The version of Record of this manuscript has been published and is available in Journal of Statistical Computation and Simulation. Published online: 10 Mar 2022 DOI: 10.1080/00949655.2022.2044480

MULTIRESOLUTION APPROXIMATON AND CONSISTENT ESTIMATION OF A MULTIVARIATE DENSITY FUNCTION

Rosa M. García-Fernández[a] (corresponding author)

Department of Quantitative Methods for Economics and Business . University of Granada

Federico Palacios-González .University of Granada

ABSTRACT. In this paper we extend multiresolution analysis structures on $\mathbb{R}^q$ to approximate multivariate probability density functions. We propose a consistent estimator for a multivariate multiresolution approximation (MMR) of a multivariate pdf. And we also develop an algorithm to estimate the MMR pdf that behaves well when handling big data. This algorithm performs better, in terms of running time, than traditional optimization algorithms. For large samples, the estimations are as good as those obtained by maximum likelihood. Numerical results are provided to illustrate the method.

Keywords: Multiresolution Analysis, Density Estimation, Cubic Box Spline, Frequency data count Algorithm, big data.

[a]Rosa M. García-Fernández (corresponding autor)

Department of Quantitative Methods for Economics and Business

University of Granada

Faculty of Economics and Business Sciences, office C-112

Campus Universitario de Cartuja, 18071 Granada. Spain.

Phone: 34 958248789.

E-mail: rosamg@ugr.es

MULTIRESOLUTION APPROXIMATON AND CONSISTENT ESTIMATION OF A MULTIVARIATE DENSITY FUNCTION

1.INTRODUCTION

Density estimation is a central task in statistics that has utilised parameter, semiparametric and nonparametric techniques (see for instance Silverman, 1986; Palacios-González and García-Fernández, 2014 a, b and Beylkin et al., 2019). The method developed in this paper is based on the use of multiresolution analysis to approximate multivariate density functions. The multiresolution analysis structures (MRA) are useful tools to approximate functions as well as to build wavelets. As is shown in Palacios-González and García-Fernández (2014 a, b) any squared integrable density can be approximated by elements of the multiresolution structure. The cited authors defined a univariate multiresolution family of probability density functions (MR pdf) by mixing dilations and translations of the squared integrable function known as cubic box spline (see Hernández and Weiss, 1996, p. 44 and Mallat, 1999, p. 221). The scaling function, Cubic Box Spline, generates subspaces Vj of the MR by means of non-orthonormal Riesz basis whose components are non-negative functions with compact support. The coefficients of the MR pdf are estimated by an algorithm based on frequency data count which is faster and easier to apply than the EM algorithm (see Palacios-González and García- Fernández, 2020).

The estimation of multivariate densities has received a lot of attention lately mainly due to its wide application to different fields of knowledge (see for instance Kasahara and Shimotsu, 2014). Some publications of the last decade on estimation of multivariate densities include Hunter and Levine (2015), Bonhomme et al. (2016), Bouzebda and Didi (2017); Li (2017); Zheng and Wu (2019); Luini, E and Arbenz, P. (2020) among others. Our approach provides an alternative tool to approximate and estimate multivariate square integrable densities.

Specifically, the main contributions of this paper are: (i) to extend the use of multiresolution analysis structures to the multivariate case to approximate probability density functions; (ii) to define a consistent estimator for a multivariate multiresolution approximation of a multivariate pdf and (iii) to develop an algorithm to estimate the MMR pdf. To build the multivariate multiresolution analysis structures we utilize multivariate

Cubic Box Spline function (see Wojtaszczyk, 1997 p. 105 and Eckley, 2001 p. 26). The multiresolution approximation of the multivariate pdf is made over the V_j spaces of the multivariate multiresolution structure (MMR). For any MMR approximation of a multivariate pdf a consistent estimator is defined. To estimate the MMR pdf, we propose an algorithm based on a straightforward process of data frequency count (FDC algorithm). The complexity of the algorithm is manageable and, as the empirical illustrations will show, it is fast enough to be applied to the field of big data (see section 5).

This paper is organized as follows. In section 2, the multivariate probability density functions are defined. Section 3 describes the approximation of a continuous multivariate density function. In section 4, a straightforward and consistent estimator of the density function is proposed. Section 5 develops an algorithm to estimate the MMR density. In addition, in this section the use of the Bayesian Information Criterium (BIC; Schwarz, 1978) to select the resolution level is described. In section 6, the algorithm is applied to the simulation models one and five of Zheng and Wu (2020). We estimate the processing time using different sample sizes. The ISEs for the joint and marginal distributions are also obtained. Section 7 shows an application to real data and section 8 concludes.

2.MULTIVARIATE SCALING FUNCTION CUBIC BOX SPLINE AND MULTIVARIATE MULTIRESOLUTION PDF

Let θ be a Cubic Box Spline function (see Hernández and Weiss, 1996 and Palacios-González and García-Fernández 2014a). Function θ is a symmetric probability density function with expected value zero and support in the interval $[-2,2]$. Given $\boldsymbol{x} = (x_1, x_2, \dots x_q) \in R^q$, the expression:

$$\Theta(\boldsymbol{x}) = \prod_{i=1}^{q} \theta(x_i) \qquad (1)$$

defines a Multivariate Cubic Box Spline (MCBS) . $\Theta(\boldsymbol{x})$ is the scaling function that allows us to build the Multivariate Multiresolution Analysis Structures used in this work (see Appendix A.1). $\Theta(\boldsymbol{x})$ is a multivariate probability density function with support in the hypercube $\prod_{i=1}^{q}[-2,2]$.

For each vector of integer numbers $\mathbf{k} = (k_1, k_2, \ldots, k_q) \in Z^q$ and each level of resolution $j \in Z$, we consider the following linear transformations of the MCBS function:

$$\Lambda_{j,\mathbf{k}}(\mathbf{x}) = 2^{qj} \prod_{i=1}^{q} \theta\left(2^j x_i - k_i\right) \qquad (2)$$

These linear transformations are multivariate density functions with support in the hypercube:

$$\prod_{i=1}^{q} \left[\frac{k_i - 2}{2^j}, \frac{k_i + 2}{2^j}\right]. \qquad (3)$$

2.1. PROBABILITY DENSITY FUNCTION IN A MULTIVARIATE MULTIRESOLUTION STRUCTURE

For any level of resolution $j \in Z$ a multivariate density function over a multiresolution structure is defined (see A.2 in appendix A):

$$f(\boldsymbol{x}) = \sum_{k \in Z^q} c_{\mathbf{k}} \Lambda_{j,\mathbf{k}}(\mathbf{x}) \qquad (4)$$

where $\{c_{\mathbf{k}}\}_{\mathbf{k} \in Z^q}$ is any set of real numbers that verifies the conditions $c_{\mathbf{k}} \geq 0 \; \forall \mathbf{k} \in Z^q$ and $\sum_{\mathbf{k} \in Z^q} c_{\mathbf{k}} = 1$ (see Theorem A.3 and Corollary A.3.1). In short, we will say that any density function like (4) is a Multivariate Multiresolution (MMR) density.

3. MMR APPROXIMATION OF A MULTIVARIATE PROBABILITY DENSITY FUNCTION

3.1 DEFINITION (see appendix B)

For all $j \in Z$ and all $\mathbf{k} = (k_1, k_2, \ldots k_q) \in Z^q$ the set $D_{\mathbf{k}}^j = \prod_{i=1}^{q} \left(\frac{k_i - 0.5}{2^j}, \frac{k_i + 0.5}{2^j}\right] \subset R^q$ is defined.

Note that each set $D_{\mathbf{k}}^j$ is a hypercube of volume $\frac{1}{2^{qj}}$ whose centre is the point $\frac{\mathbf{k}}{2^j}$. In addition, observe that the class of sets $\{D_{\mathbf{k}}^j\}_{\mathbf{k} \in Z^q}$ defines a partition of R^q.

3.2. DEFINITION (see Appendix B)

Let **X** be a q-dimensional random vector with probability density function f. For any level of resolution j, let $\{c_{\mathbf{k}}^j\}_{\mathbf{k} \in Z^q}$ be the sequence of real numbers built as follows:

$$c_{\mathbf{k}}^{j} = P\left(\mathbf{X} \in D_{\mathbf{k}}^{j}\right) = \int_{D_{\mathbf{k}}^{j}} f(\mathbf{x})\, d\mathbf{x} \qquad (5)$$

Using (5), the function

$$f_j(\mathbf{x}) = \sum_{\mathbf{k} \in Z^q} c_{\mathbf{k}}^{j} \Lambda_{j,\boldsymbol{k}(\boldsymbol{x})} \qquad (6)$$

is defined.

