discuss the transmetrics as of ?, which are discussed next.

#### 3. Transmetrics as of Hovda (2014)

In the paper by Hovda (2014), a family of pseudometrics is defined as the absolute value of the difference between two semimetrics. In that paper the concept of equivalent distance of the pseudometric is used. However, this is essentially the same as designing a transmetric by using theorem 2.2, where  $V_t$  is chosen to be the volume of either the *m*-sphere or the *m*-cube. Both choices are valid, but the mathematical formulation may be simpler depending on which kernel is chosen.



Figure 1. Generalized unit balls are shown in figure (a), (b) and (c). Balls in the pseudometric spaces ( $\mathbb{R}^3$ ,  $d_{p,0,p}$ ) are shown in figure (d) and (e). Here, one octant is removed. In figure (f), examples of the univariate generalized normal distribution are shown. This parametric family of symmetric unimodal distributions includes all normal and Laplace distributions, and as limiting cases it includes all continuous uniform distributions on bounded intervals of the real line. It includes a shape factor  $\beta$  and a scale factor  $\alpha$ . Various choices of shape factors are shown, where the scale factor is one and the mean is zero.

$$\begin{split} d_{t,V_{t},m}(\boldsymbol{x},\boldsymbol{y}) &= \left(\frac{V_{\mathbb{B}_{dp,c,p}(\boldsymbol{x},d_{p,c,p}(\boldsymbol{x},\boldsymbol{y}))}}{V_{t}}\right)^{\frac{1}{m}}, \text{ where} \\ d_{p,c,p}(\boldsymbol{x},\boldsymbol{y}) &= |d_{s,p}(\boldsymbol{x},\boldsymbol{c}) - d_{s,p}(\boldsymbol{y},\boldsymbol{c})|, \\ d_{s,p}(\boldsymbol{x},\boldsymbol{y}) &= \left(\sum_{i=1}^{m} |x_{i} - y_{i}|^{p_{i}}\right)^{\frac{1}{p}}, \\ V_{\mathbb{B}_{dp,c,p}(\boldsymbol{a},\epsilon)} &= \begin{cases} 2V_{p} \sum_{i \in O} {m \choose i} d_{s,p}(\boldsymbol{a},\boldsymbol{c})^{m-i} \epsilon^{i}, & \text{if } \epsilon < d_{s,p}(\boldsymbol{a},\boldsymbol{c}) \\ V_{p} \sum_{i=0}^{m} {m \choose i} d_{s,p}(\boldsymbol{a},\boldsymbol{c})^{m-i} \epsilon^{i}, & \text{if } \epsilon \ge d_{s,p}(\boldsymbol{a},\boldsymbol{c}) \end{cases} \\ \text{ and } \\ V_{p} &= 2^{m} \frac{\prod_{i=1}^{m} \Gamma(1 + \frac{1}{p_{i}})}{\Gamma(1 + \frac{m}{p})}. \end{split}$$

Here, *c* is a chosen point (for instance the mean or the mode of the distribution), *p* is a vector of positive  $p_i$ s and  $\tilde{p}$  is the harmonic average of the  $p'_i s$  (i. e.  $\tilde{p} = (1/m \sum_{i=1}^m p_i^{-1})^{-1}$ ). Moreover, *O* is the set of positive odd numbers that are smaller than or equal to *m*.

The unit ball of the semimetric space ( $\mathbb{R}^m$ ,  $d_{s,p}$ ) is called the generalized unit ball. The volume of this ball is  $V_p$ , which is proven by Wang (2005), Gao (2013) and Hovda (2014). In general  $d_{s,p}$  is a semimetric, because the triangle inequality does not hold when at least one of the  $p_i$ s are less than one. Depending on the  $p_i$ s, the generalized balls describe a wide range of multi-dimensional geometrical objects, where spheres, cubes, cylinders and stars are special cases. A few three dimensional objects are shown in figure 1(a), 1(b) and 1(c).

In the family of pseudometric spaces ( $\mathbb{R}^3$ ,  $d_{p,c,p}$ ), the balls can be viewed as generalizations of the spherical shell and two examples are shown in figure 1(d) and 1(e).

The formula for  $V_{\mathbb{B}_{d_{p,c,p}}(\boldsymbol{a},\epsilon)}$  is proven in Hovda (2014). The continuity of the ball at  $\epsilon = d_{s,p}(\boldsymbol{a}, \boldsymbol{c})$  is seen, when noting the fact that  $\sum_{i=0}^{m} {m \choose i} = 2^{m}$  and  $\sum_{i \in O} {m \choose i} = 2^{m-1}$ . In the case when  $\epsilon < d_{s,p}(\boldsymbol{a}, \boldsymbol{c})$ , the volume is a polynomial with respect to either  $\epsilon$  or  $d_{s,p}(\boldsymbol{a}, \boldsymbol{c})$ . The same is true when  $\epsilon \ge d_{s,p}(\boldsymbol{a}, \boldsymbol{c})$ , but the polynomial is different. This demonstrates how the

volumes of the balls are dependent on the locations of the centers. Notice also that the derivative of the volume is not continuous at  $\epsilon = d_{s,p}(a, c)$ , meaning that the derivative of the transmetric is also not continuous.

As discussed in Hovda (2014), a subset of the family of associated distributions is the generalized normal distributions (also known as the exponential power distribution). Some univariate cases are shown in figure 1(f). In Hovda (2014). In this paper several distributions are used to compare the new estimators with classical kernel density estimators. Because there is only one transmetric involved, the distributions can be shown using two dimensional plots. It is shown that the bias of multivariate distributions is significantly lower away from the mode, given that the underlying distribution is an associated distribution. In the present paper, more asymptotic results are given in section 4.1 and 4.2.

#### 4. Discussion of Bias, Variance and Mean Integrated Squared Error

Density estimators are commonly measured by global error criteria. Among these, we have chosen the mean integrated squared error MISE, which is given below

$$MISE(\hat{f}_{\mathbf{H},d_t}) = E\left(\int_{\mathbb{R}^m} (\hat{f}_{\mathbf{H},d_t}(\mathbf{x}) - f(\mathbf{x}))^2 d\mathbf{x}\right)$$
  
= 
$$\int_{\mathbb{R}^m} \left(\text{Bias}(\hat{f}_{\mathbf{H},d_t}(\mathbf{x}))\right)^2 d\mathbf{x} + \int_{\mathbb{R}^m} \text{Var}(\hat{f}_{\mathbf{H},d_t}(\mathbf{x})) d\mathbf{x}.$$
(8)

This expression is also known as the  $L^2$  risk function. In common kernel density estimation, there are two main paths for discussing the bias, the variance and the mean integrated squared error. For certain distributions, such as the mixture of normals, these statistics can be described analytically (Simonoff, 1996), but for more general distributions, analytic expressions are derived from an asymptotic argumentations (Rosenblatt, 1956; Cacoullos, 1966; Epanechnikov, 1969; Deheuvels, 1977).

