DOCTOR OF SCIENCES

QAMML: probability distributions for functional data

Cuvelier, Etienne

Award date:
2009

Awarding institution:
University of Namur

Facultés Universitaires Notre-Dame de la Paix
Faculté d'Informatique

Thèse présentée en vue de l'obtention du grade
de docteur en SCIENCES

QAMML: Probability Distributions For Functional Data

Etienne Cuvelier

Soutenance publique le vendredi 25 septembre 2009.
Jury:

Jean-Marie Jacquet (Président)
Monique Noirhomme-Fraiture (Promoteur)
Ingrid Van Keilegom
Edwin Diday
Jean-Paul Leclercq
Marcel Rémon

Cette thèse est dédiée à la mémoire de tous ceux que j'aimais et qui se sont arrêtés sur le bord du chemin de la vie : ma maman, mon beau-frère Olivier, mon ami Sylvain, ma soeur Laurence, ma grand-mère, mon grand-père, mon beau-frère Benoît...

Dedication To
The Inhabitants of SPACE IN GENERAL
This Work is Dedicated
By a Humble Native of Flatland
In the Hope that
Even as he was Initiated into the Mysteries
Of THREE Dimensions
Having been previously conversant
With ONLY TWO
So the Citizens of that Celestial Region
May aspire yet higher and higher
To the Secrets of FOUR FIVE OR EVEN SIX Dimensions
Thereby contributing
To the Enlargment of THE IMAGINATION
And the possible Development
Of that most and excellent Gift of MODESTY
Among the Superior Races
Of SOLID HUMANITY

Flatland, A Romance of Many Dimensions [Abbott, 1884]

Remerciements

Je tiens tout d'abord à remercier Monique Noirhomme-Fraiture qui, en m'engageant en tant qu'assistant, et ce malgré mon parcours atypique, m'a permis de réaliser mon rêve : faire de la recherche universitaire et couronner celle-ci par une thèse de doctorat. Elle m'a guidé et conseillé précieusement pendant toutes ces années, et ce, tout en me laissant une très grande liberté scientifique. Je l'en remercie vivement encore une fois.

Je remercie aussi les Facultés Universitaires Notre-Dame de la Paix, et plus particulièrement la Faculté d'Informatique, pour m'avoir donné tout le temps nécessaire à l'aboutissement de ces recherches.

Je tiens également à remercier Edwin Diday, car, non seulement le point de départ de cette thèse provient de ses travaux, mais c'est lui qui, lors d'une discussion à CSDA 2005, m'a suggéré de travailler avec une infinité de dimensions. Cette idée, m'a paru tout à fait folle, mais est certainement un tournant majeur dans l'orientation de mes recherches.

Je remercie aussi Jean-Marie Jacquet, Ingrid Van Keilegom, Jean-Paul Leclercq, Marcel Rémon, qui, avec Edwin Diday, ont accepté de faire partie de mon jury de thèse, leurs lectures attentives furent précieuses.

Je remercie ma jeune collègue, Marie, pour avoir partagé stoïquement son bureau, lors de sa première année aux Facs, avec un brontosaure comme moi, en fin de thèse, de surcroît.

Enfin mes remerciements les plus intenses vont vers celle sans qui je n'aurais pas pu réaliser le rêve fou de la recherche scientifique, mon épouse Sylvie. Aucune des lignes qui suivent n'auraient d'existence sans elle...

Un merci affectueux aussi, à mes enfants, Baptiste et Mathilde, qui ont quelques fois été, les victimes involontaires de mon stress, ainsi qu'à mes beaux-parents pour leur soutien discret, mais précieux.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

> Recherche scientifique: la seule forme de poésie qui soit rétribuée par l'État.
>
> Jean Rostand

The permanent growing of storage capacities and the development of live streaming of informations will increase the need for mathematical tools to modelize functional data. Actually there is virtually no limits for the storage capacities: scalable architectures of storage already allow to store Petabytes (million gigabytes) of information. And new technical capabilities induce new behaviors: these storage capabilities have allowed the development of automatic data collection. In the data analysis context, this evolution leads to meet data tables increasing in the two dimensions: the number of records in the table is growing, and dimensionality of tables follows also this evolution. The development of high dimensional data analysis and data mining algorithms for large databases, illustrate the natural evolving of research interests influenced by this evolution. But, even if the efficiency of algorithms can be improved more and more, there is limits to increasing efficiency, and the order of the computational complexity of an algorithm treating N records, can not be less then $O(N)$, because it must read each record at least once. But in most of the time, the lower bound for the computational complexity is higher than this ideal limit. Thus, a supervised classification of N objects with m attributes, by a simple decision tree has a computational complexity of $O(m \times N \times \log N)$ [Han and Kamber, 2006], and an unsupervised classification of N objects, with m dimensions, in K groups with the EM algorithm has a complexity of $O(m^2 \times N \times K)$ [Hand et al., 2001]. Of course these two algorithms are classical ones, and we can suppose that more efficient methods can decrease that complexity, but it is more than likely that the best lower bound (unknown nowadays) will be reached asymptotically.

Another path of research is to search how to define summaries of data

which minimize the loss of information. This summarizing process can be done using a natural grouping involved by the values of qualitative variables, or as the result of a classification process. Thus, to summarizing at best each group or concept underlying in a set of data, for a given categorical variable, the found values can be aggregated in one new type of variable: a modal variable, which contains the categorical values associated with their frequencies. To summarize a numerical variable, one solution is to retain its range of variation into the group, and create an interval type of variable. Then to summarize at best underlying concepts of a set of data, we use new kind of variables, called *symbolic data*, which can be: numerical variables, categorical single valued variables, categorical multivalued variables, interval variables, modal variables. Creating symbolic data, summarizing existing dataset is not enough, because the constant motion of collecting information leads to create more and more symbolic data, then it is necessary to develop data analysis techniques for symbolic data: that is the purpose of symbolic data analysis. The main data analysis techniques were adapted to symbolic data, and the results can be found in [Bock and Diday, 2000] and [Diday and Noirhomme-Fraiture, 2008].

However, in case of summarizing a numerical value for a given concept, there is a more efficient solution to manage a lesser loss of information instead of using interval data: retain the probability distribution of the data. The cumulative distribution function is the more effective and complete summary of the values found in a group or concept. Create such a summary of the values found for a given group or concept can be made easily, but then, there is a need to develop/adapt data analysis tools for this new kind of symbolic data: the continuous probability functions. But, if continuous probability distributions can be included in the list of symbolic data, they are also functional data. Indeed continuous probability distributions are functional data with specific properties, but the methods developed for any functional data will be applicable to this specific kind of symbolic data. Thus, the research for new data analysis techniques for probability distributions will always take benefit to be done also in the functional data analysis framework.

Functional data analysis is the part of data analysis which is interested specifically by the functional data. Functional data arise naturally in the study of many phenomenons, and are any continuous phenomenon which can be measured for any values of a varying parameter t, this parameter t can be the time, but not in all the cases. Examples: measures of the heights of a child over its childhood, the recording of any reproducible sound (phoneme, musical note, ...), the workload of applications server during the day, measures of a physical phenomenon varying for a given parameter, ... The study of functional data could become mainstream due to the increasing interest for the data analysis of streaming data, which become ubiquitous in our networked world. That is why we have decided during this work to

extend our research to functional data.

But a specificity of functions is that they belong to an infinite dimensional space, which increases the difficulty to define certain concepts, like the probability distribution for functional data. However probability distributions would be valuable tools, because they can be seen as the Swiss army knifes of data analysis: they are used in many procedure, unsupervised classification by mixture decomposition, bayesian supervised classification, regression functions, statistical inference,... Then, there were a strong interest to develop probability distributions for functional data.

This is the cornerstone of this thesis: the definition, the construction and the use in the data analysis framework, of probability distributions directly defined in the infinite dimensional space of functional data.

The seminal works of this thesis were made in the symbolic data analysis framework for the classification of probability distributions using mixture decomposition, because in addition to perform an unsupervised classification, the mixture decomposition classification, requires to use a probability distribution for this kind of symbolic and functional data. In [Diday, 2001] and [Diday, 2002] Edwin Diday has defined a finite dimensional approximation of these special functional distributions: the point joint distributions of distributions, using the copulas. In [Vrac, 2002] and [Diday and Vrac, 2005], point joint distributions of distributions built with the Frank copula of dimension two, were used in the mixture decomposition classification of climatological data. In [Cuvelier and Noirhomme-Fraiture, 2003] we have proposed to use the Clayton copulas of dimension higher than two, and we have used it in [Cuvelier and Noirhomme-Fraiture, 2005] and [Cuvelier and Noirhomme-Fraiture, 2008a]. Then we decided to focus our investigations in avoiding any approximation trying to define a probability distribution in the functional space of probability functions, and we have extended it in the whole functional space [Cuvelier and Noirhomme-Fraiture, 2006] and [Cuvelier and Noirhomme-Fraiture, 2007] in defining the QAMML distributions and the Gâteaux density. And, in [Cuvelier and Noirhomme-Fraiture, 2008b] we have extended the QAMML to generalized QAMML distributions.

Now let us give the whole structure of this work. In the chapter 2 we briefly introduce the two mains parts of the data analysis where we have met functional data: the Symbolic Data Analysis and the Functional Data Analysis.

In the chapter 3, we recall the probabilistic framework for infinite dimensional spaces. We start with the very basics of probability and real random variables in finite dimensional spaces. Then, our first original contribution is to define clearly the concept of probability distribution for functional random variable using the non strict piecewise order on functions. The two following sections explain the classical construction to define probability distribution

in infinite dimensional spaces using projections into finite dimensional spaces with separable functions and the Daniell-Kolmogorov's extension theorem.