3.3 LEMMA

All $f_j(\mathbf{x})$ defined by (6) are MMR densities.

Proof. For each q-dimensional random vector $\mathbf{X}$ and all level of resolution $j \in Z$, it is verified that

$1 = P(\mathbf{X} \in R^q) = \sum_{\mathbf{k} \in Z^q} P\left(\mathbf{X} \in D_{\mathbf{k}}^{j}\right) = \sum_{\mathbf{k} \in Z^q} c_{\mathbf{k}}^{j}$. Hence, the coefficients $c_{\mathbf{k}}^{j}$, given by (5), are greater than or equal to 0 and add 1. Therefore, functions (6) are densities as those defined in section 2.1.

3.4 DEFINITION

The density functions $\{f_j\}_{j \in Z}$, obtained according to (6), are called f density approximations. Such denomination is based on Theorem 3.5.

3.5. THEOREM

It is verified that: $\lim_{j \to \infty} f_j(\mathbf{x}) = f(\mathbf{x})$

Proof. See B.5 and B.5.1 in appendix B.

4.ESTIMATION OF A MULTIVARIATE PROBABILITY DENSITY FUNCTION

In this section, a consistent estimator for any MMR approximation of a multivariate pdf for any continuous random variable with finite vector of means and finite variance and covariance matrix, is defined (Corollary C.1.2 in appendix C).

This estimator is calculated easily and quickly for any level of resolution using frequencies count on the sets $\boldsymbol{D}_{\mathbf{k}}^{\boldsymbol{j}}$.

Let $\mathbf{x}_t$ $t = 1,2, \dots, n$ be a finite random sample derived from a q-dimensional random vector $\mathbf{X}$ with probability density function f.

Let $n_{\mathbf{k}}$ $\mathbf{k} \in Z^q$ be the number of elements of the sample that belongs to $D_{\mathbf{k}}^{j}$. For any level of resolution $j \in Z$ and for all $\mathbf{k} \in Z^q$

$$\hat{c}_{\mathbf{k}}^{n,j} = \frac{n_{\mathbf{k}}}{n} \qquad (7)$$

is defined.

Finally, the estimator

$$\hat{f}_{n,j}(\mathbf{x}) = \sum_{\mathbf{k}\in \boldsymbol{Z}^q} \hat{c}_{\mathbf{k}}^{n,j} \ \Lambda_{j,k}(\mathbf{x}) \qquad (8)$$

is defined.

Note that when each sample data belongs to a different $D_{\mathbf{k}}^{j}$ and the frequency is equal to one in those n sets and zero in the rest, there will be at most n non-zero coefficients $\hat{c}_{\mathbf{k}}^{n,j}$

If for any $D_{\mathbf{k}}^{j}$ the corresponding $n_{\mathbf{k}}$ is greater than 1 then the number of non-zero coefficients $\hat{c}_{\mathbf{k}}^{n,j}$ must be less than n and the sum (8) always contains a finite number of nonzero addends.

4.1 THEOREM

For any level of resolution $j \in Z$ and all $\mathbf{x} \in R^q$ it is verified that (8) is a consistent estimator of (6).

Proof. See Lemma C1 and Corollary C.1.1 (appendix C).

We want to highlight the importance of Theorems 3.5 and 4.2. On the one hand, all square integrable function, and consequently all f, has a good enough approximation, f_j , in the space V_j when the level of resolution rises conveniently. On the other hand, each approximation $f_j(\mathbf{x})$ of $f(\mathbf{x})$ has an estimator $\hat{f}_{n,j}(\mathbf{x})$ that is consistent. Both results can be combined in the following Theorem.

4.2. THEOREM

$$\lim_{j\to\infty} \hat{f}_{n,j}(\mathbf{x}) \xrightarrow{p} \lim_{j\to\infty} f_j(\mathbf{x}) = f(\mathbf{x})$$

for almost all $\boldsymbol{x} \in R^q$.

Proof. See C.2 in appendix C.

The sequence of estimates, obtained from a succession of samples of increasing size, corresponds to a sequence of random variables with compact support[1] that converges in law towards a variable whose distribution is indistinguishable from the sampled variable

[1] Any MMR function with a finite number of nonzero coefficients has compact support.

(Theorem C.2). This occurs when the distribution that generates the sample has compact support or its range is R^q .

5. A MULTIVARIATE FREQUENCY DATA COUNT ALGORITHM TO ESTIMATE A MMR DENSITY

5.1. DEFINITION

Given a level of resolution j, for all $x \in R$ the values:

$$k(x) = Max\{k \in Z \mid k \leq 2^j x\}$$
$$r(x) = 2^j x - k(x) \qquad (9)$$

are defined (see Palacios-Gonzalez and García-Fernández,2020).

Observe that $k(x)$ and $r(x)$ depend on the level of resolution j, although for convenience of notation, it will not appear explicitly. Note that

$$x = \frac{k(x) + r(x)}{2^j} \quad \forall x \in R \qquad (10)$$

and that $\forall x \in R$ it is verified

$$r(x) = 0 \Leftrightarrow x = \frac{k(x)}{2^j}. \qquad (11)$$

5.2 LEMMA

Let $\mathbf{x} = (x_1, x_2, \ldots x_q) \in R^q$ be arbitrary. For all $i = 1,2,\ldots,q$

$$k_i = \begin{cases} k(x_i) & if\ r(x_i) \leq 0.5 \\ k(x_i) + 1 & if\ r(x_i) > 0.5 \end{cases} \qquad (12)$$

and $\mathbf{k} = (k_1, k_2, \ldots k_3)$, it is verified that $\mathbf{x} \in D_{\mathbf{k}}^j$.

Proof. See appendix D

5.3 STEPS OF THE ALGORITHM

To estimate the MMR density, according to (7) and (8), we propose the following frequency data count (FDC) algorithm.

Let us consider a sample $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ from a q dimensional random vector $\mathbf{X}$ and a level of resolution $j \in Z$. We represent by x_t^i $\forall i = 1,2,\ldots,q$ the i-th component of the sample element $\mathbf{x}_t$ $\forall t = 1,2,\ldots,n$.

FDC ALGORITHM

Step 1. Define a matrix $KData$ of dimension $n \times q$ to store the vector $\mathbf{k} = (k_1, k_2, \dots, k_q)$ associated with each sample data according to Lemma 5.2; define a matrix, $coef1$, of dimension $n \times (q+1)$ to count frequencies in $KData$, and define a matrix, $Usado$, of dimension n to control if each row of $KData$ has been used in the count.

Note that in the first q columns of each line of the $coef1$ there is a vector $\mathbf{k} = (k_1, k_2, \dots, k_q)$ of $KData$, which is different from the rest. Column $q+1$ contains the number of times that vector $\mathbf{k}$ is repeated in $KData$. Consequently, unless all $\mathbf{k}$ of $KData$ are different there will always be empty rows in $coef1$.

Step 2. For each $\mathbf{x}_t$ from $t = 1$ to n and each $i = 1$ to q make (see (12))

a) $k = int(2^j \times x_t^i)$

b) $r = x_t^i - k$

c) If $r > 0.5$ then $KData(t,i) = k+1$ else $KData(t,i) = k$

Step 3. From $t = 1$ to $n-1$, if $Usado(t) = 0$ make

a) $h = h + 1$

b) Copy row t of $KData$ in row h of $coef1$

c) $coef1(h, q+1) = coef1(h, q+1) + 1$

d) $Usado(t) = 1$

e) From $u = t+1$ to n, if $Usado(u) = 0$ make

e1) Compare row t of $KData$ with row u of $KData$

e2) If both are equal, then

e3) $coef1(h, q+1) = coef1(h, q+1) + 1$

e4) $Usado(u) = 1$

Step 4. If $Usado\ (n) = 0$ then make

a) $h = h + 1$

b) copy row n of $KDatos$ in row h of $coef1$

c) $coef1(h, q+1) = coef1(h, q+1) + 1$

Step 5. Define a $coef$ matrix of dimension $h \times (q + 1)$ to copy non-empty rows of $coef1$ in $coef$.