In the case of transmetric density estimation, we can present this asymptotic result for the variance.

**Theorem 4.1.** The asymptotic variance of  $\hat{f}_{\mathbf{H},d_t}(\mathbf{x})$  is

$$Var(\hat{f}_{\mathbf{H},d_{t}}(\mathbf{x})) \approx n^{-1} |\mathbf{H}|^{-\frac{1}{2}} f(\mathbf{x}) \prod_{j=1}^{q} R(K_{j}), \text{ where}$$
$$R(K_{j}) = \int_{\mathbb{R}^{m_{j}}} K_{j}^{2}(d_{t,j}(\mathbf{x}_{j}, \mathbf{y}_{j})) d\mathbf{y}_{j},$$

which is independent of x.

In the case when  $d_{i,j}$  is a transmetric with an order, theorem 2.1 can be used to derive analytical expressions for  $R(K_j)$ . In the trivial case when  $\hat{f}_{\mathbf{H},d_i} = \hat{f}_{\mathbf{H},d_i}$ , the expression for the variance is identical to what is reported in Deheuvels (1977).

This result may be desirable for modelling purposes, as the asymptotic variance of the estimator is everywhere proportional to the distribution itself and the determinant of the bandwidth matrix. However, it is far more difficult to find asymptotic results for the bias and the mean integrated squared error.

In section 4.1, we present some asymptotic arguments for the bias in the case when the transmetric density estimator is applied on an associated distribution. In section 4.2, an asymptotic expression for MISE is derived when q is equal to one. These results ties well with the results of Hovda (2014).

#### 4.1 Asymptotic Arguments for the Bias

In this section we describe the asymptotic bias in the case when all transmetrics in  $d_t$  are of the type described in section 3 and in Hovda (2014). It is an important observation that  $d_{t,j}(x_j, y_j)$  is equal to  $|x_j - y_j|$ , when the dimension and the order  $m_j$  is one and  $c_j$  is either a very high or a very low. This means that the common kernel density estimator  $\hat{f}_{\mathbf{H},d}(\mathbf{x})$  is a special case. With regards to the associated distributions we have the following results.

**Lemma 4.1.** Let  $f_{\sim_{\mathbf{H},d_t}}(\mathbf{x})$  be an associated distribution of  $\hat{f}_{\mathbf{H},d_t}$  and let each  $d_{t,j}$  be of the type described in section 3 and designed by a pseudometric  $d_{p,j,c_j,p_j}$ . Also, let each  $d_{p,j,c_j,p_j}$  be constructed by a semimetric  $d_{s,j,p_j}(\mathbf{x}_j, \mathbf{c}_j)$ . Then  $f_{\sim_{\mathbf{H},d_t}}(\mathbf{x})$  is equal to a function  $g(\boldsymbol{\epsilon})$ , which satisfies

$$\int_{\mathbb{R}^m} g(\boldsymbol{\epsilon}) d\boldsymbol{x} = 1, \tag{9}$$

where each  $\epsilon_j = d_{s,j,p_i}(\mathbf{x}_j, \mathbf{c}_j)^{\tilde{p}_j}$ . Here  $\tilde{p}_j$  is the harmonic average of  $\mathbf{p}_j$ .

Lemma 4.2. The m-dimensional generalized normal distribution

$$f_{\sim_{\mathbf{H},d_t}}(\mathbf{x}) = |\mathbf{H}|^{\frac{1}{2}} \prod_{i=1}^m \frac{p_i}{2\Gamma(\frac{1}{p_i})} \exp\left(-|x_i - c_i|^{p_i}\right)$$
(10)

is an associated distributions of  $d_t$ , when each  $d_{t,j}$  is of the type described in section 3 and designed by a pseudometric  $d_{p,j}$ .

It is important to note that the family of generalized normal distributions is only a special case of the *total* family of associated distributions. The bias of the estimator is described next.

**Theorem 4.2.** Let each  $d_{t,j}$  of  $d_t$  be of the type described in section 3 and designed by a pseudometric  $d_{p,j}$  and denote  $\alpha_{x,j} = d_{s,j,p_j}(\mathbf{x}_j, \mathbf{c}_j), \ \boldsymbol{\alpha}_x = \{\alpha_{x,1}, \alpha_{x,2}, ..., \alpha_{x,q}\}, \ \alpha_{y,j} = d_{s,j,p_j}(\mathbf{y}_j, \mathbf{c}_j), \ \boldsymbol{\alpha}_y = \{\alpha_{y,1}, \alpha_{y,2}, ..., \alpha_{y,q}\}, \ \beta_{x,j} = (V_{p_j}/V_{t,j})^{1/m_j}\alpha_{x,j} \text{ and } \beta_{y,j} = (V_{p_j}/V_{t,j})^{1/m_j}\alpha_{y,j}.$  Here,  $V_{t,j}$  and  $V_{p_j}$  are the volumes of unit balls in the transmetric space  $(\mathbb{R}^{m_j}, d_{t,j})$  and the semimetric space  $(\mathbb{R}^{m_j}, d_{s,j})$ , respectively. Moreover, define  $f_{\mathbf{H}}(\boldsymbol{\alpha}_y) = f_1(|\mathbf{H}|^{1/(2m)}\boldsymbol{\alpha}_y) = f(\mathbf{y})$ , then

$$\begin{split} E(\hat{f}_{\mathbf{H},d_{t}}(\boldsymbol{x})) &\approx f_{\mathbf{H}}(\boldsymbol{\alpha}_{\boldsymbol{x}}) + (\boldsymbol{\psi} - \boldsymbol{\alpha}_{\boldsymbol{x}})D_{f_{\mathbf{H}}}(\boldsymbol{\alpha}_{\boldsymbol{x}}) + \\ &\frac{1}{2}tr\Big(H_{f_{\mathbf{H}}}(\boldsymbol{\alpha}_{\boldsymbol{x}})\Big[(\boldsymbol{\psi} - \boldsymbol{\alpha}_{\boldsymbol{x}})(\boldsymbol{\psi} - \boldsymbol{\alpha}_{\boldsymbol{x}})^{T} - diag^{2}(\boldsymbol{\psi}) + diag(\boldsymbol{\phi})\Big]\Big), \end{split}$$