In the chapter 4 we consider the attempts to build the probability distribution for functional data defined in the preceding chapter with finite dimensional distributions. For this, our second original contribution is to define the surfaces of margins and densities, which gives any margin of a given functional random variable for any t in the domain of this random variable. These concept were formerly defined in the symbolic data analysis framework for the probability distributions: surfaces of distributions of distributions and surfaces of densities of distributions [Vrac, 2002], we simply extend it in the whole functional case. We define also the associated functional quantile. Then we recall the concept of copulas which permits the decomposition any finite dimensional distribution in two parts: the margins in one hand, and the dependence structure (the copula) in the other hand. We recall also the family of Archimedean copulas, very well adapted to capture the dependence structure for finite dimensional distributions of discretized functional random variables. In this field, another original contribution is to give formulas for the joint density, valid for any finite dimension, of the following Archimedean families: Clayton, Frank, Gumbel-Hougaard and Joe. Then we recall the two steps parameter estimation when using joint distributions build upon copulas. We conclude this chapter with an original contribution, showing that in almost cases, the attempts to build the probability distribution for functional data as limit of finite dimensional distributions, almost always leads to failure.

The chapter 5 contains the core of our contributions. Firstly, we show that any discretized version of a functional data defines a piecewise function, build upon functional quantiles defined in the chapter 3.6: the quantilized function. Then, due to the impossibility to define a probability distribution for functional data as the limit of classical finite dimensional distributions, we use a classical procedure in mathematics when there is no solution to a specific problem in a given framework: we define a new framework designed to give a solution to this problem. We call it fractal distributions for functional quantiles. After this, we introduce the existing tools to build this new kind of distributions: aggregation operations and quasi-arithmetic means. We are then able to define the Quasi-Arithmetic Mean of Margins (QAMM) distributions which are fractal distributions for quantilized functions. And the limit of this new kind of distributions leads us directly to the Quasi-Arithmetic Mean of Margins Limit (QAMML), which are probability distributions directly defined in the infinite dimensional space of functional data. And, like new solutions bring new problems, we give our solution to define a density function for QAMML distributions using the Gâteaux derivative: the Gâteaux density. Then we give an extended version of QAMML distribution: the generalized QAMML which is able to capture a more extended range of dependence structures. And we conclude this chapter on the pa-

rameter estimation of QAMML distribution using the Gâteaux density.

In the chapter 6 we show the QAMML and the Gâteaux density in action, using them in three direct applications: the unsupervised classification by mixture decomposition in the Symbolic Data Analysis framework, the Bayesian supervised classification of a classical functional dataset, and we conclude with a statistical computation: the functional confidence interval.

Then we end this thesis with a conclusion, and discuss about the perspectives and open questions.

Chapter 2

Functional Data

> I believe that numbers and functions of Analysis are not the arbitrary result of our minds; I think that they exist outside of us, with the same character of necessity as the things of objective reality, and we meet them or discover them, and study them, as do the physicists, the chemists and the zoologists.
>
> David Hilbert (1862-1943)

2.1 Introduction

In this chapter we briefly introduce the two domains of the data analysis where we have met functional data: the Functional Data Analysis (FDA), obviously, and the Symbolic Data Analysis (SDA).

For the sake of comparison, we can say that symbolic data analysis is a more general framework for data analysis, because it extends the classical framework to new kind variables, included probabilistic functional data (*cdf*: continuous cumulative distribution functions, and *pdf*: probability distribution functions), while functional data analysis is the more general framework for the functional data, because it deals only with functional data, but with all kind of functional data. Therefore FDA and SDA have a common intersection, but none of both contains the other completely (see figure 2.1).

We begin this chapter with the introduction to symbolic data analysis, because the seminal work of this thesis concerned a data analysis task on probability distributions in the SDA framework. Then we have extended the study to all kind of functional data, that is why we close this introductory chapter with the functional data analysis.

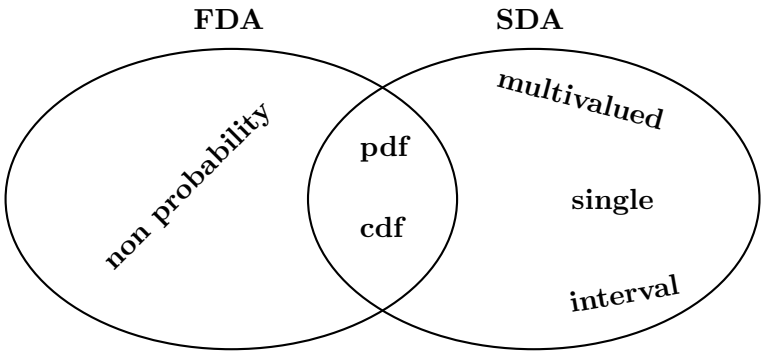


Figure 2.1: The Functional Data Analysis (FDA) and the Symbolic Data Analysis (SDA).

2.2 Symbolic Data Analysis

Nowadays data analysis techniques permit to discover and to structure the knowledge content in databases, and it is a common fact that the storage capacities are still growing. And this growth is a permanent challenge for data analysis to adapt, tune existing techniques or develop new techniques well adapted to these huge mass of data. A path of research is, of course, to develop techniques increasingly effective on very large databases, but another way is to ask whether it is necessary to maintain and amass huge amounts of data. Once "nuggets" of knowledge are extracted of data warehouses, using techniques like clustering, hierarchical classification, and so on, why do not only work on these results? But how to store groups, clusters or category found? The Symbolic Data Analysis can be helpful in this task because: *the aim of Symbolic Data Analysis is to generalize data analysis and statistics to higher-level units described by symbolic data* ([Diday and Noirhomme-Fraiture, 2008]). Aristote already distinguishes two kind of objects: first-level objects called "individuals" (e.g. this apple) and second level objects, called concepts (e.g. the variety of apple called Elstar). A concept can be described in extension (the set of all objects described by the concept) or in intention (what are the shared properties of the objects described by the concept), and symbolic data try to describe the intension of a concept. To give an example of what are symbolic data, we propose to use a synthetic and didactic example, trying to approximate the concept underlying four common varieties of apples: Elstar, Granny Smith, Golden Delicous and Jonagold.

Let us suppose that we have collected in the table 2.1 a large record of

Id	Variety	Size(mm)	Weight (g)	harvest's month	Source	...
124587	Elstar	65	182	September	Belgium	...
124588	Elstar	62	172	September	Belgium	...
⋮	⋮	⋮	⋮	⋮	⋮	⋮
278945	Elstar	58	168	Augustus	France	...
278946	Elstar	59	170	Augustus	France	...
⋮	⋮	⋮	⋮	⋮	⋮	⋮
3789456	Granny Smith	70	190	Augustus	Italy	...
3789457	Granny Smith	72	210	Augustus	Spain	...
⋮	⋮	⋮	⋮	⋮	⋮	⋮
458971	Golden Delicious	67	180	October	US	...
458972	Golden Delicious	69	185	October	Californy	...
⋮	⋮	⋮	⋮	⋮	⋮	⋮

Table 2.1: A classical data table containing informations about apple varieties.

data about four varieties of apples sold in our country[1]: the variety of the apple, its weight and size, the harvest's month, the producing country, ...

Variety	Size(mm)		Weight(g)	
	min	max	min	max
Elstar	58	72	147	188
Granny Smith	65	79	167	197
Golden Delicious	59	75	145	179
Jonagold	62	77	152	178

Table 2.2: Summary about sizes and weights for apples varieties.

Variety	Harvest's month		
	Aug	Sept	Oct
Elstar	21%	77%	2%
Granny Smith	33%	40%	27%
Golden Delicious	47%	20%	33 %
Jonagold	11%	77%	12%

Table 2.3: Summary about sizes and weights for apples varieties.

If we try to modelize each variety of apple with the classical and basics techniques of descriptive statistics, then:

- for the weight and size we can compute the mean and the standard deviation, or the minimum and maximum of these variables, as in table 2.2;

[1]All the data used in this example are not real data, they are just illustrative data.

Variety	Source			
	BEL	FRA	ITA	US
Elstar	41%	28%	21%	10%
Granny Smith	33%	20%	17%	30%
Golden Delicious	12%	18%	40%	30%
Jonagold	31%	38%	20%	11%

Table 2.4: Summary about harvest months for apples varieties.

- for the harvest's month and for the producing countries, we can compute the proportion for each category, as in tables 2.3 and 2.4.

To give a first idea of the underlying concept of a variety it would be more interesting to merge all these informations in one entity:

$$\begin{aligned}
\text{Elstar} &= (\text{Size} = [58;72]) \wedge (\text{Weight} = [147,188]) \\
&\wedge \quad (\text{Harvest} = (\text{Aug} :21\%, \text{Sep} :77\%, \text{Oct} :2\%)) \\
&\wedge \quad (\text{Source} = \text{(BEL:41\%,FRA:28\%,ITA:21\%,US:10\%)}) \\
\text{GrannySmith} &= (\text{Size} = [65;79]) \wedge (\text{Weight} = [167;197]) \\
&\wedge \quad (\text{Harvest} = (\text{Aug} :33\%, \text{Sep} :40\%, \text{Oct} :27\%)) \\
&\wedge \quad (\text{Source} = \text{(BEL:33\%,FRA:20\%,ITA:17\%,US:30\%)}) \\
\text{GoldenDelicious} &= (\text{Size} = [59;75]) \wedge (\text{Weight} = [145;179]) \\
&\wedge \quad (\text{Harvest} = (\text{Aug} :47\%, \text{Sep} :20\%, \text{Oct} :33\%)) \\
&\wedge \quad (\text{Source} = \text{(BEL:12\%,FRA:18\%,ITA:40\%,US:30\%)}) \\
\text{Jonagold} &= (\text{Size} = [62;77]) \wedge (\text{Weight} = [152;178]) \\
&\wedge \quad (\text{Harvest} = (\text{Aug} :11\%, \text{Sep} :77\%, \text{Oct} :12\%)) \\
&\wedge \quad (\text{Source} = \text{(BEL:31\%,FRA:38\%,ITA:20\%,US:11\%)}) .
\end{aligned}$$

That what we call a symbolic description or symbolic data in SDA. The table 2.5 show the corresponding symbolic data table.