Step 6. Output of the estimated coefficients that are in $coef$ matrix

Observe that this algorithm is $O(n \times Max(q,h))$ being normally $q \ll n$, where h is the number of estimated coefficients (with non null frequency). Note that, without additional complications, the split and conquer method (Xueying and Minge, 2012) can be applied by splitting up the sample in p_1 subsamples of size n_1 equal to the integer[2] part of n/p_1 that will be processed in parallel resulting in an algorithm $O\left(n_1 \times Max(q,h)\right)$. If p_1 is high enough it will result in $n_1 \times Max(q,h)$ being a relatively low value.

For instance, if $n = O(10^7)$, $q = O(10^2)$ and $h = O(10^3)$, $p_1 = O(10^3)$ then $n_1 = O(10^4)$ and the resulting algorithm will be $O(10^6)$. In run time, this means that a sample with 10 million of individuals, with 100 data for each individual, could be processed in fractions of a second.

5.4. SELECTION OF THE LEVEL OF RESOLUTION

So far, the best level of resolution to obtain a good estimation has not been considered. We resolve this issue by employing the parsimony principle underlying in the *Bayesian Information Criterium* (BIC; Schwarz 1978).

The greater the level of resolution, the more accurate the approximation of a probability density function by a MRA density is (see Theorem 3.5). Consequently, given a fixed sample of size n, the sample likelihood given by $\hat{f}_{n,j}(\mathbf{x})$ increases with the level of resolution j. But for every unit we increase the level of resolution, the number of non-zero coefficients $\hat{c}_{\mathbf{k}}^{n,j}$ increases (it can be duplicated).

For instance, if we want to estimate a density by means of a MRA function as defined in (8) we have to use a level of resolution that achieves a trade-off between the maximum likelihood of the sample and the number of non-null coefficients of the estimator. We solve this problem using the Bayesian Information Criterion of Schwarz (1978) that we enunciate as follows.

[2] Obviously, unless n is a multiple of p_1, the last subsample will contain less than n_1 elements.

For any level of resolution j, let p_j be the number of non-null values of the sequence of coefficients $\{\hat{c}_{\mathbf{k}}^{n,j}\}_{k\in Z}$. Given a sample of size n, the coefficients are obtained by expression (7). Let L_j be the likelihood that the estimator (8) assigns to the sample. We select the level of resolution that minimizes $BIC(j) = -2logL_j + p_jLn(n)$.

6. SIMULATION RESULTS

We focus on models 1 and 5 used in the publication of Zheng and Wu (2020) to test our algorithm. These models are multivariate mixtures densities with independent marginal distributions. That is, given a p-dimensional random vector $\mathbf{X} = (x_1, x_2, \dots x_p)$ the density function is given by $f(\mathbf{x}) = \sum_{i=1}^{m} \pi_i f_i(\mathbf{x})$ where $\pi_i > 0\ \forall i = 1,2, \dots m$ and $\sum_{i=1}^{m} \pi_i = 1$ with $f_i(\mathbf{x}) = \prod_{j=1}^{p} f_i^j(x_j)$. In model 1, the random vector has dimension $p = 3$, and the mixture has two components ($m = 2$) the first, is the product of three independent $N(0,1)$ and the second is the product of $N(3,1)$, $N(4,1)$ and $N(5,1)$. The coefficients of the mixture are $\pi_1 = 0.60$ and $\pi_2 = 0.4$.

In model 5, the random vector has dimension $p = 3$ and $m = 3$. The three components of the mixture are the product of a Normal pdf, a double exponential pdf defined by $f_2^j(x_2) = \frac{1}{2} e^{-\left|x_2 - \mu_2^j\right|}$ and a pdf of a noncentral t distribution with ten degrees of freedom. The first component is the product of a $N(0,1)$ a double exponential with $\mu_1^2 = 3$ and a t with expected value $\mu_1^3 = 6$. The second component is the product of a $N(0,1)$ a double exponential with $\mu_2^2 = 4$ and a t with expected value $\mu_2^3 = 5$. The third is the product of a $N(0,1)$ a double exponential with $\mu_3^2 = 8$ and a t with expected value $\mu_3^3 = 10$. The parameters of the mixture are $(\pi_1,\ \pi_2,\ \pi_3) = (0.2,\ 0.3,\ 0.5)$.

This means that the first marginal density of model 5 is the mixture of three $N(0,1)$ (three times the same distribution) which is obviously $N(0,1)$. The second marginal density is the mixture of three different double exponential densities and the third is the mixture of three noncentral t densities (also different).

Table 6.1 shows the computation time to estimate the coefficients $c_{\mathbf{k}}$ of density (4) using the algorithm developed in section 5.2. The sample sizes are n =500; 1,000; 2,000; and 10,000.

Table 6. 1. Processing time in seconds to estimate the joint pdf

	Sample Size			
Models	500	1,000	2,000	10,000
1	0.0008	0.0015	0.0072	0.0140
5	0.0009	0.0019	0.0017	0.0019

Table 6.2 contains the root squared of the average ISE obtained using 100 samples of the marginal densities for each sample size.

Table 6.2. $\sqrt{ISE}$

$\sqrt{ISE}$		Sample size			
Model	Marginal	500	1000	2000	10000
1	1	0.0434	0.0392	0.0328	0.0164
5	1	0.0405	0.0275	0.0237	0.0227
1	2	0.0450	0.0413	0.0339	0.0178
5	2	0.0446	0.0353	0.0304	0.0250
1	3	0.0453	0.0399	0.0327	0.0182
5	3	0.0313	0.0248	0.0155	0.0096

Figures 6.1, 6.2 and 6.3 show the marginal densities (gray colour) and their estimations (black colour) for $n = 10{,}000$.

Figure 6.1. Marginal density 1 and its estimation

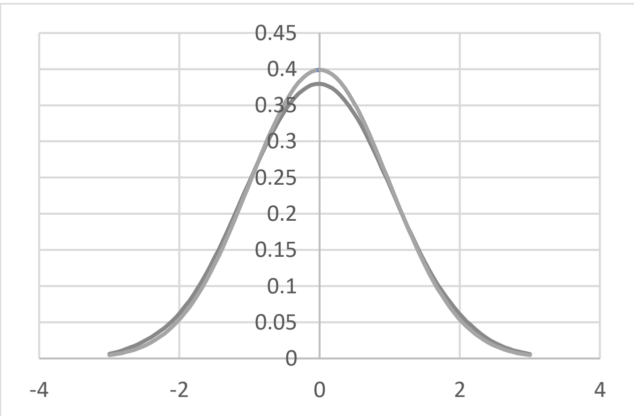


Figure 6.2. Marginal density 2 and its estimation

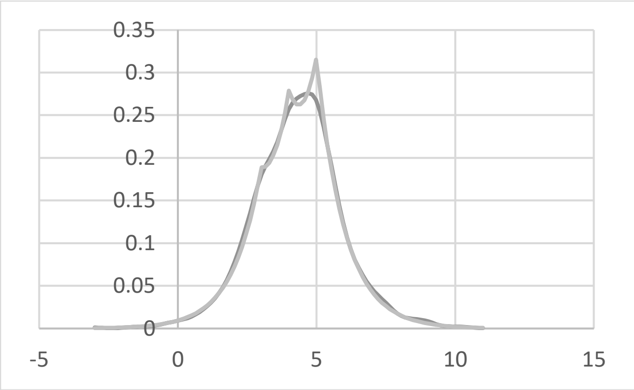


Figure 6.3. Marginal density 3 and its estimation

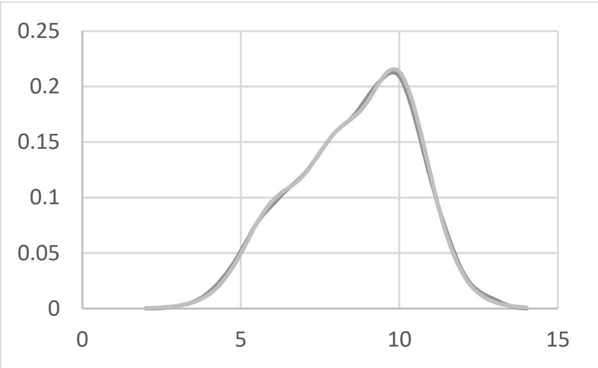


Table 6.3 shows the $\sqrt{\overline{ISE}}$ for a random vector with $p = 3$. The average ISE is obtained using 100 samples for each sample size of the joint densities of models 1 and 5.