where  $D_{f_{\rm H}}(\alpha_x)$  is a vector of the first order partial derivatives of  $f_{\rm H}$  and  $H_{f_{\rm H}}(\alpha_x)$  is the Hessian matrix. The entries of  $\psi$  and  $\phi$  are given by

$$\psi_{j} = m_{j} \frac{V_{i,j}^{\frac{m_{j}+1}{m_{j}}}}{V_{p_{j}}^{\frac{m_{j}+2}{m_{j}}}} \int_{0}^{\infty} K_{j}(U_{m_{j}}(\beta_{x,j},\beta_{y,j}))\beta_{y,j}^{m_{j}}d\beta_{y,j}$$

$$\phi_{j} = m_{j} \frac{V_{i,j}^{\frac{m_{j}+2}{m_{j}}}}{V_{p_{j}}^{\frac{m_{j}}{m_{j}}}} \int_{0}^{\infty} K_{j}(U_{m_{j}}(\beta_{x,j},\beta_{y,j}))\beta_{y,j}^{m_{j}+1}d\beta_{y,j}.$$
(11)

where

$$U_{s}(u,v) = \begin{cases} \left(2\sum_{i\in O_{s}} {s \choose i} u^{s-i} |v-u|^{i}\right)^{\frac{1}{s}}, & \text{if } |v-u| < u \\ \left(\sum_{i=0}^{s} {s \choose i} u^{s-i} |v-u|^{i}\right)^{\frac{1}{s}}, & \text{if } |v-u| \ge u, \end{cases}$$

where  $O_s$  is the set of positive odd numbers that are smaller than or equal to s.

Note that the expression for the expected value is only affected by the bandwidth matrix in the way that  $\alpha_x$  is constructed. We will get back to this in section 4.2.

In the trivial case when the estimator  $\hat{f}_{\mathbf{H},d_t}(\mathbf{x}) = \hat{f}_{\mathbf{H},d}(\mathbf{x})$ , each  $P_j$  is equal to  $\{j\}$ , each  $c_j$  goes to  $-\infty$ , each  $dV_{\mathbb{B}_{d_{s,j,p_j}}(c_j,\alpha_{y,j})}/d\alpha_{y,j}$  is equal to one,  $\boldsymbol{\psi} = \boldsymbol{\alpha}_{\mathbf{x}}$  and each  $\phi_j = \alpha_{x,j}^2 + \mu_2(K_j)$ , where  $\mu_2(K_j) = \int_{-\infty}^{\infty} u^2 K_j(u) du$ . This means that if all  $K_j(u)$  are the same,  $E(\hat{f}_{\mathbf{H},d}(\mathbf{x})) = f_{\mathbf{H}}(\alpha_x) + \frac{1}{2}\mu_2(K)tr(H_{f_{\mathbf{H}}})$ . Since  $H_{f_{\mathbf{H}}} = \mathbf{H}H_f$ , the bias is identical to what is reported in ?.

In the more general case, when the estimator is not just  $\hat{f}_{\mathbf{H}}(\mathbf{x})$ , some results are found when choosing the kernels to be uniform.

**Lemma 4.3.** Let all parameters be the same as in theorem 4.2, and also let the kernels be  $m_j$ -dimensional uniform kernels given by equation (5) and each  $V_{t,j} = 2^{m_j}$ . Then,

$$\psi_{j} = \begin{cases} \left(\frac{2^{m_{j}}}{V_{p_{j}}}\right)^{\frac{1}{m_{j}}} \frac{\sqrt{3}m_{j}}{m_{j}+1} & for \quad 0 \leq \beta_{x,j} < \frac{\sqrt{3}}{2} \\ \left(\frac{2^{m_{j}}}{V_{p_{j}}}\right)^{\frac{1}{m_{j}}} \frac{m_{j}}{m_{j}+1} \left(\beta_{x,j} + r_{m_{j}}(\beta_{x,j}) + \frac{2}{3^{\frac{2}{2}}} r_{m_{j}}(\beta_{x,j})(\beta_{x,j} - r_{m_{j}}(\beta_{x,j}))^{m_{j}} \right) & for \quad \frac{\sqrt{3}}{2} \leq \beta_{x,j} < \infty \end{cases}$$

$$\phi_{j} = \begin{cases} \left(\frac{2^{m_{j}}}{V_{p_{j}}}\right)^{\frac{2}{m_{j}}} \frac{3m_{j}}{m_{j}+2} & for \quad 0 \leq \beta_{x,j} < \frac{\sqrt{3}}{2} \\ \left(\frac{2^{m_{j}}}{V_{p_{j}}}\right)^{\frac{2}{m_{j}}} \frac{m_{j}}{m_{j}+2} \left((\beta_{x,j} + r_{m_{j}}(\beta_{x,j}))^{2} + \frac{4}{3^{\frac{m_{j}}{2}}}\beta_{x,j}r_{m_{j}}(\beta_{x,j})(\beta_{x,j} - r_{m_{j}}(\beta_{x,j}))^{m_{j}} \right) & for \quad \frac{\sqrt{3}}{2} \leq \beta_{x,j} < \infty. \end{cases}$$

$$(12)$$

where the function  $r_s(u)$  is defined by the formula

$$U_s(u, u + r_s(u)) = U_s(u, u - r_s(u)) = \sqrt{3}.$$

Analytical solutions of  $r_s(u)$  is only possible for s < 5, but since the functions have only s and u as variables, they can easily be put into a table for fast lookup on a computer. In order to describe this function more we have found these approximations

$$r_{s}(u) \approx \begin{cases} \sqrt{3} - u & \text{for } 0 \le u < \frac{\sqrt{3}}{2} \\ \frac{2}{\sqrt{3}s}(u - \frac{\sqrt{3}}{2})^{2} - (u - \frac{\sqrt{3}}{2}) + \frac{\sqrt{3}}{2} & \text{for } \frac{\sqrt{3}}{2} \le u < \sqrt{3} \\ \frac{3^{\frac{5}{2}}}{2su^{s-1}} & \text{for } \sqrt{3} \le u < \infty. \end{cases}$$
(13)

The top expression in equation (13) is exact and can easily be derived. The bottom expression is an approximation that is based on the fact that the volume of the ball in the far field is

$$V_{\mathbb{B}_{d_{p,j,e_j,p_j}(x_j,\epsilon)}} \approx \frac{dV_{\mathbb{B}_{d_{s,j,p_j}(e_j,\beta_{x,j})}}}{d\beta_{x,j}} 2\epsilon \quad \text{for} \quad \beta_{x,j} > \sqrt{3}.$$

The middle expression is a function that is fitted to ensure continuity between the other two expressions, including the derivative at  $\sqrt{3}/2$ .