Variety	Size(mm)	Weight (g)	Harvest's month	Source	...
Elstar	[58;72]	[147,188]	(Aug: 21%, Sep:77%,Oct:2%)	(BEL: 41%,FRA: 28%, ITA: 21%, US: 10%)	...
Granny Smith	[65;79]	[167;197]	(Aug: 33%, Sep:40%,Oct:27%)	(BEL: 33%,FRA: 20%, ITA: 17%, US: 30%)	...
Golden Delicious	[59;75]	[145;179]	(Aug: 47%, Sep:20%,Oct:33%)	(BEL: 12%,FRA: 18%, ITA: 40%, US: 30%)	...
Jonagold	[62;77]	[152;178]	(Aug: 11%, Sep:77%,Oct:12%)	(BEL: 31%,FRA: 38%, ITA: 20%, US: 11%)	...

Table 2.5: A symbolic data table containing informations about apple varieties.

To give a more precise approximation of the concept, we can also use knowledge, external to the original data table 2.1, like tastes and classical uses shown in the table 2.6, and dominant color for each variety: Elstar are red, Granny Smith are green, Golden Delicous are yellow and Jonagold are red.

Variety	Taste			Uses		
	Sweet	Tart	Sour	Fresh	Dessert	Cooked
Elstar	True	False	True	True	True	False
Granny Smith	False	True	True	True	False	False
Golden Delicious	True	False	False	True	False	True
Jonagold	True	True	True	True	True	True

Table 2.6: Informations about tastes and uses for apples varieties.

Then using these additional informations the symbolic description can be updated as follow:

$$
\begin{aligned}
\text{Elstar} &= (\text{Size} = [58;72]) \wedge (\text{Weight} = [147,188]) \\
&\wedge \ (\text{Harvest} = (\text{Aug} :21\%, \text{Sep} :77\%, \text{Oct} :2\%)) \\
&\wedge \ (\text{Source} = {}_{(\text{BEL:41\%,FRA:28\%,ITA:21\%,US:10\%})}) \\
&\wedge \ \text{Color} = \text{red} \wedge \text{Taste} = \{\text{sweet}, \text{sour}\} \\
&\wedge \ \text{Uses} = \{\text{fresh}, \text{dessert}\} \\
\text{GrannySmith} &= (\text{Size} = [65;79]) \wedge (\text{Weight} = [167;197]) \\
&\wedge \ (\text{Harvest} = (\text{Aug} :33\%, \text{Sep} :40\%, \text{Oct} :27\%)) \\
&\wedge \ (\text{Source} = {}_{(\text{BEL:33\%,FRA:20\%,ITA:17\%,US:30\%})}) \\
&\wedge \ \text{Color} = \text{green} \wedge \text{Taste} = \{\text{tart}, \text{sour}\} \\
&\wedge \ \text{Uses} = \{\text{fresh}\} \\
\text{GoldenDelicious} &= (\text{Size} = [59;75]) \wedge (\text{Weight} = [145;179]) \\
&\wedge \ (\text{Harvest} = (\text{Aug} :47\%, \text{Sep} :20\%, \text{Oct} :33\%)) \\
&\wedge \ (\text{Source} = {}_{(\text{BEL:12\%,FRA:18\%,ITA:40\%,US:30\%})}) \\
&\wedge \ \text{Color} = \text{yellow} \wedge \text{Taste} = \{\text{sweet}\} \\
&\wedge \ \text{Uses} = \{\text{fresh}, \text{cooked}\} \\
\text{Jonagold} &= (\text{Size} = [62;77]) \wedge (\text{Weight} = [152;178]) \\
&\wedge \ (\text{Harvest} = (\text{Aug} :11\%, \text{Sep} :77\%, \text{Oct} :12\%)) \\
&\wedge \ (\text{Source} = {}_{(\text{BEL:31\%,FRA:38\%,ITA:20\%,US:11\%})}) \\
&\wedge \ \text{Color} = \text{red} \wedge \text{Taste} = \{\text{sweet}, \text{tart}, \text{sour}\} \\
&\wedge \ \text{Uses} = \{\text{fresh}, \text{dessert}, \text{cooked}\}.
\end{aligned}
$$

Even if these symbolic descriptions of our varieties of apples can be still improved adding more and more informations, let us conclude this example in listing the different types of symbolic variables met.

This didactic example permits to us to introduce (almost) all kind of symbolic variables:

- numerical variables (not seen in the example of apples, but similar to the classical case),
- categorical single valued variables (e.g., Color = red),
- categorical multi-valued variables: (e.g., Taste = {sweet, sour}),
- interval variables (e.g., Size = [58; 72]),
- modal variables (e.g., Harvest = (Aug :21%, Sep :77%, Oct :2%)).

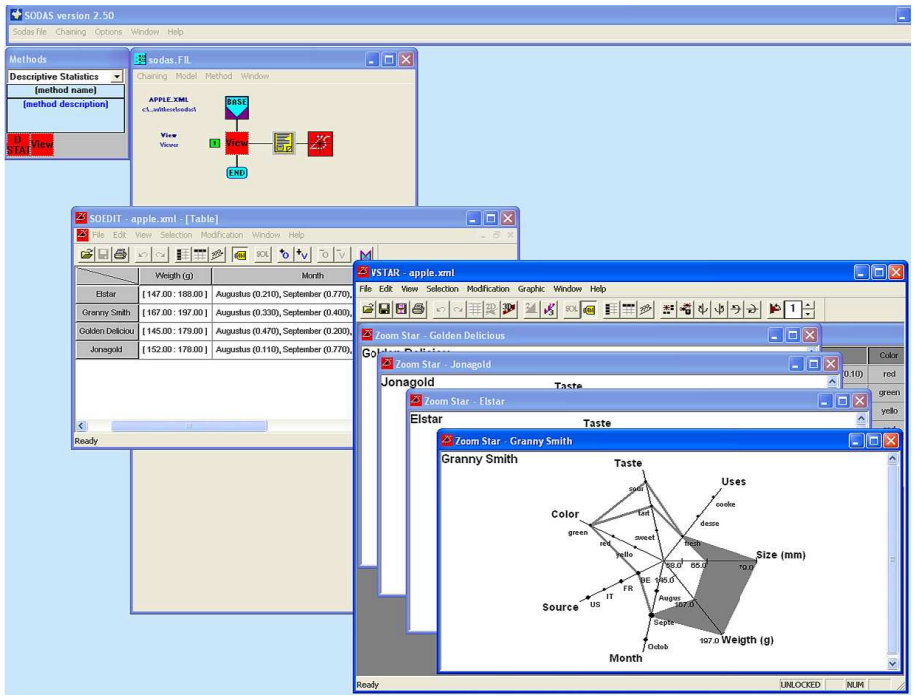
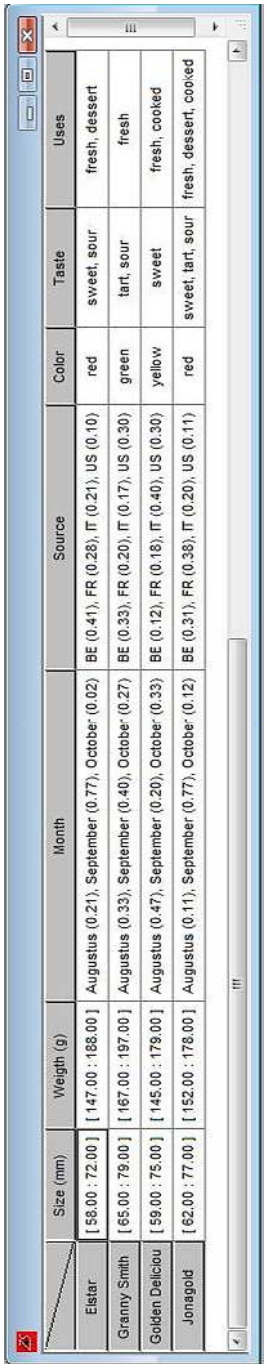


Figure 2.2: A screenshot of the SODAS 2 software.

For all these kind of symbolic data, main classical data analysis techniques were adapted to the symbolic context: statistical descriptions, unsupervised divisive classification, hierarchical and pyramidal clustering, principal component analysis, generalized canonical analysis, bayesian decision trees, factor discriminant analysis, linear regression, visualizations...

A complete review of the symbolic data analysis can be found in [Bock and Diday, 2000], [Billard and Diday, 2006] and [Diday and Noirhomme-Fraiture, 2008].

Moreover all the developed methods for the symbolic data analysis were implemented in two original softwares called SODAS, results of two European projects involving fifteen partners. These software permit also to create, edit and visualize symbolic data. The figure 2.2 gives a screen shot



	Size (mm)	Weigth (g)	Month	Source	Color	Taste	Uses
Elstar	[58.00 : 72.00]	[147.00 : 188.00]	Augustus (0.21), September (0.77), October (0.02)	BE (0.41), FR (0.28), IT (0.21), US (0.10)	red	sweet, sour	fresh, dessert
Granny Smith	[65.00 : 79.00]	[167.00 : 197.00]	Augustus (0.33), September (0.40), October (0.27)	BE (0.33), FR (0.20), IT (0.17), US (0.30)	green	tart, sour	fresh
Golden Deliciou	[59.00 : 75.00]	[145.00 : 179.00]	Augustus (0.47), September (0.20), October (0.33)	BE (0.12), FR (0.18), IT (0.40), US (0.30)	yellow	sweet	fresh, cooked
Jonagold	[62.00 : 77.00]	[152.00 : 178.00]	Augustus (0.11), September (0.77), October (0.12)	BE (0.31), FR (0.38), IT (0.20), US (0.11)	red	sweet, tart, sour	fresh, dessert, cooked

Figure 2.3: The symbolic data table containing informations about apple varieties in the SODAS sofware.

of the SODAS 2 software. This latter is freely avalaible at the following url: `http://www.info.fundp.ac.be/asso/`.