Table 6.3. $\sqrt{\overline{ISE}}$

$\sqrt{\overline{ISE}}$	Sample size			
	500	1,000	2,000	10,0000
Model 1	0.0358	0.0333	0.0327	0.0322
Model 5	0.0218	0.0174	0.0145	0.0113

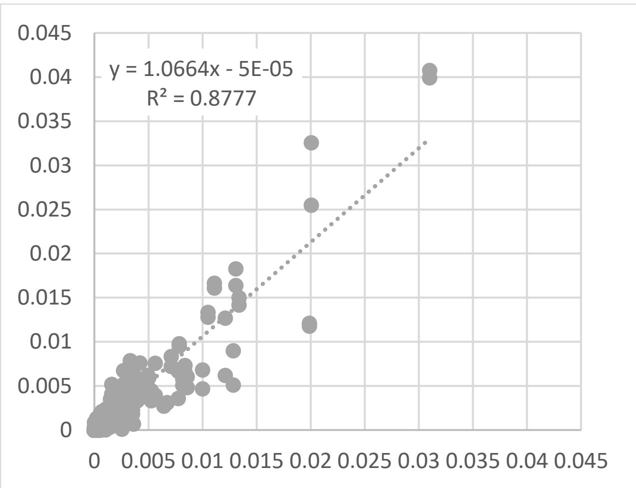
Note that the running times are shorter than those provided by Zheng and Wu (2020).

6.1 AN ALTERNATIVE TO EVALUATE THE QUALITY OF FIT OF A MULTIVARIATE DENSITY

The multivariate numerical integration can be a laborious and time-consuming task and, in some instances, makes it difficult to calculate the ISE of the joint density due to the course of dimensionality. Moreover, the ISE could be less precise when the dimension of the space increases. To avoid this, we propose the following alternative to the ISE. We apply it to evaluate the goodness of fit of the joint fitted density of model 5.

We make a scatter plot where the real values of the density are represented in the abscissa axis and in the estimations in the ordinate axis. The real values and the estimates of the density are obtained on a regular grid of points that covers the area of the three-dimensional space containing most of the probability mass. In figures 6.1, 6.2 and 6.3 this area is the parallelepiped $[-3,3] \times [-3,\ 11] \times [2,14]$. Each interval has been broken down into nine segments of equal length. In this way we obtain a grid with 1,000 point of the parallelepiped[3]. The values $\left(f(\mathbf{x}_i), \hat{f}(\mathbf{x}_i)\right)\ i = 1,2,\ldots,1{,}000$ are obtained on a grid of points of the three-dimensional space $\mathbf{x}_i = (x_i^1, x_i^2, x_i^3)\ i = 1,2,\ldots,1{,}000$, f represents the density of the model 5 and $\hat{f}$ is its estimation using a sample of size 10,000. The values $\left(f(\mathbf{x}_i), \hat{f}(\mathbf{x}_i)\right)$ are shown in the scatter plot 6.4. A perfect estimation of the density f would produce a cloud of points located on the equation $\hat{f} = f$.

Figure 6.4. Values $\left(f(\mathbf{x}_i), \hat{f}(\mathbf{x}_i)\right)$



[3] The partition in 9 segments of equal length is obtained by means of 10 equidistant points on the interval $[-3,3]$. Analogously, 10 equidistant points are obtained on the intervals $[-3,11]$ and $[2,14]$. The cartesian product of the three sets of 10 points provide a regular grid of 1,000 points on the parallelepiped $[-3,3] \times [-3,\ 11] \times [2,14]$.

The coefficient of determination of the fit should be higher as the quality of the estimates improves. The previous figure contains the points cloud and $\hat{f} = 1.0664f + 0{,}00005$. The fit is made by least squares and the coefficient of determination is equal to $R^2 = 0.8777$. The heteroskedasticity is due to bias of the estimated highest values of the density.

7. REAL DATA ILLUSTRATION

In this section, the developed approach is applied to estimate the joint density of the variables: spending, wealth and income of households, that is $z = f(spending, wealth, income)$.

As well as being illustrative, the empirical application also shows how the estimated multivariate densities provide a wider view of the economic position of households. The better the knowledge of the economic and financial situation of households, the more effective the economic policy will be. Therefore, the proposed methodology can contribute to a better design of redistributive policies, as well as having other applications.

The sample data comes from the Spanish Survey of Household Finances (EFF) for the year 2014, which was conducted by the Bank of Spain (Banco de España, 2017). The EFF provides information on assets, debt income and spending. The sample size is 6,120 households. Annual expenditure on non-durable goods (food and other) are expressed in ten thousand euros. Net wealth (gross wealth less debts) is expressed in million euros and household income is calculated as the sum of labor and no-labor incomes for all household members in 2013. It is expressed in hundred thousand euros.

Table 1 shows the 10, 50 and 90th percentiles of each variable their mean and their standard deviation (SD).

Table 7.1

Percentile	Expenditure ($\times 10^4$ €)	Net Wealth ($\times 10^6$ €)	Income ($\times 10^5$ €)	MMR pdf
10	0.6000	0.0041	0.0942	1.6101
50	1.2372	0.2389	0.3030	1.2201
90	3.6000	2.4117	0.9839	0.0111
Mean	1.7831	1.4080	0.5095	0.0516

SD	2.0124	9.2982	0.9090	

The joint density $f(spending, wealth, income)$ can be estimated on any triplet of values of them (for example, see the last column of Table 1 that shows the value of the three-dimensional density function on each triplet of the values of the variables contained in columns 2-4) but its graphical representation is not possible since it is immersed in a four-dimensional space. For this reason, we analyse the marginal densities of wealth and spending. In addition, we show the bi-dimensional conditional densities to the 10, 50 and 90 percentiles of income variable (Figures 7.2 – 7.4). The percentiles are in Table 7.1.

Figure 7.1. Wealth (y-axis) and Spending (x-axis) marginal density

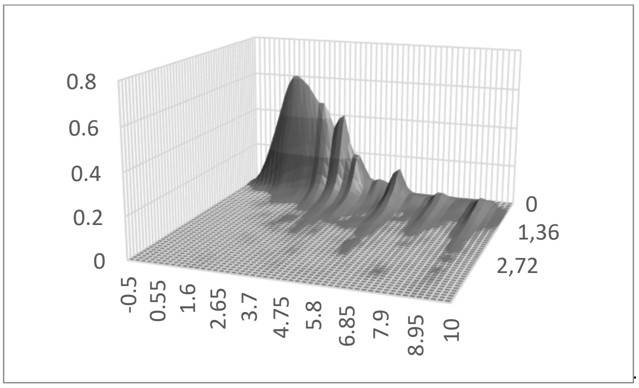


Figure 7.1 displays the marginal probability density function (z-axis: points in space lying above the origin) of the variable wealth (y -axis: points lying to the right of the origin) and spending (x-axis: points in the space lying back from the origin) without taking into account income. It is a multimodal density. The sequence of the different modes indicates a positive correlation between spending and wealth showing higher levels of spending as wealth increases.

Figure 7.2. Wealth (y-axis) and Spending (x-axis) density conditioned by the 10th income percentile: 0.0942 (9,420 euros).

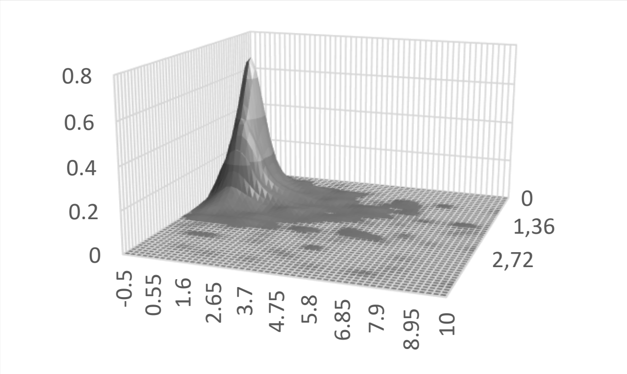
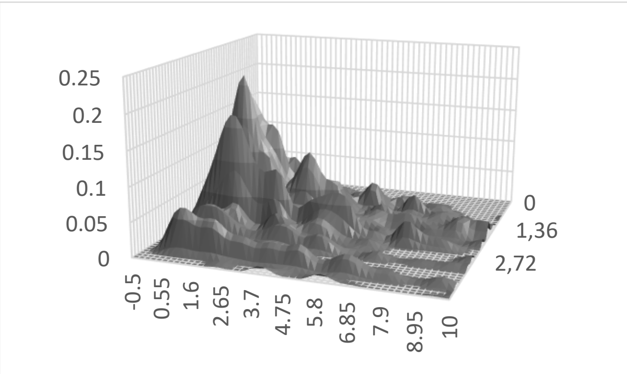


Figure 7.2 shows $\hat{f}(wealth, spending|income = 0.0942)$ (note that 0.0942 x 100,00 = 9,420 euros. It is almost unimodal, and we can observe that the higher the wealth of the household the greater its spending is.