# 4.2 Asymptotic Arguments for MISE Where q is Equal to One

In this section we consider the special case of the model in section 4.1, when q is equal to one. This is exactly the same case as in section 3 and in Hovda (2014). Here,  $\alpha_x = \alpha_{x,1}$  and  $\alpha_y = \alpha_{y,1}$ . This means that  $f(\mathbf{y}) = f_{\mathbf{H}}(\alpha_y)$ , then  $E(f_{\mathbf{H},d_t}(\mathbf{x}))$  and  $\sigma^2(f_{\mathbf{H},d_t}(\mathbf{x}))$  can be approximated by inserting equation (12) into theorem (4.2) and using theorem (4.1) so that

$$E(f_{\mathbf{H},d_{t}}(\mathbf{x})) \approx f_{1}(|\mathbf{H}|^{\frac{1}{2m}}\alpha_{x}) + |\mathbf{H}|^{\frac{1}{2m}}M_{1}(\beta_{x},m) \left. \frac{df_{1}}{du} \right|_{u=|\mathbf{H}|^{\frac{1}{2m}}\alpha_{x}} + |\mathbf{H}|^{\frac{1}{m}}M_{2}(\beta_{x},m) \left. \frac{d^{2}f_{1}}{du^{2}} \right|_{u=|\mathbf{H}|^{\frac{1}{2m}}\alpha_{x}}$$

$$\sigma^{2}(f_{\mathbf{H},d_{t}}(\mathbf{x})) \approx |\mathbf{H}|^{-\frac{1}{2}}R(K_{1})f_{1}(|\mathbf{H}|^{\frac{1}{2m}}\alpha_{x})n^{-1},$$
(14)

where

$$M_{1}(\beta_{x},m) = \begin{cases} \left(\frac{2^{m}}{V_{p}}\right)^{\frac{1}{m}} \left(\frac{\sqrt{3m}}{m+1} - \beta_{x}\right) & \text{for } 0 \leq \beta_{x} < \frac{\sqrt{3}}{2} \\ \left(\frac{2^{m}}{V_{p}}\right)^{\frac{1}{m}} \frac{m}{m+1} \left(-\frac{1}{m}\beta_{x} + r_{m}(\beta_{x}) + \frac{2}{3}}{\frac{2}{3}\frac{m}{2}} r_{m}(\beta_{x})(\beta_{x} - r_{m}(\beta_{x}))^{m}\right) & \text{for } \frac{\sqrt{3}}{2} \leq \beta_{x} < \infty \end{cases}$$

$$M_{2}(\beta_{x},m) = \begin{cases} \left(\frac{2^{m}}{V_{p}}\right)^{\frac{2}{m}} \left(\frac{\beta_{x}^{2}}{2} - \frac{\sqrt{3m}}{m+1}\beta_{x} + \frac{3m}{2(m+2)}\right) & \text{for } 0 \leq \beta_{x} < \frac{\sqrt{3}}{2} \\ \left(\frac{2^{m}}{V_{p}}\right)^{\frac{2}{m}} \frac{1}{(m+1)(m+2)} \left(\beta_{x}^{2} - \frac{mr_{m}(\beta_{x})\beta_{x}}{2} + \frac{m(m+1)}{2}r_{m}(\beta_{x})^{2} - \frac{2mr_{m}(\beta_{x})\beta_{x}}{3^{\frac{m}{2}}} \left(\beta_{x} - r_{m}(\beta_{x})\right)^{m}\right) & \text{for } \frac{\sqrt{3}}{2} \leq \beta_{x} < \infty \end{cases}$$

$$R(K_{1}) = \frac{1}{2^{m}3^{\frac{m}{2}}}.$$

$$(15)$$

The variance of  $\sigma^2(f_{\mathbf{H},d_t}(\mathbf{x}))$  is exactly the same as given in Epanechnikov (1969), when a uniform kernel is assumed.

When it comes to the bias, the results are different in several ways. The bias when  $\beta_x = 0$ , differs from the results of Epanechnikov (1969) in two ways. First the ball where the kernel is non zero has the shape of  $\mathbb{B}_{d_{s,p}(0,\sqrt{3})}$  and not the *m*-cube. The second difference is that the bias of the new estimator is dependent on the first derivative of *f*, which is not the case in Epanechnikov (1969).

However, the greatest difference is due to fact that the bias decreases as a function of  $\beta_x$ , regardless of the first and second derivatives. When *m* is at least 2 and  $\beta_x$  is large compared to  $r_m(\beta_x)$ , which is the case when  $\beta_x > \sqrt{3}$ , then  $(\beta_x - r_m(\beta_x))^m \approx \sum_{i=0}^2 {m \choose i} (-r_m(\beta_x)^i \beta_x^{m-i})$ . This means that for  $\sqrt{3} \le \beta_x < \infty$ , i. e. for any **x** such that  $\mathbf{H}^{-\frac{1}{2}} \mathbf{x} \notin \mathbb{B}_{d_{x,p}(\mathbf{0},\sqrt{3})}$ ,

$$M_{1}(\beta_{x},m) \approx \left(\frac{2^{m}}{V_{p}}\right)^{\frac{1}{m}} \frac{m(m-1)r_{m}(\beta_{x})^{2}}{2(m+1)\beta_{x}} = \left(\frac{2^{m}}{V_{p}}\right)^{\frac{1}{m}} \frac{(m-1)3^{m}}{8m(m+1)\beta_{x}^{2m-1}}$$

$$M_{2}(\beta_{x},m) \approx \left(\frac{2^{m}}{V_{p}}\right)^{\frac{2}{m}} \frac{mr_{m}(\beta_{x})^{2}}{(m+1)(m+2)} = \left(\frac{2^{m}}{V_{p}}\right)^{\frac{2}{m}} \frac{3^{m}}{4m(m+1)(m+2)\beta_{x}^{2m-2}}.$$
(16)

It is now clear that the bias is vanishing rapidly as  $\beta_x$  increases above  $\sqrt{3}$ . This result comply with the Monte Carlo simulations that is conducted on various distributions in Hovda (2014).

Assuming large n, i.e. small bandwidths, we can compute the asymptotic mean integrated squared bias AMISB based on the approximations in equation (16).