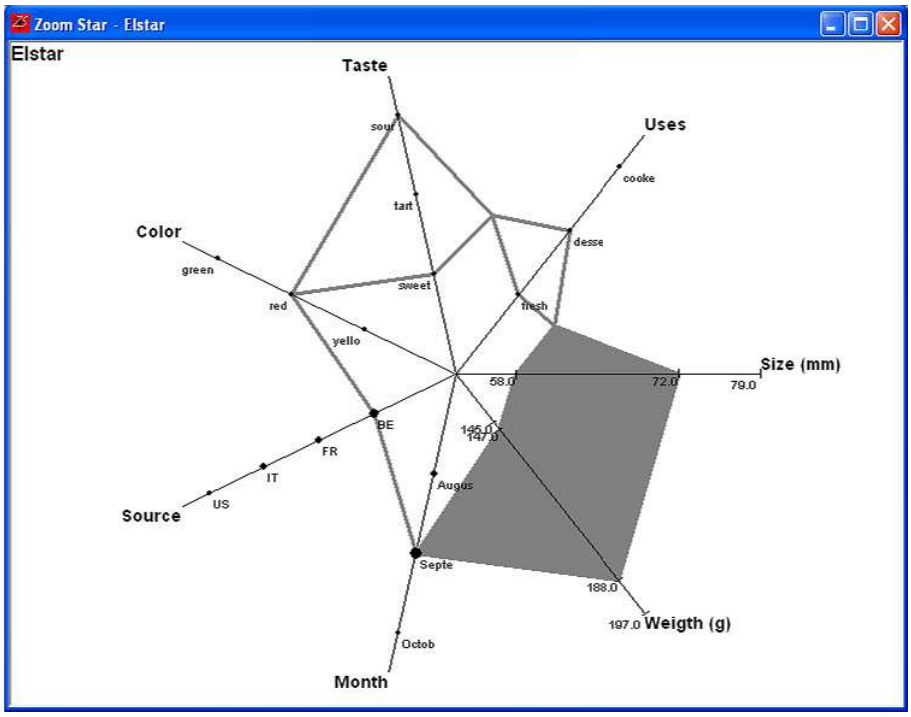


Figure 2.4: The 2D zoom star of the Elstar symbolic data.

For the sake of illustration the figure 2.3 shows the symbolic data table for the varieties of apples in the SODAS 2 software.

Figures 2.4 to 2.8 show the two possible kind visualizations for symbolic data: the 2D zoom star and the 3D zoom star. In the zoom star visualization, each variable is represented on a radial axis. In the 2D cases (figure 2.4 and 2.6), categorical or modal variables are equally distant dots, the size of which is proportional to the weight associated with the category, the limits of intervals and the dots of larger size are joined and the internal surface is colored. The 3D representation (figure 2.5 and 2.7) is more suited when categorical or modal variables are majority, because these latter variables are represented by bar charts. In this case the form is lost but interactivity is greater because users can select the best viewing angle.

Let us suppose now that, in the table 2.1, there is a variable which gives the content in sugar in g (by 100 g of apple). This content in sugar is clearly a continuous random variable. Let us also suppose that, due to the different climates of the producing countries, cumulative distributions functions (cdf) of this concentration are given by the figure 2.8. We can see clearly in this figure that, if, for this numerical variable, we only take the minimum and maximum of the values, we will loss a lot of information about the underlying

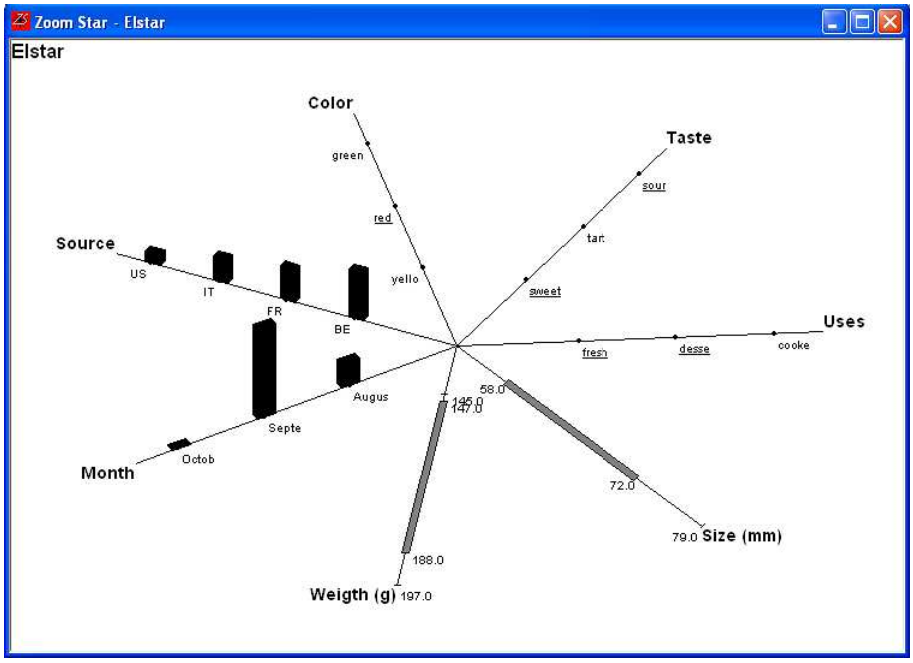


Figure 2.5: The 3D zoom star of the Elstar symbolic data.

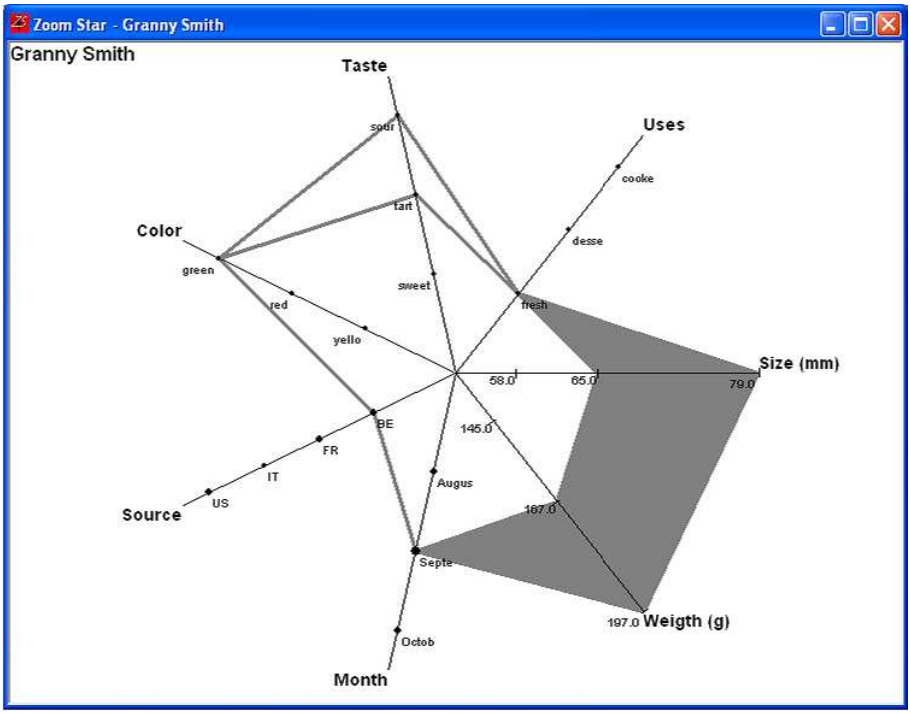


Figure 2.6: The 2D zoom star of the Granny symbolic data.

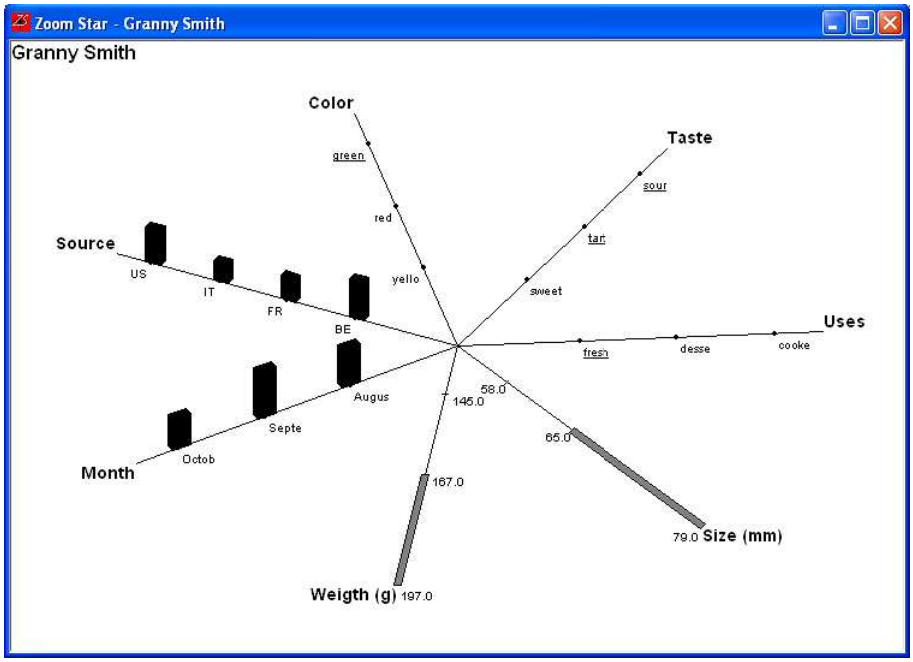


Figure 2.7: The 3D zoom star of the Granny symbolic data.