Figure 7.3 Wealth (y-axis) and Spending (x-axis) density conditioned by the 50th income percentile: 0.3030 (x 100,000=30,030 euros).



The density shown in Figure 7.3 for the 50th income percentile presents higher dispersion for spending and wealth. Figure 7.4. Wealth (y-axis) and Spending (x-axis) density conditioned by the 90th income percentile: 0.9839 (x 100,000 = 98,390 euros)

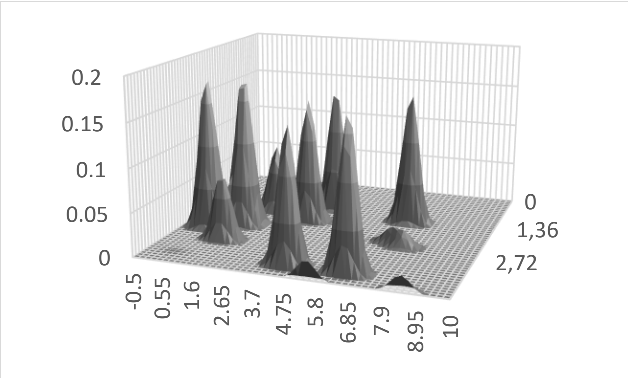


Figure 7.4 shows the same data as figure 7.3, but in a more marked way. For high incomes the wealth and spending capacity are more diverse. In addition, it should be pointed out that there are few rich households in the sample[4]. The estimated density is higher in Figure 7.4 (between 0.17 and 0.18) while in Figure 7.2 the highest peak shows a density between7.3 and 7.4.

Figure 7.5. Regression function of spending on income and wealth

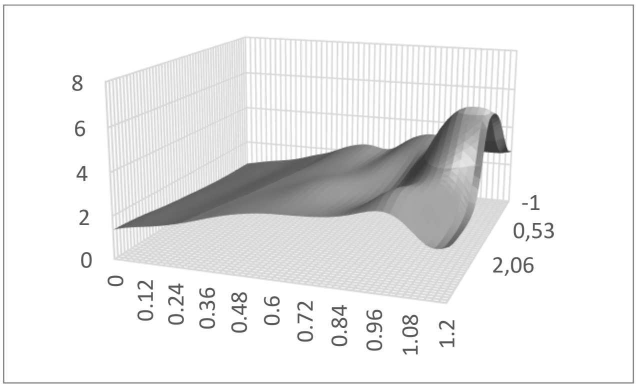


Figure 7.5 represents the regression graph of the expected spending (z-axis) on income (y-axis) and wealth (x-axis). An almost linear increase in spending can be observed when income increases to 0.432 (y-axis; 43,200 €) for any level of wealth. A smaller rise in expected spending is also observed when wealth increases at any level of income smaller than 0.432 (43,200 €). For higher incomes, a greater interaction between income and wealth is observed. When the household income is null, the annual spending goes from 7230 euros (0. 723 in z-axis) to 13,725 euros (1,3725 in z-axis) for households with higher wealth.

[4] Note that the z-axis scale is different for each figure. We have preferred to loss in comparability to improve the clarity of the figures.

Approximately, the average of both quantities corresponds to the autonomous level of consumption of the Keynes' Theory of Consumption (Keynes, 1936), that is the intercept of the linear equation that explains consumption as a function of income (without considering wealth). In a model that also considers wealth; the autonomous level of consumption grows slightly with wealth as it has been indicated previously.

8. CONCLUSIONS

This paper introduces the multivariate multiresolution approximation to probability density functions and the estimation of these approximations. A new algorithm, based on a process of data frequency count (DFC algorithm), is developed. It provides a consistent estimation of the MMR densities for big data. In the context of massive data, the quality of the estimations is almost equivalent to those obtained by the maximum likelihood. But the processing time is smaller. In contrast, the processing time to maximize of the likelihood function, by means of optimization software or applying the EM algorithm, is higher when the number of variables is greater than three and the size of the sample is too big. For instance, we need several hours to estimate by maximum likelihood the coefficients of the multivariate MR density introduced in section 7 (we are not more explicit in this regard for reasons of space and because this is not the purpose of this article). The developed algorithm provides a good estimation of the coefficients of the density in 0.03 seconds using a sample of 16,299 data. After estimating the joint density, we can obtain the marginal densities, the conditional densities as well as the regression functions applying elemental rules of probability calculus[5]. (see section 7).

To sum up, the MMR approximations of multivariate densities and the FDC algorithm are powerful statistical tools to estimate multivariate densities of continuous variables as for instance sales, time of using internet, consumption, etc. The empirical illustration shows the capacity of the proposed approach to provide a better understanding of the joint behaviour of the income, wealth and spending of the households.

[5] We will go deeper on these topics in future publications for space reasons.

REFERENCES

Banco de España (2017): Survey of Household Finances (EFF) 2014: Methods, Results and Changes since 2011. Analytical article, 24, January 2007.

Beylkin, G., Monzon, L., Satkauskas, I., 2019. On computing distributions of products of random variables via Gaussian multiresolution. Appl Comput Harmonic Anal 47(2):306–337.

Bonhomme, S., Jochmans, K., and Robin, J.M., 2016. Estimating Multivariate Latent-Structure Models. The Annals of Statistics, 44: 540–563.

Bouzebda, S. and Didi, S., 2017. Multivariate wavelet density and regression estimators for stationary and ergodic discrete time processes: Asymptotic results. Communications in Statistics-Theory and Methods, 46 (3): pp. 1367-1406.

Eckley, I.A., 2001. Wavelet methods for time series and spatial data. Thesis of University of Bristol.

Hernández, E., and Weiss, G. 1996. A first course on wavelet. CRC Press, Washington.

Hunter, D. R., and Levine, M.,2015. Semi-Parametric Estimation for Conditional Independence Multivariate Finite Mixture Models. Statistics Surveys, 9: pp. 1-31.

Kasahara, H., and Shimotsu, K., 2014. Non-Parametric Identification and Estimation of the Number of Components in Multivariate Mixtures. Journal of the Royal Statistical Society, Series B, 76: pp. 97–111

Keynes, J. M., 1936. The General Theory of Employment Interest and Money.

Macmillan, London.

Li, DW., 2017. Kernel estimations for multivariate density functional with bootstrap. Communications in Statistics- Theory and Methods, 46 (9): pp. 4631-4641.

Luini, E. and Arbenz, P.,2020. Density estimation of multivariate samples using Wasserstein distance. Journal of Statistical Computation and Simulation, 20(2): pp. 181-210

Mallat, S., 1999. A Wavelet Tour of Signal Processing. Cambridge: Academic Press.

Palacios-González, F., and García-Fernández, R.M., 2014a.A flexible family of density Functions. Statistics: A Journal of Theoretical and Applied Statistics, 49: pp. 680–704.

Palacios-González, F., and García-Fernández R.M., 2014b. Mixtures of Mixtures Based on Multiresolution Analysis Theory. Communications in Statistics, Simulation and Computation, 43: pp.723–742

Palacios-González, F., and García-Fernández R.M., 2020. A faster algorithm to estimate multiresolution densities. Computational Statistics, 35: pp. 1207-1230.

Pollard, D., 2003. A user's guide to measure theoretic probability. Cambridge University Press.

Schwarz, G. E., 1978. Estimating the dimension of a model. Annals of Statistics 6(2): pp. 461–464, DOI: 10.1080/02331888.2014.883398.

Silverman, B. W., 1986. Density Estimation for Statistics and Data Analysis, vol. 86 of Monographs on Statistics and Applied Probability. Chapman and Hall, London.