**Theorem 4.3.** Let all parameters be the same as in lemma 4.3 and also let q be equal to one. We define  $u_1$  as

$$mV_p \int_0^{u_1} u^{m-1} f_1(u) du = L \tag{17}$$

where L is a chosen limit, then for  $|\mathbf{H}|^{\frac{1}{2m}} \leq \frac{u_1}{\sqrt{3}} \left(\frac{V_p}{2^m}\right)^{\frac{1}{m}}$ , the asymptotic mean integrated squared bias

$$AMISB(f_{\mathbf{H},d_{t}}) \approx |\mathbf{H}|^{2} A(m, \boldsymbol{p}, f_{1}), \quad where$$

$$A(m, \boldsymbol{p}, f_{1}) = \left(\frac{(m-1)^{2} 3^{2m} 2^{4m-6}}{m^{2}(m+1)^{2} V_{\boldsymbol{p}}^{3}}\right) \int_{u_{1}}^{\infty} u^{-3m+1} \left(\frac{df_{1}(u)}{du}\right)^{2} du.$$
(18)

*Moreover, the optimal*  $|\mathbf{H}|^{\frac{1}{2}}$  *and optimal AMISE are* 

$$H_{AMISE}(f_{\mathbf{H},d_t}) = 4^{-\frac{1}{5}}A(m, \mathbf{p}, f_1)^{-\frac{1}{5}}R(K_1)^{\frac{1}{5}}n^{-\frac{1}{5}} = B_1(m, \mathbf{p}, f_1)n^{-\frac{1}{5}}$$

$$AMISE_{opt}(f_{\mathbf{H},d_t}) = 5 \cdot 4^{-\frac{4}{5}}A(m, \mathbf{p}, f_1)^{\frac{1}{5}}R(K_1)^{\frac{4}{5}}n^{-\frac{4}{5}} = B_2(m, \mathbf{p}, f_1)n^{-\frac{4}{5}},$$

$$where \quad R(K_1) = \frac{1}{2^{m_3}\frac{m}{2}}.$$
(19)

Notice that  $A(m, p, f_1)$  is independent of  $|\mathbf{H}|$ , given that  $|\mathbf{H}|$  is small enough, which is related to *L*, which is the accuracy of the approximation parameter. For instance, if we choose *L* to be 0.05, the bias in a region that includes the center and that has 0.05 probability is ignored in the estimation.

The important result here is that  $AMISE_{opt}(f_H)$  is proportional to *n* to the power of -4/5, i.e. the convergence order is 4/5. The convergence order is independent of the number of dimensions, which means that the new estimators defy the curse of dimensionality when *q* is equal to one and *f* is an associated distribution of  $f_{H,d_i}$ . It is also an important result that  $A(m, p, f_1)$  is straightforward to compute, which is relevant for finding pilot bandwidths as part of parameter estimation. theorem

For comparison, the common kernel density estimators, the H<sub>AMISE</sub> and AMISE<sub>opt</sub> are published in (Epanechnikov, 1969) as

$$H_{AMISE}(f_{H}) = \left(\frac{m \prod_{k=1}^{m} R(K_{k})}{Mn}\right)^{\frac{m}{m+4}}$$

$$AMISE_{opt}(f_{H}) = n^{-1}H_{AMISE}^{-1} \prod_{k=1}^{m} R(K_{k}) + \frac{1}{4}H_{AMISE}^{\frac{4}{m}}M, \text{ where}$$

$$M = \int \sum_{k=1}^{m} \frac{\partial^{2} f(\mathbf{x})}{\partial^{2} x_{k}} d\mathbf{x}.$$
(20)

Here, the convergence order of  $AMISE_{opt}(f_H)$  is 4/(4 + m), which is dependent of the number of dimensions.

# 4.2 Graphical Escription of B<sub>1</sub> and B<sub>2</sub>

In order to understand the behavior of  $B_1(m, p, f_1)$  and  $B_2(m, p, f_1)$ , we will discuss the multi-dimensional generalized normal distribution defined in equation (10). We let all elements of p be equal to p and the center be **0**. The integral part of equation (24) is derived as

$$\frac{p\Gamma(2-\frac{3m}{p},u_1)}{2^{2m+2-\frac{3m}{p}}\Gamma(1+\frac{1}{p})^{2m}},$$

which is found by substitution of variables to fit the definition of the incomplete upper gamma function  $\Gamma(\cdot, \cdot)$ . The  $u_1$  parameter is found by inserting  $f_1$  into equation (19) resulting in

$$u_1 = \left(\gamma^{-1}\left(\frac{m}{p}, \frac{p2^m\Gamma(1+\frac{1}{p})^mL}{mV_p}\right)\right)^{\frac{1}{p}},$$

where  $\gamma^{-1}(\cdot, \cdot)$  is the inverse of the lower incomplete gamma function.

In figure 2,  $log(B_1(m, p, f_1))$  and  $log(B_2(m, p, f_1))$  are plotted for various *m* and *p*, where the approximation limit *L* is set to 0.05 and 0.2. Based on the plots, it can be seen that for *p* larger than two, it is a fair simplification that  $B_1(m, p, f_1)$  and  $B_2(m, p, f_1)$  are exponential functions of *m* alone.

The effect of increasing *L* from 0.05 to 0.2 is minor, but can be seen as an upward shift of the values of  $log(B_1(m, p, f_1))$  and  $log(B_2(m, p, f_1))$ . Using the equations (17), (18) and (19), it can be determined what population sizes that these approximations are valid for. In the case of L = 0.05, the minimum required population sizes vary between  $4.4 \cdot 10^6$  and  $5.8 \cdot 10^7$ , when the number of dimensions range from two to eight. In the case of L = 0.2, the required population sizes are in the range of  $1.5 \cdot 10^5$  to  $3.2 \cdot 10^5$ , when the number of dimensions also range from two to eight.

# 5. Proofs

#### Proof of Theorem 2.1

We consider the uniform tagged partition  $\epsilon_0, \epsilon_1, ..., \epsilon_n$ , with the intervals  $t_0, t_1, ..., t_{n-1}$  on the interval [0, U]. We define a set of functions  $g_i$ s such that

$$g_i(\epsilon) = \begin{cases} G_i & \epsilon \in t_i \\ 0 & \epsilon \notin t_i, \end{cases}$$



Figure 2.  $log_{10}(B_1)$  and  $log_{10}(B_2)$  as functions of p and m. In the top row, the approximation limit L is set to 0.05, while in the bottom row L is set to 0.2. Increasing L from 0.05 to 0.2 can be seen as an upward shift of the values of  $log(B_1(m, p, f_1))$  and  $log(B_2(m, p, f_1))$ . For p larger than two, a rule of thumb is that  $B_1(m, p, f_1)$  and  $B_2(m, p, f_1)$  are exponential functions of m.