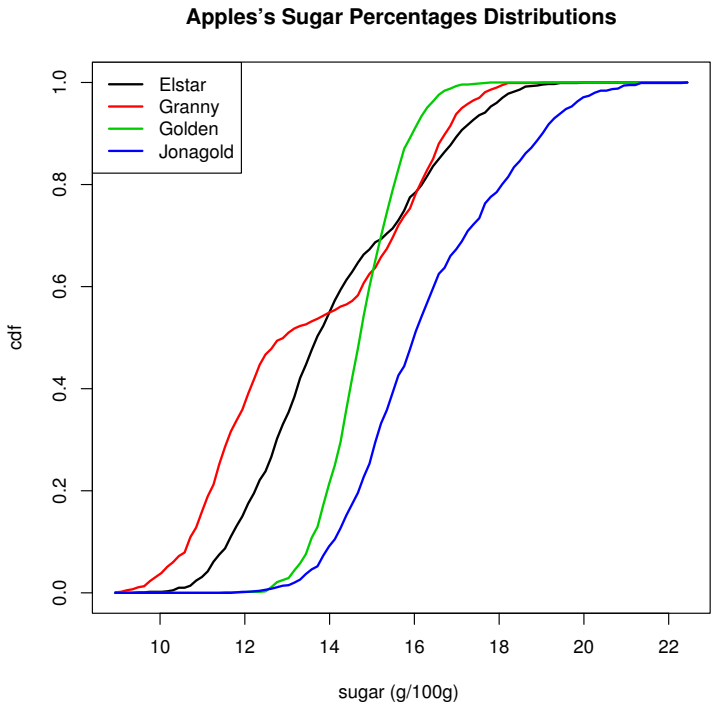


Figure 2.8: The cdfs of the apples varieties.

structure of the variable. Then to minimize this loss of information, the best thing to do is to add directly these *cdfs* in the symbolic descriptions of the varieties of apples:

$$
\begin{aligned}
\text{Elstar} &= (\text{Size} = [58;72]) \wedge (\text{Weight} = [147,188]) \\
&\wedge \ (\text{Harvest} = (\text{Aug}:21\%, \text{Sep}:77\%, \text{Oct}:2\%)) \\
&\wedge \ (\text{Source} = (\text{BEL}:41\%, \text{FRA}:28\%, \text{ITA}:21\%, \text{US}:10\%)) \\
&\wedge \ \text{Color} = \text{red} \wedge \text{Taste} = \{\text{sweet}, \text{sour}\} \\
&\wedge \ \text{Uses} = \{\text{fresh}, \text{dessert}\} \\
&\wedge \ F_{\text{Elstar}} \\
\text{GrannySmith} &= (\text{Size} = [65;79]) \wedge (\text{Weight} = [167;197]) \\
&\wedge \ (\text{Harvest} = (\text{Aug}:33\%, \text{Sep}:40\%, \text{Oct}:27\%)) \\
&\wedge \ (\text{Source} = (\text{BEL}:33\%, \text{FRA}:20\%, \text{ITA}:17\%, \text{US}:30\%)) \\
&\wedge \ \text{Color} = \text{green} \wedge \text{Taste} = \{\text{tart}, \text{sour}\} \\
&\wedge \ \text{Uses} = \{\text{fresh}\} \\
&\wedge \ F_{\text{Granny}} \\
\text{GoldenDelicious} &= (\text{Size} = [59;75]) \wedge (\text{Weight} = [145;179]) \\
&\wedge \ (\text{Harvest} = (\text{Aug}:47\%, \text{Sep}:20\%, \text{Oct}:33\%)) \\
&\wedge \ (\text{Source} = (\text{BEL}:12\%, \text{FRA}:18\%, \text{ITA}:40\%, \text{US}:30\%)) \\
&\wedge \ \text{Color} = \text{yellow} \wedge \text{Taste} = \{\text{sweet}\} \\
&\wedge \ \text{Uses} = \{\text{fresh}, \text{cooked}\} \\
&\wedge \ F_{\text{Golden}} \\
\text{Jonagold} &= (\text{Size} = [62;77]) \wedge (\text{Weight} = [152;178]) \\
&\wedge \ (\text{Harvest} = (\text{Aug}:11\%, \text{Sep}:77\%, \text{Oct}:12\%)) \\
&\wedge \ (\text{Source} = (\text{BEL}:31\%, \text{FRA}:38\%, \text{ITA}:20\%, \text{US}:11\%)) \\
&\wedge \ \text{Color} = \text{red} \wedge \text{Taste} = \{\text{sweet}, \text{tart}, \text{sour}\} \\
&\wedge \ \text{Uses} = \{\text{fresh}, \text{dessert}, \text{cooked}\}. \\
&\wedge \ F_{\text{Jonagold}}
\end{aligned}
$$

where F_{Elstar}, F_{Granny}, F_{Golden}, and F_{Jonagold} are the continuous probability distributions shown in figure 2.8. Then we can introduce a new kind of symbolic variable: continuous probability distribution functions.

This last type of symbolic variable containing a *cdf* is clearly a functional data[2], and even if probability distributions are a very particular type of functional data, mathematical tools developed for functional data in general

[2]In the SDA framework, this thesis focus only symbolic variables expressed by continuous *cdf*, so we do not consider at all other probabilistic symbolic variables, as modal variables or discrete probability distributions.

will be useful for this new kind of symbolic variables. That is why we have found interesting to extend the development of our probabilistic tools to all kind of functional data.

2.3 Functional Data Analysis

The functional data analysis is the part of data analysis which study specifically functional data. We call *functional data*, any phenomenon $\mathcal{X}(t)$ which can be measured for any value of a real parameter t belonging to a range $[t_{\min}, t_{\max}]$. Functional data arise naturally in the real life, and can be considered as a subset of stochastic process, but, of course, the adjective functional is not chosen innocently: it supposes that, for a given data $\mathcal{X}$, the value in a given t, depends of this latter, and is related to the values in other ts. Of course, in most of the cases, t represents the time, but it is not always the case.

To introduce to functional data analysis, let us introduce functional data with four examples of functional datasets, coming from different studies. These examples, and their complete descriptions and analysis, can be found in three fundamentals books about functional data analysis: [Ramsay and Silverman, 2005], [Ramsay and Silverman, 2002] and [Ferraty and Vieu, 2006].

The figure 2.9 shows 15 of the heights curves of 54 girls measured at 31 different times, between 1 and 18 years (see [Tuddenham and Snyder, 1954], [Ramsay and Silverman, 2005] or [Ramsay and Silverman, 2002]) [3]. Even if the data are stored in a 54×31 matrix, they are clearly discrete version of continuous curves, where the height depends obviously of the time t. Moreover, although we can observe some variations between different curves, they share more or less the same shape, then we can talk about the height for a given girl, this function depending from the time.

The second example shown in the figure 2.10 (see [Ramsay and Silverman, 2002])[4] displays, in the top panel, the position of the center of the lower lip of a speaker pronouncing the syllable "bob" for 32 replications, the middle panel displays the corresponding accelerations of the lip and the bottom panel contains electromyogram (EMG) recordings for a facial muscle that depresses the lower lip, the depressor labii inferior. These data were collected to understand the link between the neural activity sends to the labii inferior and the real movement of the lip when pronouncing a word. In this example, t is the time measured in milliseconds, and again if data are stored in a 32×501 matrix, they

[3] Dataset available at: `ftp://ego.psych.mcgill.ca/pub/ramsay/FDAfuns/Matlab/growth.zip`.

[4] Dataset available at: `ftp://ego.psych.mcgill.ca/pub/ramsay/FDAfuns/Matlab/lip.zip`.

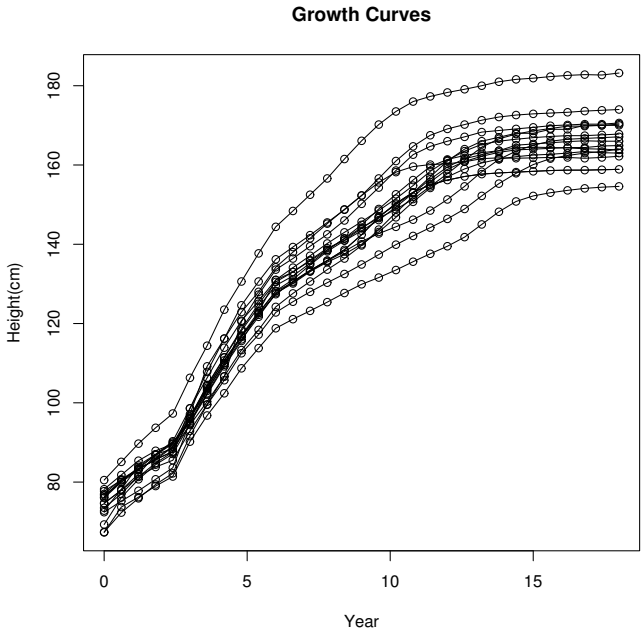


Figure 2.9: Heights of the first 15 females in the Berkeley Growth Study. Circles indicate the ages at which the measurements were taken.

are clearly, by the nature of the observed phenomenon, discrete versions of continuous curves. And lip positions and lip accelerations curves share a common shape, with eventually some slight shifts.

The third example shown in the figure 2.11 (see [Ferraty and Vieu, 2003])[5] comes from a quality control problem in the food industry. For each meat sample the data consists of a 100 channel spectrum of absorbances (absorbance is the $-\log_{10}$ of the transmittance measured by the spectrometer). These data were recorded on a Tecator Infractec Food and Feed Analyzer working in the wavelength range 850-1050 nm by the near-infrared (NIR) transmission principle. In this example, the parameter t represents the wavelength for which the measurement was made. Even if data are discretized, they come clearly from continuous curves, but are stored in 215×100 matrix.

Our last example comes from a speech recognition problem (see [Ferraty and Vieu, 2006] or [Hastie et al., 1995])[6]. Data (see figure 2.12) are log-periodograms corresponding to recordings of speakers of 32 ms duration, and are stored in a 4509×256 matrix. Here also, even if we have to deal

[5]Dataset available at: `http://lib.stat.cmu.edu/datasets/tecator`.

[6]The dataset is available at: `http://www-stat.stanford.edu/ElemStatLearn`, which is the website of [Hastie et al., 2001].

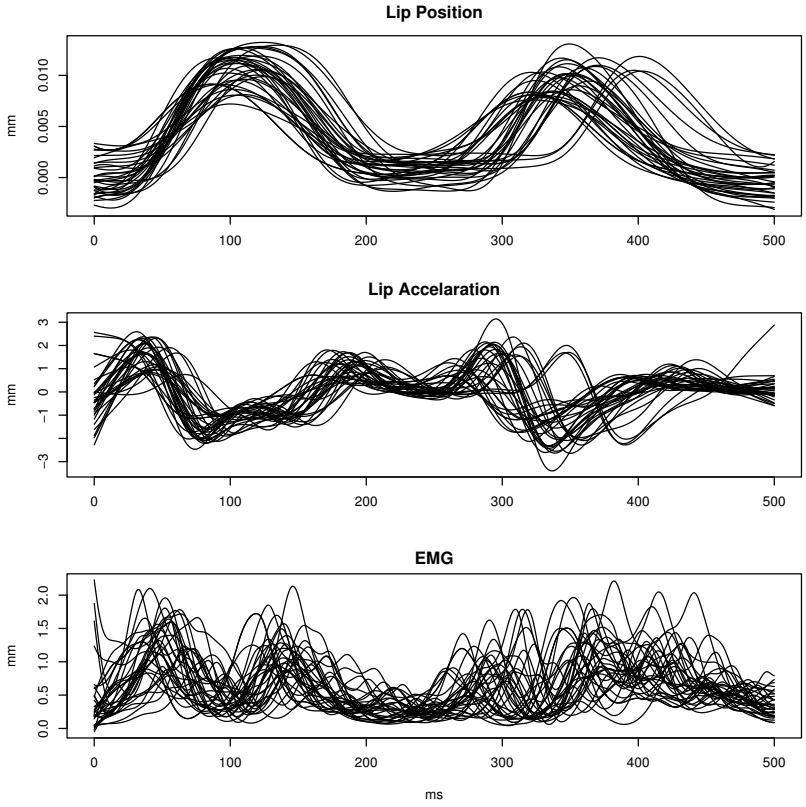


Figure 2.10: Position of the center of the lower lip of a speaker pronouncing the syllable "bob" for 32 replications, corresponding accelerations, electromyogram (EMG) from a facial muscle linked to the lower lip.