Xueying, C., and X. Minge, 2012. A Split-and-Conquer Approach for Analysis of Extraordinarily Large Data. Technical report 2012-01, Center for Discrete Mathematics and Theoretical Computer Science (DIMACS), http://dimacs.rutgers.edu/TechnicalReports/TechReports/2012/2012-01.pdf

Zheng, C.W. and Wu, Y.C., 2019. Nonparametric estimation of Multivariate mixtures. Journal of the American Statistical Association: 115 (531), pp. 1456-1471.

Wojtaszczyk, P., 1997. A mathematical introduction to wavelets. Cambridge University Press.

APPENDIX

A. PROBABILITY DENSITIES FUNCTIONS IN A MULTIRESOLUTION ANALYSIS STRUCTURE

In this section we review the multivariate MRA structures defined by the Multivariate Scaling Function $\Theta(\mathbf{x}) = \prod_{i=1}^{q} \theta(x_i)$, where θ is a function named cubic box spline.

A.1.DILATIONS AND TRANSLATION OF A MULTIVARIATE SCALING FUNCTION

Given a vector $\mathbf{k} = (k_1, k_2, \dots, k_q) \in Z^q$, an integer value $j \in Z$ and a multivariate scaling function Θ, dilation and translation of Θ are defined as follows: $\Theta_{j,\mathbf{k}}(\mathbf{x}) = \prod_{i=1}^{q} \theta_{j,k_i}(x_i) = \prod_{i=1}^{q} 2^{j/2}\theta(2^j x_i - k_i)$. Similar to the unidimensional case (see for instance Palacios-González and García-Fernández, 2020) we assign to each function $\Theta_{j,\mathbf{k}}$ a point of R^q that is the translation center of Θ. This point is: $\frac{\mathbf{k}}{2^j} = \left(\frac{k_1}{2^j}, \frac{k_2}{2^j}, \dots, \frac{k_q}{2^j}\right)$.

The function $\Theta_{j,\mathbf{k}}$ is symmetrical (in R^{q+1}) around the vertical axis that rises on such point. It takes non negative values and its support is $\prod_{i=1}^{q}\left[\frac{k_i-2}{2^j},\frac{k_i+2}{2^j}\right]$.

Moreover,

$$\int_{R^q} \Theta_{j,\mathbf{k}}(\mathbf{x})d\mathbf{x} = 2^{-qj/2} \qquad (A.1.1)$$

Note that the set $\left\{\left(\frac{\mathbf{k}}{2^j}\right)\right\}_{\mathbf{k}\in Z^q} = \left\{\left(\frac{k_1}{2^j}, \quad \frac{k_2}{2^j}, \quad \dots, \quad \frac{k_q}{2^j}\right)\right\}_{\mathbf{k}\in Z^q}$ define a regular grid of points on R^q which is thicker the higher the level of resolution is. Moreover, the grid defined for a level of resolution j is always contained within a grid with a higher level of resolution.

A.2 A MRA STRUCTURE GENERATES ON R^q BY A MULTIVARIATE SCALING FUNCTION Θ.

Any multivariate scaling function defines a MRA structure on R^q. The scaling function is used to build a sequence $\{V_j\}_{j\in Z}$ of functional linear spaces $g: R^q \to R$ that verify that $V_j \subset V_{j+1} \quad \forall j \in Z.$

Each linear space V_j is generated by the Riesz basis (p. 44 in Hernández and Weiss (1996)) $\{\Theta_{j,\mathbf{k}}\}_{\mathbf{k}\in Z^q}$. That is, V_j is made up by functions with the following form:

$$f(\mathbf{x}) = \sum_{\mathbf{k}\in Z^q} b_{\mathbf{k}}\Theta_{j,\mathbf{k}}(\mathbf{x}) \qquad \forall \mathbf{x} \in R^q \qquad (A.2.1)$$

where $b_{\mathbf{k}} \in R \ \forall \ \mathbf{k} \in Z^{\boldsymbol{q}}$**.**

This is the usual way in which functions are defined within the MRA structure. However, in the present work, two small notation changes will be made to facilitate the handling of the probability density functions. If we call

$a_{\mathbf{k}} = 2^{qj/2} b_{\mathbf{k}}$

Then, we can rewrite the function f defined in (A.2.1) as follows:

$f(\mathbf{x}) = \frac{1}{2^{qj/2}}\sum_{\mathbf{k}\in Z^q} a_{\mathbf{k}}\Theta_{j,\mathbf{k}}(\mathbf{x})$ (A.2.3)

Writing $c_{\mathbf{k}} = \frac{a_{\mathbf{k}}}{2^{qj}}$ and $\Lambda_{j,k(x)} = \prod_{i=1}^{q} 2^j\theta(2^j x_i - k_i)$ we obtain a third expression equivalent to (A.2.1) and to (A.2.3). That is, $f(\mathbf{x}) = \sum_{\mathbf{k}\in Z^q} c_{\mathbf{k}}\Lambda_{j,\mathbf{k}}(\mathbf{x})$. The latter expression (see 4) is the most used and it facilitates the proof of theorem 4.2. Expression (A.2.3) facilitates the proof of theorem enunciated in 3.5.

A.3.THEOREM

Given a level of resolution j, any function $f(\mathbf{x})$, obtained by means of (A.2.3) using a set of values $a_{\mathbf{k}}$ such that

$a_{\mathbf{k}} \geq 0 \; \mathbf{k} \in Z^q$ (A.3.1)

and

$$\sum_{\mathbf{k}\in Z^q} a_{\mathbf{k}} = 2^{qj} \qquad (A.3.2)$$

is a multivariate probability density function in the V_j space of the MRA structure.

Proof. On the one hand, it is evident that if (A.3.1) is true, then $f(\mathbf{x}) = \frac{1}{2^{qj/2}} \sum a_{\mathbf{k}} \Theta_{j,\mathbf{k}}(\mathbf{x}) \geq 0 \quad \forall \mathbf{x} \in R^q$. On the other hand, given that $\forall \mathbf{x} \in R^q$ the series $\sum a_{\mathbf{k}} \Theta_{j,\mathbf{k}}(\mathbf{x})$ is absolutely convergent, we have that

$$\int_{R^q} f(\mathbf{x}) d\mathbf{x} = \frac{1}{2^{qj/2}} \sum_{\mathbf{k}\in Z^q} a_{\mathbf{k}} \int_{R^q} \Theta_{j,\mathbf{k}}(\mathbf{x}) d\mathbf{x}\,.$$

Considering (A.1.1) we can state that $\int_{R^q} f(\mathbf{x}) d\mathbf{x} = \frac{1}{2^{qj}} \sum_{\mathbf{k}\in Z^q} a_{\mathbf{k}}$. So, if (A.3.2) is true, then $\int_{R^q} f(\mathbf{x}) d\mathbf{x} = 1$ and consequently f is a true density function.

A.3.1. COLLORARY

Any function (A.2.3) written as in (4) whose coefficients $c_{\mathbf{k}}$ satisfy the conditions:

$c_{\mathbf{k}} \geq 0 \;\; \forall \mathbf{k} \in Z^q$ and $\sum_{\mathbf{k}\in Z^q} c_{\mathbf{k}} = 1$ is a probability density function.

B. APPROXIMATION OF ONE DENSITY BY ANOTHER OF THE MULTIVARIATE STRUCTURE

Given $\mathbf{x} = (x_1, x_2, \dots, x_q) R^q$ the vectors $\mathbf{k}(\mathbf{x}) = (k(x_1), k(x_2), \dots, k(x_q))$ and $\mathbf{r}(\mathbf{x}) = (r(x_1), r(x_2), \dots, r(x_q))$ are defined using (9). It is verified (see (9) y (10)) that $\mathbf{r}(\mathbf{x}) = 2^j \mathbf{x} - \mathbf{k}(\mathbf{x})$ and $\mathbf{x} = \frac{\mathbf{k}(\mathbf{x}) + \mathbf{r}(\mathbf{x})}{2^j}$. We assume that the operations above indicated are the sum and the difference of two vectors and the product of a scalar by a vector.