where  $G_i = g((\epsilon_i + \epsilon_{i+1})/2)$ . Therefore,  $g(\epsilon) = \lim_{n \to \infty} \sum_{i=1}^n g_i(\epsilon)$  almost everywhere. We recognize that  $d_{t,V_t,u}(\mathbf{x}, \mathbf{y}) = \epsilon(\mathbf{y}) = \epsilon$  and since  $g(\epsilon)$  is bounded, we can write

$$\begin{split} \int_{\mathbb{B}_{d_{t,V_{t,u}}}(\boldsymbol{x},U)} g(\boldsymbol{\epsilon}) d\boldsymbol{y} &= \int_{\mathbb{B}_{d_{t,V_{t,u}}}(\boldsymbol{x},U)} \lim_{n \to \infty} \sum_{i=1}^{n} g_{i}(\boldsymbol{\epsilon}) d\boldsymbol{\epsilon} \\ &= \lim_{n \to \infty} \sum_{i=1}^{n} G_{i} \left( \int_{\mathbb{B}_{d_{t,V_{t,u}}}(\boldsymbol{x},\epsilon_{i+1})} d\boldsymbol{\epsilon} - \int_{\mathbb{B}_{d_{t,V_{t,u}}}(\boldsymbol{x},\epsilon_{i})} d\boldsymbol{\epsilon} \right) \\ &= \lim_{n \to \infty} \sum_{i=1}^{n} G_{i} \left( V_{\mathbb{B}_{d_{t,V_{t,u}}}(\boldsymbol{x},\epsilon_{i+1})} - V_{\mathbb{B}_{d_{t,V_{t,u}}}}(\boldsymbol{x},\epsilon_{i})} \right). \end{split}$$

This is recognized to be equal to the Riemann-Stieltjes integral. This integral coincide with the Riemann integral, because  $V_{\mathbb{B}_{d_t V, \mu}}$  is differentiable and the derivative is continuous. In other words,

$$\int_{\mathbb{B}_{d_{t,V_{t,u}}}(\mathbf{x},U)} g(\epsilon)d\mathbf{y} = \int_{0}^{U} g(\epsilon)dV_{\mathbb{B}_{d_{t,V_{t,u}}}(\mathbf{x},\epsilon)}$$
$$= \int_{0}^{U} g(\epsilon)\frac{dV_{\mathbb{B}_{d_{t,V_{t,u}}}(\mathbf{x},\epsilon)}}{d\epsilon}d\epsilon = uV_{t}\int_{0}^{U} g(\epsilon)\epsilon^{u-1}d\epsilon$$

This concludes the first property. The second property is a result of the fact that g has compact support, which implies that

$$\lim_{U\to\infty}\int_{\mathbb{B}_{d_{t,V_{t,u}}}(\mathbf{x},U)}g(\epsilon)d\mathbf{y}=\int_{\mathbb{R}^m}g(\epsilon)d\mathbf{y}.$$

#### Proof of Theorem 2.2

The ball of  $\mathbb{B}_{d_{t,V_{t,u}}}(\boldsymbol{x}, \boldsymbol{\epsilon})$  is given by

$$\mathbb{B}_{d_{t,V_{t},u}}(\boldsymbol{x}, \boldsymbol{\epsilon}) = \{ \boldsymbol{y} \in \mathbb{R}^{m} | d_{t,V_{t},u}(\boldsymbol{x}, \boldsymbol{y}) \leq \boldsymbol{\epsilon} \} \\ \{ \boldsymbol{y} \in \mathbb{R}^{m} | V_{\mathbb{B}_{d_{p}}(\boldsymbol{x}, d_{p}(\boldsymbol{x}, \boldsymbol{y}))} \leq V_{t} \boldsymbol{\epsilon}^{u} \}$$

which proves that  $V_{\mathbb{B}_{d_{t,V_{t,m}}}(\mathbf{x},\epsilon)} = V_t \epsilon^u$ . The volume is therefore invariant to  $\mathbf{x}$ . The proof is completed by noting that the criteria of topology, nonnegativity and mild identity are obvious.

#### Proof of Theorem 4.1

The variance is given by

$$Var(\hat{f}_{\mathbf{H},d_t}(\mathbf{x})) = n^{-1} \left[ |\mathbf{H}|^{-1} \int_{\mathbb{R}^m} \prod_{j=1}^q K_j^2(d_{t,j}(\mathbf{x}_j, \mathbf{y}_j)) f(\mathbf{y}) d\mathbf{y} - (E(\hat{f}_{\mathbf{H},d_t}(\mathbf{x})))^2 \right]$$

and by changing variables, taylor expanding around x and neglecting all higher terms

$$Var(\hat{f}_{\mathbf{H},\boldsymbol{d}_{t}}(\boldsymbol{x})) = n^{-1} |\mathbf{H}|^{-\frac{1}{2}} f(\boldsymbol{x}) \prod_{j=1}^{q} \int_{0}^{\infty} K_{j}^{2}(d_{t,j}(\boldsymbol{x}_{j},\boldsymbol{y}_{j})) d\boldsymbol{y}_{j}$$

Proof of lemma 4.1

We recognize that the kernel of  $f_{\sim_{\mathbf{H},d_t}}(\mathbf{x})$  has the form

$$\left\{\left\{\boldsymbol{x} \in \mathbb{R}^{m} \mid \bigcap_{j=1}^{q} d_{s,j,\boldsymbol{p}_{j}}(\boldsymbol{x}_{j},\boldsymbol{c}_{j}) = \epsilon_{j}\right\} \forall \boldsymbol{\epsilon} \in \mathbb{R}^{q+}\right\},\$$

where  $\mathbb{R}^{q^+}$  indicates the positive octant in  $\mathbb{R}^q$ . Each level set is therefore uniquely determined by  $\epsilon$ , thus  $f_{\sim \mathbf{H},d_l}(\mathbf{x}) = g(\epsilon)$ . The proof is completed by noting the fact that  $f_{\sim \mathbf{H},d_l}(\mathbf{x})$  must integrate to one.