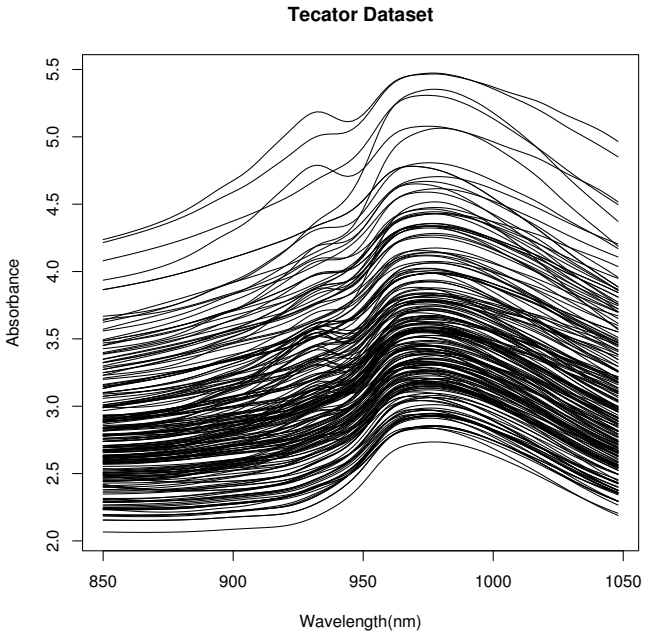


Figure 2.11: The Spectromectric Curves from the Tecator Dataset.

with discretized data, they come from a continuous phenomenon. The study concerns five speech frames corresponding to five phonemes transcribed as follows:

- "sh" as in "she" (group 1);
- "iy" as in "she" (group 2);
- "dcl" as in "dark" (group 3);
- "aa" as the vowel in "dark" (group 4);
- "ao" as the first vowel in "water" (group 5).

In these four examples, it is clear that we are facing to discrete versions of continuous curves. To illustrate the additional functional nature of these data, it is interesting to perform a Principal Component Analysis (PCA) on each dataset containing the discretized versions of these functions. If we plot the two first components of each PCA, then in all cases we can see that there is clearly a scale factor, which traduce the fact that all correlations are positives (except in very few cases for the phonemes data) in the matrix of correlations used for the PCA. This strong relation between variables illustrate why we can call them functional data: any variable depends of an underlying parameter t and moreover is strongly related to the others.

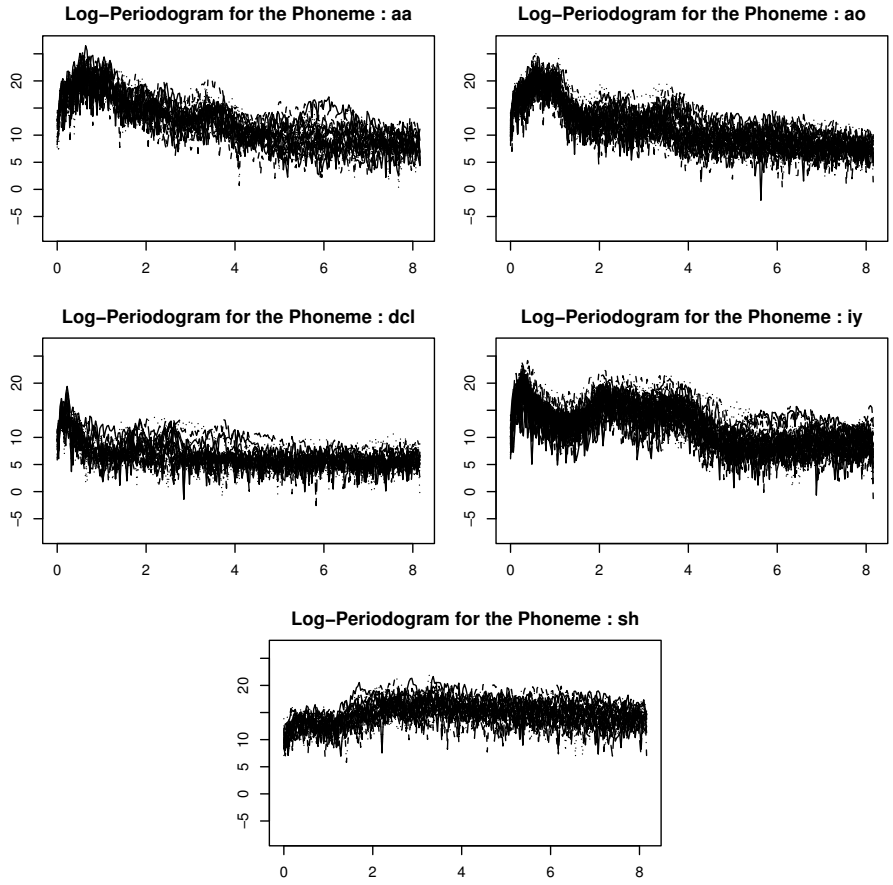


Figure 2.12: The phonemes data.

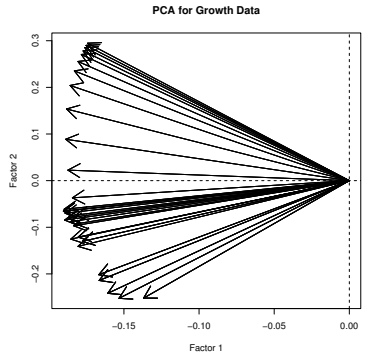


Figure 2.13: The two first factors of the standard PCA for the Growth data.

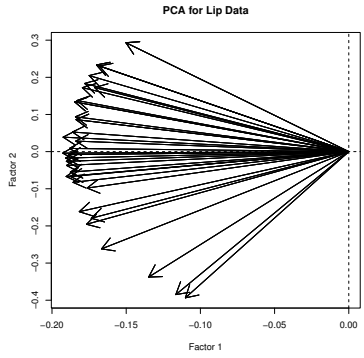


Figure 2.14: The two first factors of the standard PCA for the Lip position data.

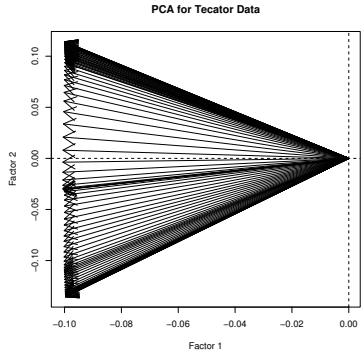


Figure 2.15: The two first factors of the standard PCA for the Tecator data.

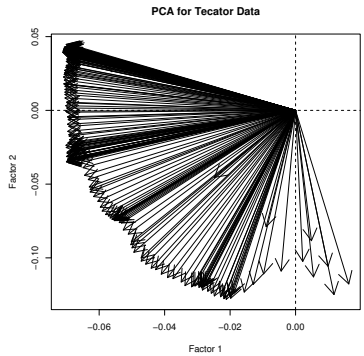


Figure 2.16: The two first factors of the standard PCA for the phonemes data.

Now, let us introduce the most basic and developed tasks of the functional data analysis.

The first goal of the functional data analysis (fda) is to represent and to store data in a way that aid further analysis and reflect the functional nature of these data. As we have seen in the preceding examples, usually, functional data are stored in a multidimensional way: for a given $\mathcal{X}$, we know the values $\{\mathcal{X}(t_1), \ldots, \mathcal{X}(t_n)\}$, for a given subset $\{t_1, \ldots, t_n\}$, and often these values are affected by some noise: in fact, most of the time we know $\{y_1, \ldots, y_n\}$ where

$$y_i = \mathcal{X}(t_i) + \epsilon_i, \tag{2.1}$$

where the noise ϵ_i contributes to the roughness of the raw data. Then the first goal of fda, can be seen as "given $\{y_1, \ldots, y_n\}$, how to find the best estimation for $\mathcal{X}$?" And the more familiar technique of fitting model to data is the *least square estimation*, minimizing the sum of square errors (SSE):

$$SSE = \sum_{i=1}^{n} \left[y_i - \widehat{\mathcal{X}}(t_i) \right]^2. \tag{2.2}$$

Of course, find the minimum for the SSE criterion is only possible when we have a parametric model for $\widehat{\mathcal{X}}$. And in our case any type of function can be met, then the linear model and other simple models are inadequates. We need a very flexible model for $\widehat{\mathcal{X}}$. The usual functional models are based on linear combinations of basis functions:

$$\widehat{\mathcal{X}}(t) = \sum_{k=1}^{K} c_k \phi_k(t). \tag{2.3}$$

The most familiar basis functions are the *monomials*:

$$1, t, t^2; t^3, \ldots, t^k, \ldots$$

or the *Fourier series* terms

$$1, \sin(\omega t), \cos(\omega t), \sin(2\omega t), \cos(2\omega t), \ldots.$$

But the most flexible are based on *spline functions*, and more precisely on B-splines.