Let us use the following notation: $\mathbf{l} = (l_1, l_2, \dots, l_q) \in Z^q$, $C_1 = \prod_{i=1}^{q} [0,1)$, $C_2 = Z^q \cap \prod_{i=1}^{q} [-2,2)$, $C_2^0 = Z^q \cap \prod_{i=1}^{q} (-2,2)$, $C_{\mathbf{x}}^j = \prod_{i=1}^{q} \left[\frac{k(x_i)}{2^j}, \frac{k(x_i)+1}{2^j}\right) \quad \forall \mathbf{x} \in R^q$ and $D_{\mathbf{x}}^j = \prod_{i=1}^{q} \left(\frac{k_i - 0.5}{2^j}, \frac{k_i + 0.5}{2^j}\right] \quad \forall \mathbf{k} \in Z^q$.

B.1.LEMMA

For all $\mathbf{x} = (x_1, x_2, \dots, x_q) \in R^q$ it is verified that $\frac{k(x_i)}{2^j} \leq x_i < \frac{k(x_i)+1}{2^j} \quad \forall i = 1,2,\dots,q$ and $0 \leq r(x_i) < 1 \quad \forall i = 1,2,\dots,q$ (see Palacios-González and García-Fernández, 2020). This allows us to affirm that $\forall j \in Z$ and $\forall \mathbf{x} \in R^q$, $\mathbf{x} \in C_{\mathbf{x}}^j$ and $\mathbf{r}(\mathbf{x}) \in C_1$.

B.1.1. COROLLARY

For all $\mathbf{x} = (x_1, x_2, \dots, x_q) \in R^q$ it is verified that $\lim_{j\to\infty} \frac{k(x_i)}{2^j} = x_i \quad \forall i = 1,2,\dots,q$ (Palacios-González and García-Fernández, 2020). Hence, we can affirm that $\lim_{j\to\infty} \frac{\mathbf{k}(\mathbf{x})}{2^j} = \mathbf{x}$. As a consequence, we have the following corollary.

B.1.2. COROLLARY

For all $\mathbf{x} = (x_1, x_2, \dots, x_q) \in R^q$ it is verified that $\lim_{j\to\infty} r(x_i) = 0 \quad \forall\, i = 1,2,\dots,q$ (see Palacios-González and García-Fernández, 2020). As a consequence:

$\lim_{j\to\infty} \mathbf{r}(\mathbf{x}) = (0,0,\dots,0)$ (B.1.1)

B.2. LEMMA

For all $\mathbf{x} = (x_1, x_2, \dots, x_q) \in R^q$ and all $\mathbf{l} = (l_1, l_2, \dots, l_q) \in Z^q$ it is verified $\theta_{j,k(x_i)-l_i}(x) = 2^{j/2}\theta(r(x_i) + l_i) \quad \forall j, l_i \in Z \text{ and } \forall i = 1,2,..,q$ (see Palacios-González and García-Fernández, 2020). This allows us to write $\Theta_{j,\mathbf{k}-\mathbf{l}}(\mathbf{x}) = 2^{qj/2}\Theta(\mathbf{r}(\mathbf{x}) + \mathbf{l})$.

B.2.1. COROLLARY

For all $x \in R$ and all $l \in Z$, it is verified that if $r(x) > 0$ then $\theta_{j,k(x)-l}(x) \neq 0 \Leftrightarrow -2 \leq l \leq 1$ and if $r(x) = 0$ then, $\theta_{j,k(x)-l}(x) \neq 0 \Leftrightarrow -1 \leq l \leq 1$ (see Palacios-González and García-Fernández, 2020). As a consequence:

$\Theta_{j,\mathbf{k}-\mathbf{l}}(\mathbf{x}) \neq 0 \Leftrightarrow \mathbf{l} \in C_2$ (B.1.2)

That is, if and only if $-2 \leq l_i \leq 1 \quad \forall i = 1,2,\dots,q$. In the particular case in which $\mathbf{r}(\mathbf{x}) = \mathbf{0}$ we can state that $\Theta_{j,\mathbf{k}-\mathbf{l}}(\mathbf{x}) \neq 0 \Leftrightarrow \mathbf{l} \in C_2^0$. That is, if and only if $-1 \leq l_i \leq 1 \quad \forall i = 1,2,\dots,q$.

B.3.LEMMA

It is verified $\sum_{l=-2}^{1} \theta(l + r) = 1 \quad \forall r \mid 0 \leq r < 1$ (see Palacios-González and García-Fernández, 2020). As a consequence, for all $\mathbf{r} \in C_1$ we can state:

$$\sum_{\mathbf{l} \in C_2} \Theta(1 + \mathbf{r}) = 1 \quad (\text{B}.3.1)$$

Proof.

$$\sum_{\mathbf{l}\in C_2}\Theta(1+\mathbf{r}) = \sum_{l_1=-2}^{1}\sum_{l_2=-2}^{1}\dots\sum_{l_q=-2}^{1}\theta(1+r_1)\theta(1+r_2)\dots\theta\left(1+r_q\right) =$$

$$\left(\sum_{l_1=-2}^{1}\theta(1+r_1)\right)\left(\sum_{l_2=-2}^{1}\theta(1+r_2)\right)\dots\left(\sum_{l_q=-2}^{1}\theta\left(1+r_q\right)\right) = 1\times 1\times\dots\times 1 = 1$$

B.4.PROPOSITION

Let $f(\mathbf{x})$ be a function as that defined in (A.2.3). For any level of resolution j, for all $\boldsymbol{x}\in R^q$, it is held that

$$f(\mathbf{x}) = \sum_{\mathbf{l}\in C_2} a_{\mathbf{k}(\mathbf{x})-\mathbf{l}}\Theta(\mathbf{r}(\mathbf{x})+\mathbf{l}) \qquad \text{(B. 4.1)}$$

If $\boldsymbol{r}(\boldsymbol{x}) = \mathbf{0}$ we can write (B.4.1) more accurately as $f(\mathbf{x}) = \sum_{\mathbf{l}\in C_2^0} a_{\mathbf{k}(\mathbf{x})-\mathbf{l}}\Theta(\mathbf{l})$.

Proof. It is evident that making a change of origin to the point $\mathbf{k}(\mathbf{x})$ in Z^q, the function $f(\mathbf{x})$,(A.2.3), can be rewritten as:

$$f(\mathbf{x}) = \frac{1}{2^{qj/2}}\sum_{\mathbf{l}\in Z^q} a_{\mathbf{k}(\mathbf{x})-\mathbf{l}}\Theta_{j,\mathbf{k}(\mathbf{x})-\mathbf{l}}(\mathbf{x}) \qquad \text{(B. 4.2)}$$

Taking into consideration lemma B.2 and expression (B.1.2) the proposition is true. This means that locally, on each real value $\mathbf{x}$, the function $f(\mathbf{x})$ will be calculated by a linear convex combination with a finite number of addends between 3^q and 4^q, depending on the number of non-null components of vector $\mathbf{r}(\mathbf{x})$.

Note that the function (B.4.2) can be rewritten as follows:

$$f(\mathbf{x}) = \sum_{\mathbf{l}\in C_2} c_{\mathbf{k}(\mathbf{x})-\mathbf{l}}\Lambda_{j,\mathbf{k}(\mathbf{x})-\mathbf{l}}(\mathbf{x}) \qquad \text{(B. 4.3)}$$

B.5.THEOREM

Let $\mathbf{X}$ be a continuous random vector. Let $a_{\mathbf{k}}^{j} = 2^{qj}P(\mathbf{X}\in D_{\mathbf{k}}^{\boldsymbol{j}})$ and let $\left\{f_j\right\}_{j\in Z}$ be a sequence of functions generated according (A.2.3). It is verified that

$\lim_{j\to\infty} f_j(\mathbf{x}) = f(\mathbf{x})\ \forall\mathbf{x}\in R$ (B.5.1)

Proof. Taking into account corollary B.1.1 and considering that the volume of $D_{\mathbf{k}}^{j}$ is $Vol(D_{\mathbf{k}}^{j}) = \frac{1}{2^{qj}}$ for all $\mathbf{k} \in Z^q$, we have that $\lim_{j\to\infty} a_{\mathbf{k}}^{j} = \lim_{j\to\infty} \frac{P(\mathbf{X}\in D_{\mathbf{k}}^{j})}{Vol(D_{\mathbf{k}}^{j})}$.