# Proof of lemma 4.2

We consider the special case when  $g(\epsilon)$  is written as a product of univariate functions  $g_j(\epsilon_j)$ , and each  $g_j(\epsilon_j)$  has the form of the exponential function  $a_j \exp(-\epsilon_j)$ . In this case the product of the coefficients  $a_j$  is found from equation (9) and theorem 2.1, such that

$$\begin{split} &\int_{\mathbb{R}^m} \prod_{j=1}^q a_j \exp(-\epsilon_j) d\mathbf{x} = 1 \\ |\mathbf{H}|^{-\frac{1}{2}} \prod_{j=1}^q a_j \int_{\mathbb{R}^{m_j}} \exp(-\epsilon_j) d\mathbf{x}_j = 1 \\ |\mathbf{H}|^{-\frac{1}{2}} \prod_{j=1}^q a_j \int_0^\infty \exp(-\epsilon_j) \frac{dV_{\mathbb{B}_{d_{s,jp_j}}(c_j, \epsilon_j^{1/\tilde{p}_j})}{d\epsilon_j}}{d\epsilon_j} d\epsilon_j = 1 \\ |\mathbf{H}|^{-\frac{1}{2}} \prod_{j=1}^q \frac{a_j V_{p_j} m_j}{\tilde{p}_j} \int_0^\infty \exp(-\epsilon_j) \epsilon_j^{m_j/\tilde{p}_j - 1} d\epsilon_j = 1 \\ |\mathbf{H}|^{-\frac{1}{2}} \prod_{j=1}^q \frac{a_j V_{p_j} m_j}{\tilde{p}_j} \Gamma\left(\frac{m_j}{\tilde{p}_j}\right) = 1 \\ |\mathbf{H}|^{-\frac{1}{2}} \prod_{j=1}^q a_j V_{p_j} \Gamma\left(1 + \frac{m_j}{\tilde{p}_j}\right) = 1 \\ \prod_{j=1}^q a_j = |\mathbf{H}|^{\frac{1}{2}} \prod_{i=1}^m \frac{p_i}{2\Gamma(\frac{1}{p_i})}. \end{split}$$

#### Vol. 5, No. 2; 2016

#### 

# Proof of theorem 4.2

We define  $d_{t1,j}(\alpha_{x,j}, \alpha_{x,j}) = d_{t,j}(\mathbf{x}_j, \mathbf{y}_j)$ , then

$$E(\hat{f}_{\mathbf{H},d_{t}}(\mathbf{x})) = |\mathbf{H}|^{-\frac{1}{2}} \int_{\mathbb{R}^{m}} \prod_{j=1}^{q} K_{j}(d_{t,j}(\mathbf{x}_{j},\mathbf{y}_{j}))f(\mathbf{y})d\mathbf{y}$$
  
$$= \int_{\mathbb{R}^{q^{+}}} \prod_{j=1}^{q} \left[ K_{j}(d_{t1,j}(\alpha_{x,j},\alpha_{y,j})) \frac{dV_{\mathbb{B}_{d_{s,j,p_{j}}}(c_{j},\alpha_{y,j})}}{d\alpha_{y,j}} \right] f_{\mathbf{H}}(\alpha_{y})d\alpha_{y},$$
(21)

where  $\mathbb{R}^{q^+}$  indicates the positive octant in  $\mathbb{R}^q$ . This is seen by changing variables to  $\mathbf{H}^{\frac{1}{2}}\mathbf{y}$  and utilizing theorem 2.1 q times. Taylor expanding around  $\alpha_x$ , gives  $f_{\mathbf{H}}(\alpha_y) = f_{\mathbf{H}}(\alpha_x) + (\alpha_y - \alpha_x)D_{f_{\mathbf{H}}}(\alpha_x) + 1/2(\alpha_y - \alpha_x)^T H_{f_{\mathbf{H}}}(\alpha_x)(\alpha_y - \alpha_x)$ . When inserted into equation (21), the first term in the expansion becomes  $f_{\mathbf{H}}(\alpha_x)$  and the second term becomes

$$\begin{split} &\left[\int_{\mathbb{R}^{q+}} \prod_{j=1}^{q} \left[ K_j(d_{t1,j}(\alpha_{x,j},\alpha_{y,j})) \frac{dV_{\mathbb{B}_{d_{x,j,p_j}}(c_j,\alpha_{y,j})}}{d\alpha_{y,j}} \right] \alpha_y d\alpha_y - \alpha_x \right] D_{f_{\mathbf{H}}}(\alpha_x) \\ &= (\psi - \alpha_x) D_{f_{\mathbf{H}}}.(\alpha_x). \end{split}$$

The third term becomes

$$\begin{split} &\frac{1}{2} \int_{\mathbb{R}^{q+}} \prod_{j=1}^{q} \left[ K_{j}(d_{t1,j}(\alpha_{x,j},\alpha_{y,j})) \frac{dV_{\mathbb{B}_{d_{x,jp_{j}}}(c_{j},\alpha_{y,j})}}{d\alpha_{y,j}} \right] \times \\ &(\alpha_{y} - \alpha_{x})^{T} H_{f_{H}}(\alpha_{x})(\alpha_{y} - \alpha_{x}) d\alpha_{y} \\ &= \frac{1}{2} tr \left( H_{f_{H}}(\alpha_{x}) \int_{\mathbb{R}^{q+}} \prod_{j=1}^{q} \left[ K_{j}(d_{t1,j}(\alpha_{x,j},\alpha_{y,j})) \frac{dV_{\mathbb{B}_{d_{x,jp_{j}}}(c_{j},\alpha_{y,j})}}{d\alpha_{y,j}} \right] \times \\ &(\alpha_{y}\alpha_{y}^{T} - 2\alpha_{x}\alpha_{y}^{T} + \alpha_{x}\alpha_{x}^{T}) d\alpha_{y} ) \\ &= \frac{1}{2} tr \left( H_{f_{H}}(\alpha_{x})(\alpha_{x}\alpha_{x}^{T} - 2\alpha_{x}\psi^{T} + \psi\psi^{T} - diag^{2}(\psi) + diag(\phi)) \right) \\ &= \frac{1}{2} tr \left( H_{f_{H}}(\alpha_{x})((\psi - \alpha_{x})(\psi - \alpha_{x})^{T} - diag^{2}(\psi) + diag(\phi)) \right), \end{split}$$

when the parameters  $\psi_i$  and  $\phi_i$  are

$$\psi_{j} = \int_{0}^{\infty} K_{j}(d_{t1,j}(\alpha_{x,j}, \alpha_{y,j})) \frac{dV_{\mathbb{B}_{d_{x,jp_{j}}}(c_{j}, \alpha_{y,j})}}{d\alpha_{y,j}} \alpha_{y,j} d\alpha_{y,j}$$
$$\phi_{j} = \int_{0}^{\infty} K_{j}(d_{t1,j}(\alpha_{x,j}, \alpha_{y,j})) \frac{dV_{\mathbb{B}_{d_{x,jp_{j}}}(c_{j}, \alpha_{y,j})}}{d\alpha_{y,j}} \alpha_{y,j}^{2} d\alpha_{y,j}.$$

The parameters  $\psi_j$  and  $\phi_j$  are manipulated to those defined in theorem 4.2, by inserting for the volumes and changing variables to the  $\beta_{y,j}$ s.