A spline is a special function defined piecewise by polynomials: let K points, called *knots*, belonging to a real interval $[a, b]$: $a = t_0 < t_1 < \ldots < t_{K-2} < t_{K-1} = b$. The function

$$S : [a, b] \leftarrow \mathbb{R}$$

is called a spline of degree n if

1. $S \in \mathrm{C}^{n-1}(a,b)$, i.e. S is $n-1$ differentiable with continuous differentials,
2. for any restriction on $[t_i, t_{i+1}]$ $(i = 0, \ldots, k-2)$, S is a polynomial of degree n:
$$S_{|[t_i,t_{i+1}]} \in [a,b]^n(t),$$
where $[a,b]^n(t)$ is the set of polynomials in t of degree n defined in $[a,b]$.

Thus:

- a spline of degree 0, is just a piecewise constant function, and its derivative is the zero function,
- a spline of degree 1, is just a piecewise linear function, and its derivative is a piecewise constant function,
- a spline of degree 2, is just a piecewise quadratic function, and its derivative is a piecewise linear function,
- ...

The most powerful splines functions are the B-spline functions: given the knots $\tau = \{a = t_0, t_1, \ldots, t_{K-2}, t_{K-1} = b\}$ defined above, a B-spline of degree n is given by the following linear combination:

$$\mathbf{S}(t) = \sum_{k=0}^{K+n-1} c_k B_{k,n}(t;\tau) \tag{2.4}$$

where n is the degree of the basis functions $B_{k,n}$. These latter are defined as follow:

$$B_{j,0}(t;\tau) = \begin{cases} 1 & \text{if} \quad t_j \leq t < t_{j+1} \\ 0 & \text{otherwise} \end{cases} \tag{2.5}$$

$$B_{j,n}(t;\tau) = \frac{t - t_j}{t_{j+n} - t_j} B_{j,n-1}(t;\tau) + \frac{t_{j+n+1} - t}{t_{j+n+1} - t_{j+1}} B_{j+1,n-1}(t) \tag{2.6}$$

where $B_{k,n}(t;\tau)$ denotes the value at t of the k^{th} function of the basis defined by the set of knots τ.

Figures 2.17 to 2.19 show the sets of basis functions for a B-spline of, respectively, order 0, 1 and 2. The number of basis functions is equal to $K + n - 1$, where K is the number of knots and n the order of the B-spline. We see that B-splines of order 0 are constant polynomials, B-splines of order 1 are piecewise straight lines and B-splines of order 2 are polynomials of degree 2. See [de Boor, 2002] and [Schumaker, 1981] for more details on spline functions.

There is many advantage to use a basis representation given by expression (2.3), with the B-splines or other functions, to store and work with functional data:

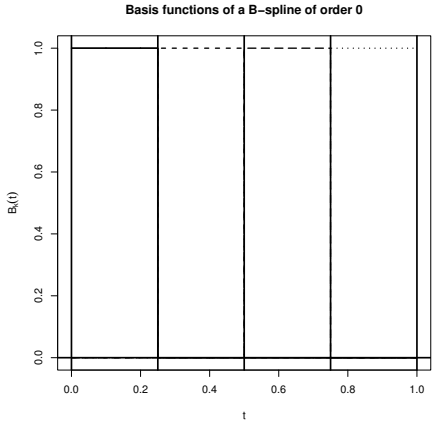


Figure 2.17: The four constant functions defining an order 0 spline with 5 knots:

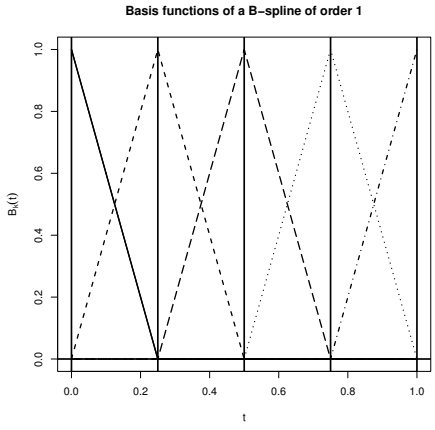


Figure 2.18: The five linear piecewise functions defining an order 1 spline with 5 knots.

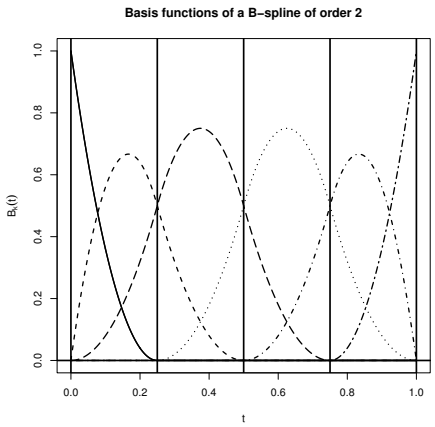


Figure 2.19: The six quadratic piecewise functions defining an order 2 spline with 5 knots.

- we can work directly with functions, and not with discretized versions of these;
- fitting functional data into a basis representation leads to the reduction of the noise ϵ_i;
- once the basis functions are chosen, the only numbers to store to record a fitted functional are the coefficients c_k of the linear combination given by expression (2.3), and then the basis representation can be seen as a dimensionality reduction, and some analysis can be made on these coefficients;
- if the chosen basis representation functions ϕ_k are differentiables, then the derivative of expression (2.3) is given by

$$\frac{d\widehat{\mathcal{X}}(t)}{dt} = \sum_{k=1}^{K} c_k \phi_k^{'}(t). \tag{2.7}$$

This last point is very interesting because, it is well known in real analysis, that the derivatives of a function contain important informations which can not be known when we work only on the original function (i.e. without the derivatives). Thus, in the three first examples of functional data, i.e. the height curves (fig.2.9), the lip positions (fig.2.10) and the spectrometric curves (2.11), once we have plotted the data connecting the different discrete values of each data with simple lines, the fineness of the discretization seems to reveals the "smoothness" of these data. In an analytical point of view, we could say that these data seems continuous, and more, differentiable.

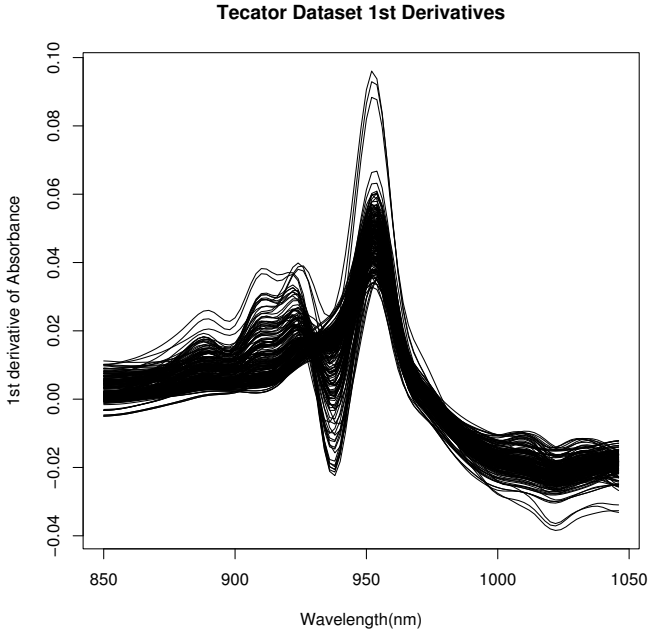


Figure 2.20: First derivatives of the Tecator Data.

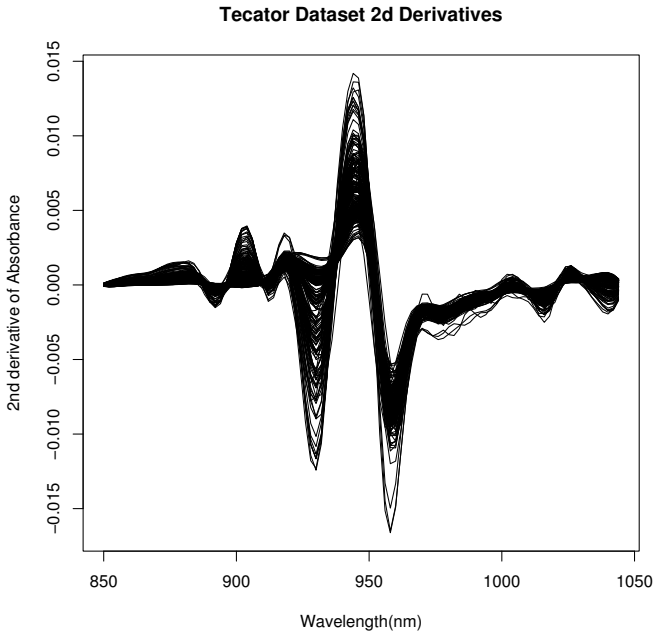


Figure 2.21: Second derivatives of the Tecator Data.

Figures 2.20 and 2.21 show the first and second derivatives of Tecator data.

When functional data are continuous and differentiable, then, often, the derivatives of the data contain also very interesting informations for data analysis tasks. For the sake of illustration the Tecator dataset shown in figure 2.11 contains, in addition to the spectrometric curves, a real variable with the percentage of fat content for each element. And the data are usually separated in two groups: the group with a low fat content (less than 20%) and the group with a high fat content (more or equal to 20%). Figures 2.22 and 2.23 show these two groups, and we see that the only visual difference is a "little jump" of the curves in the group of high fat content, in the wavelengths between 900 and 950 nm. This slight difference is very little to perform one of the classical data analysis: the supervised classification, i.e. the learning of how to assign a new record to one of the two groups using the known data (see section 6.3 in chapter 6 for more details). While, if we look at the first and the second derivatives of the two groups (respectively figures 2.24 and 2.25 and figures 2.26 and 2.27) we see that there is a more clear difference between the two groups, and not only between wavelengths between 900 and 950 nm! And, regardless of the supervised classification method used, the results are always better when the method is applied on the derivatives of these data (on the second derivatives, more precisely).