Expression $\frac{P(\mathbf{X}\in D_{\mathbf{k}}^{j})}{Vol(D_{\mathbf{k}}^{j})}$ represents the mean density by volume unit in the hypercube $D_{\mathbf{k}}^{j}$. If $j \to \infty$ then, the volume of $D_{\mathbf{k}}^{j}$ converges to zero, the hypercube converges to the point $\mathbf{x}$ and the quotient is the probability density function of $\mathbf{x}$. This allows us to affirm that $\forall \mathbf{x} \in R^q$ $\lim_{j\to\infty} a_{\mathbf{k}}^{j} = \lim_{j\to\infty} \frac{P(\mathbf{X}\in D_{\mathbf{k}}^{j})}{Vol(D_{\mathbf{k}}^{j})} = f(\mathbf{x})$. Considering that (see corollary B.1.1) $\lim_{j\to\infty} \frac{\mathbf{k}(\mathbf{x})-\mathbf{l}}{2^j} = \lim_{j\to\infty} \frac{\mathbf{k}(\mathbf{x})}{2^j} = \mathbf{x}$ $\forall \mathbf{x} \in R^q$ $\forall \mathbf{l} \in Z^q$, we can state that $\lim_{j\to\infty} a_{\mathbf{k}(\mathbf{x})-\mathbf{l}}^{j} = f(\mathbf{x})$ and according to proposition B.4:

$$\lim_{j\to\infty} f_j(\mathbf{x}) = \sum_{\mathbf{l}\in C_2} \lim_{j\to\infty} a_{\mathbf{k}(\mathbf{x})-\mathbf{l}}^{j} \lim_{j\to\infty} \Theta(\mathbf{r}(\mathbf{x})+\mathbf{l}) = f(\mathbf{x}) \sum_{\mathbf{l}\in C_2} \lim_{j\to\infty} \Theta(\mathbf{r}(\mathbf{x})+\mathbf{i}) \quad (B.5.2)$$

Given that Θ is a continuous function $\lim_{j\to\infty} \Theta(\mathbf{r}(\mathbf{x})+\mathbf{l}) = \Theta\left(\lim_{j\to\infty} \mathbf{r}(\mathbf{x})+\mathbf{l}\right) = \Theta(\mathbf{l})$. Hence, considering (B.3.1) we have that $\sum_{\mathbf{l}\in C_2} \lim_{j\to\infty} \Theta(\mathbf{r}(\mathbf{x})+\mathbf{l}) = \sum_{\mathbf{l}\in C_2^0} \Theta(\mathbf{l}) = 1$. Finally, substituting in (B.5.2), we obtain expression (B.5.1).

B.5.1. COROLLARY

Remember that if $c_{\mathbf{k}}^{j} = \frac{a_{\mathbf{k}}^{j}}{2^{qj}} = P(\mathbf{X} \in D_{\mathbf{k}}^{j})$ then the functions $f_j(\mathbf{x})$ can be rewritten as: $f_j(\mathbf{x}) = \sum_{\mathbf{k}\in Z^q} c_{\mathbf{k}}^{j} \Lambda_{j,\mathbf{k}}(x) = \sum_{\mathbf{l}\in C_2} c_{\mathbf{k}}^{j} \Lambda_{j,\mathbf{k}(\mathbf{x})-\mathbf{l}}(\mathbf{x})$ which is the form used in (6). It is obvious that with this equivalent notation the theorem remains true. That is, $\lim_{j\to\infty} f_j(\mathbf{x}) = f(\mathbf{x})$ $\forall\, \mathbf{x} \in R$. For this reason, it is stated in 3.4 that the densities f_j are approximations of the density f.

C. CONSISTENT ESTIMATION OF A MULTIVARIATE DENSITY

Let $\mathbf{X}$ be a q dimensional random vector with a finite vector of means and finite variance covariance matrix. Let $A = \left(\mathbf{X} \in \prod_{i=1}^{q}(c_i, d_i]\right)$ be an event where $c_i, d_i \in R$ and where $c_i < d_i$ for all $i = 1,2,\ldots,q$. Let $P(A) = P\left(\mathbf{X} \in \prod_{i=1}^{q}(c_i, d_i]\right)$ be the probability of A.

For the sample $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ of $\mathbf{X}$, let $f_r^n(A) = \frac{\#(\mathbf{x}_t\ t=1,2,\ldots,n \mid\ \mathbf{x}_t \in \prod_{i=1}^{p}(c_i, d_i])}{n}$ be the relative frequency of event A. It has been obtained by means of the corresponding count of frequencies in the sample. Given a sequence of samples of increasing size n, from a random variable X, it is verified that $\lim_{n\to\infty} P(|f_n^r(A) - P(A)| > \epsilon) = 0 \;\; \forall\epsilon > 0$. This result is known as the Bernouilli law of big numbers (see Pollard, 2003). It states that the relative frequency of any event converges in probability to the probability of such event. It is usually written as $f_r^n(A) \xrightarrow{p} P(A)$.

C.1.LEMMA

Taking into account the definitions of $c_{\mathbf{k}}^{j}$ (see 5) and $\hat{c}_{\mathbf{k}}^{n,j}$ (see 7) and the Bernouilli law of big numbers, we can affirm that for any level of resolution $j \in Z$ (see 3.2 and 4) it is verified that $\hat{c}_{\mathbf{k}}^{n,j} \xrightarrow{p} c_{\mathbf{k}}^{j}$.

C.1.1. COROLLARY

For all $j \in Z$, $\hat{f}_{n,j}(\mathbf{x})$ is a consistent estimator of $f_j(\mathbf{x})$. That is, it is verified

$\hat{f}_{n,j}(\mathbf{x}) \xrightarrow{p} f_j(\mathbf{x}) \quad \forall\mathbf{x} \in R^q$.

Proof. As a consequence of Lemma C.1 and expression B.4.3, on each point $\mathbf{x} \in R^q$:

$$\hat{f}_{n,j}(\mathbf{x}) = \sum_{\mathbf{l}\in C_2} \hat{c}_{\mathbf{k}(\mathbf{x})-\mathbf{l}}^{n,j} \Lambda_{j,\mathbf{k}(\mathbf{x})-\mathbf{l}}(\mathbf{x}) \quad \xrightarrow{p} \quad \sum_{\mathbf{l}\in C_2} c_{\mathbf{k}}^{j} \Lambda_{j,\mathbf{k}(\mathbf{x})-\mathbf{l}}(\mathbf{x}) = f_j(\mathbf{x})$$

Theorem B.5 and Corollary C.1.1 lead directly to enunciate the following theorem of consistent approximation of any continuous multivariate density function with finite vector of means and finite variance covariance matrix.

C.2. THEOREM

For all $\mathbf{x} \in R^q$ it is verified that

$\lim_{j\to\infty} \hat{f}_{n,j}(\mathbf{x}) \xrightarrow{p} \lim_{j\to\infty} f_j(\mathbf{x}) = f(\mathbf{x})$.

APPENDIX D

Proof of Lemma 5.2

It is evident (see Palacios-González and García-Fernández, 2020), that for all $x_i \in R$ it is verified that if $r(x_i) \leq 0.5$ then $x_i = \frac{k(x_i)+r(x_i)}{2^j} \leq \frac{k(x_i)+0.5}{2^j}$ and hence [6] $x \in \left(\frac{k(x_i)-0.5}{2^j}, \frac{k(x_i)+0.5}{2^j}\right]$, which is the interval centered in $\frac{k(x_i)}{2^j}$. If $r(x_i) > 0.5$ then $x_i =$

[6] Note that $0 \leq r(x_i) < 1$

$\frac{k(x_i)+r(x_i)}{2^j} > \frac{k(x_i)+0.5}{2^j} = \frac{(k(x_i)+1)-0.5}{2^j}$ and consequently, $x \in \left(\frac{(k(x_i)+1)-0.5}{2^j}, \frac{(k(x_i)+1)+0.5}{2^j}\right]$ which is the interval centered in $\frac{k(x_i)+1}{2^j}$. As a consequence, if

$$k_i = \begin{cases} k(x_i) & if\ r(x_i) \leq 0.5 \\ k(x_i)+1 & if\ r(x_i) > 0.5 \end{cases}$$

it is obvious that each component x_i of vector **x** belongs to the interval $\left(\frac{k_i-0.5}{2^j}, \frac{k_i+0.5}{2^j}\right]$. Hence,

$\mathbf{x} = (x_1, x_2, \ldots x_q) \in \prod_{i=1}^{q} \left(\frac{k_i-0.5}{2^j}, \frac{k_i+0.5}{2^j}\right] = D_{\mathbf{k}}^j$.