# Proof of lemma 4.3

This proof is straightforward by applying uniform kernels on equation (11).

# Proof of theorem 4.3

To keep the exposition simple we have decided to exclude the third term in equation (14) so that

$$\operatorname{AMISB}(f_{\mathbf{H},d_{t}}) \approx |\mathbf{H}|^{\frac{1}{2}} V_{\boldsymbol{p}} \times \int_{\sqrt{3} \left(\frac{2^{m}}{V_{\boldsymbol{p}}}\right)^{\frac{1}{m}}}^{\infty} \left( |\mathbf{H}|^{\frac{1}{2m}} M_{1}(\beta_{x},m) \left. \frac{df_{1}}{du} \right|_{u=|\mathbf{H}|^{\frac{1}{2m}} \alpha_{x}} \right)^{2} \alpha_{x}^{m-1} d\alpha_{x}.$$

$$(22)$$

From equation (16), we see that  $M_1(\beta_x, m)$  is

$$M_{1}(\beta_{x},m) \approx \left(\frac{2^{m}}{V_{p}}\right)^{2} \frac{(m-1)3^{m}}{8m(m+1)} \alpha_{x}^{-2m+1}$$

$$= |\mathbf{H}|^{1-\frac{1}{2m}} \left(\frac{2^{m}}{V_{p}}\right)^{2} \frac{(m-1)3^{m}}{8m(m+1)} u^{-2m+1},$$
(23)

where  $u = |\mathbf{H}|^{\frac{1}{2m}} \alpha_x$ . Changing variables to *u*, equation (22) becomes

$$AMISB(f_{\mathbf{H},d_{t}}) \approx |\mathbf{H}|^{2} \left( \frac{(m-1)^{2} 3^{2m} 2^{4m-6}}{m^{2} (m+1)^{2} V_{p}^{3}} \right) \times \int_{|\mathbf{H}|^{\frac{1}{2m}} \sqrt{3} \left(\frac{2^{m}}{V_{p}}\right)^{\frac{1}{m}}} u^{-3m+1} \left( \frac{df_{1}(u)}{du} \right)^{2} du.$$
(24)

Notice that there is a dependence of **H** on the lower limit of the integral. The expression for AMISB $(f_{\mathbf{H},d_t})$  is proven by assuming  $|\mathbf{H}|^{\frac{1}{2m}} \leq \frac{u_1}{\sqrt{3}} \left(\frac{V_p}{2^m}\right)^{\frac{1}{m}}$ .

The asymptotic mean integrated variance AMIV is easily derived from theorem 4.1 as

$$\operatorname{AMIV}(f_{\mathbf{H},d_t}) \approx |\mathbf{H}|^{-\frac{1}{2}} R(K_1) n^{-1}$$

and from equation (8) the asymptotic mean integrated squared error AMISE can be approximated as

AMISE
$$(f_{\mathbf{H},d_t}) \approx |\mathbf{H}|^2 A(m, \mathbf{p}, f_1) + |\mathbf{H}|^{-\frac{1}{2}} R(K_1) n^{-1}.$$
 (25)

The optimal  $|\mathbf{H}|^{\frac{1}{2}}$  and optimal AMISE are found by differentiating with respect to  $|\mathbf{H}|$  and setting this expression equal to zero.

# 6. Conclusion

The transmetric is defined, which is a generalization of the metric that is tailored to favorable properties needed in kernel density estimation. The idea is to use assumptions on the kernel, or the level sets, of the distribution to improve convergence orders and the simplify graphical display. Transmetric density estimation involves using multiple transmetrics, which allows an intuitive way to model a great variety of problems, while ensuring that the variance of the estimator is proportional to the distribution itself and the determinant of the bandwidth matrix.

Asymptotic results for the mean integrated squared error are difficult in the general case. For the transmetric as of Hovda (2014), the convergence order is shown to be as high as 4/5, independent of the number of dimensions.

## Acknowledgements

The author gratefully acknowledge the university for supporting this research. Moreover, the reviewers have given valuable feedback.

#### References

Cacoullos, T. (1966). Estimation of a Multivariate Density. Annals of the Institute of Statistical Mathematics, 18, 179-189. http://dx.doi.org/10.1007/BF02869528

- Deheuvels, P. (1977). Estimation non parametrique de la densite par histogrammes generalises (II). *Publ. lInst. Statist. lUniv.*, 22, 1-23.
- Delaigle, A., & Hall, P. (2010). Defining Probability Functions for a Distribution of Random Functions. The Annals of Statistics, 38, 1171-1193. http://dx.doi.org/10.1214/09-AOS741
- Epanechnikov, V. K. (1969). Non-parametric Estimation of a Multivariate Probability Density. *Theory of Probability* and its Applications, 14, 153-158. http://dx.doi.org/10.1137/1114019
- Ferraty, F., Kudraszow, N., & Vieu, P. (2012). Nonparametric Estimation of a Surrogate Density Function in Infinitedimensional Spaces. *Journal of Nonparametric Statistics*, 24, 447-464. http://dx.doi.org/10.1080/10485252.2012.671943
- Ferraty, F. Y., & Vieu, P. (2006). Nonparametric Functional Data Analysis: Theory and Practice, Springer.
- Fix, E., & Hodges, J. L. (1951). Discriminatory Analysis Nonparametric Discrimination: Consistency Properties, Technical report 4, United States Air Force School of Aviation Medicine, Randolph Field, Texas.
- Gao, F. (2013). Volumes of Generalized Balls. The American Mathematical Monthly, 120, 130.
- Hovda, S. (2014). Using pseudometrics in kernel density estimation. *Journal of Nonparametric Statistics*, 26, 669-696. http://dx.doi.org/10.1080/10485252.2014.944524
- Parzen, E. (1962). On Estimation of a Probability Density Function and Mode. *The Annals of Mathematical Statistics*, 33, 1065-1076. http://dx.doi.org/10.1214/aoms/1177704472
- Rosenblatt, M. (1956). Remarks on Some Nonparametric Estimates of a Density Function. The Annals on Mathematical Statistics, 27, 832-837. http://dx.doi.org/10.1214/aoms/1177728190
- Simonoff, J. S. (1996). Smoothing Methods in Statistics, Springer Series in Statistics. http://dx.doi.org/10.1007/978-1-4612-4026-6
- Wang, X. (2005). Volumes of Generalized Unit Balls. *Mathematics Magazine*, 78, 390-395. http://dx.doi.org/10.2307/30044198

## Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/3.0/).