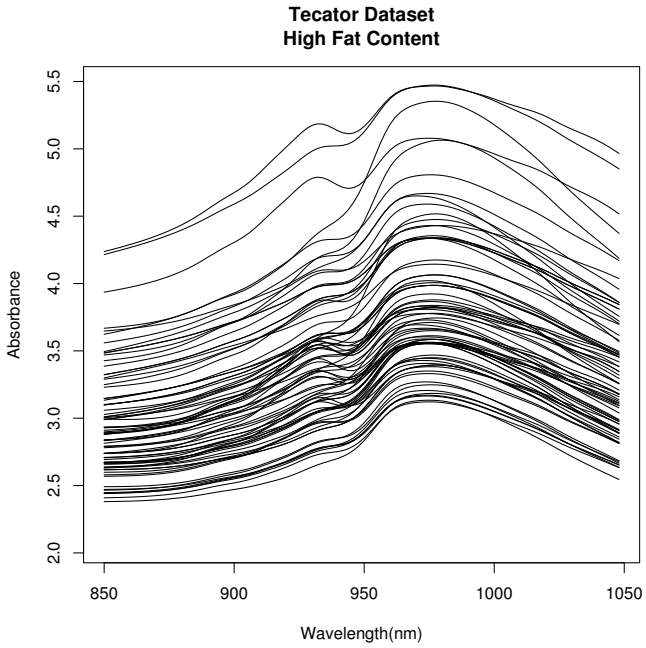


Figure 2.22: Tecator Data with high fat content.

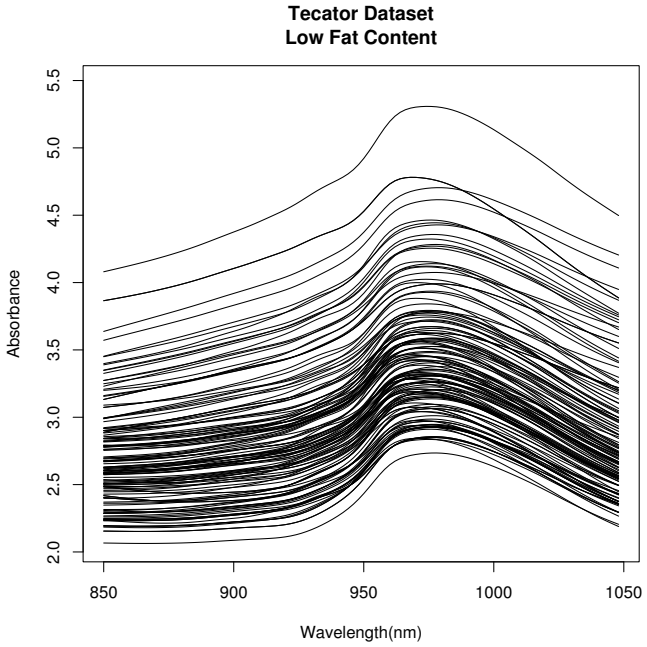


Figure 2.23: Tecator Data with low fat content.

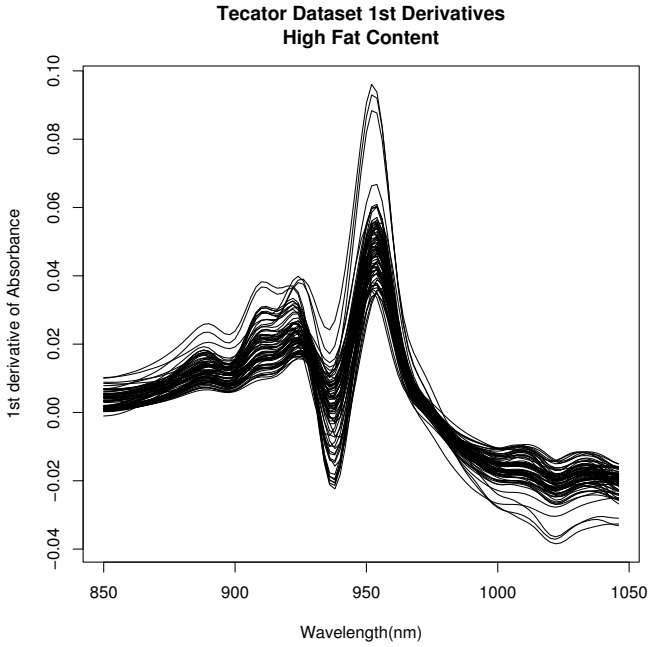


Figure 2.24: First derivatives of the Tecator Data with high fat content.

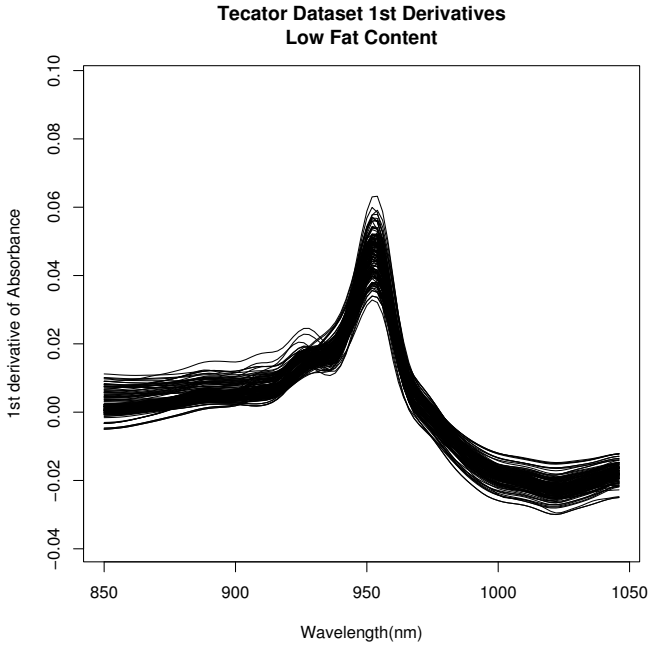


Figure 2.25: First derivatives of the Tecator Data with low fat content.

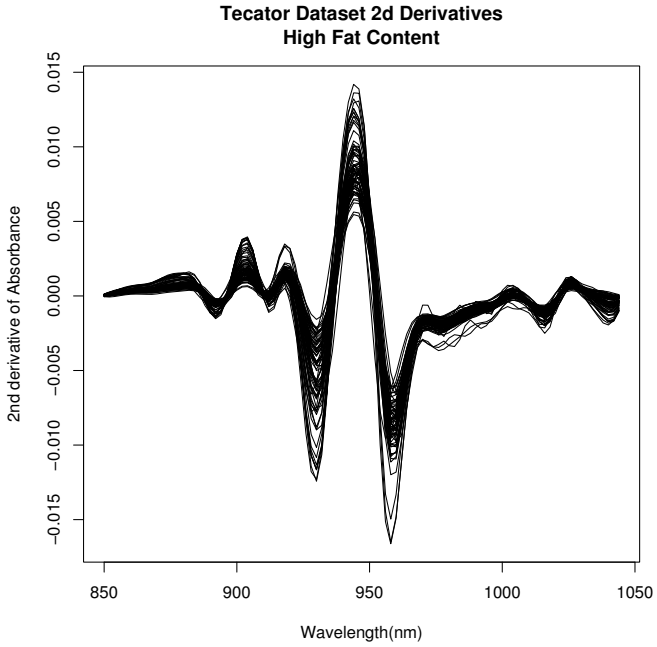


Figure 2.26: Second derivative of the Tecator Data with high fat content.

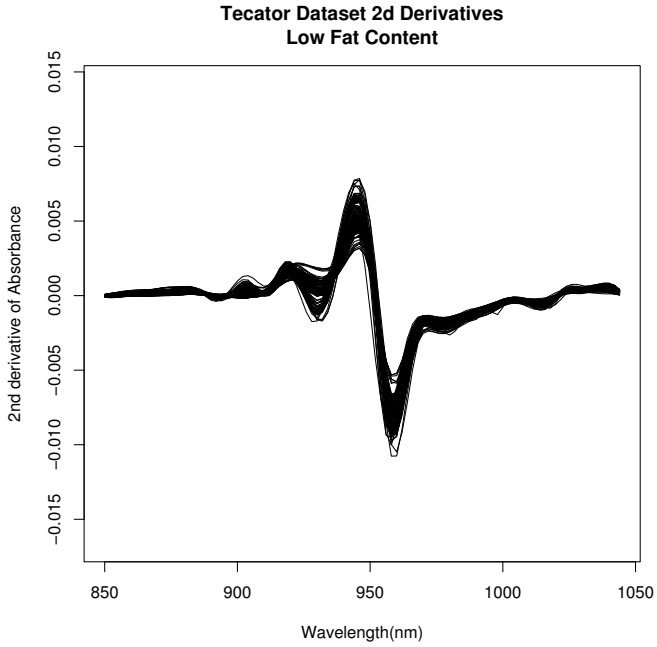


Figure 2.27: Second derivatives of the Tecator Data with low fat content.

This use of the derivatives, is also a good advocacy for a fully functional data analysis approach. Indeed, a common objection to the functional analysis is the following: *"Since functional data are always stored as discretized versions why do not directly use the classical multidimensional techniques?"*. The first argument for fda is the fact that some classical multidimensional methods are not well adapted to data with highly correlated variables, as it is the case for functional data. Another argument is what we called the "curse of dimensionality": finest is the discretization of the curve, higher is the dimension of its discrete version. And the high dimensionality of the data leads to the scarcity of the data included in a ball with fixed radius, when the dimensionality increase[7]. To counter this curse of dimensionality, it is natural to try to use dimensionality reduction techniques, but they do not always take into account the functional nature of the data. Our last argument is directly linked to the use of the derivatives: it is only if we consider and develop tools and methods which take into account the functional nature of functional data that we can hope to discover usable pieces of knowledge.

Once functional data are stored in a "functional way", some improvement can made to the fitting of the functional model given by expression (2.3) for

[7]We will encounter this problem in this thesis and we will also use call it a "volumetric behavior".

a given data. An example of improvement concern the "roughness" wanted for the estimation $\widehat{\mathcal{X}}$. For this purpose a "roughness penalty" is added to the least square criterion (2.2):

$$PEN_m(\widehat{\mathcal{X}}) = \int [\frac{d^m}{ds}\widehat{\mathcal{X}}(s)]^2 ds \tag{2.8}$$

where m is usually equal to 2 when we want to minimize the curvature of the functional.

After the preliminary steps of registering and displaying data, in most of the times users want to explore the data to see the features characterizing typical functions. And for this, in the multivariate context, a very popular analysis technique is the principle component analysis (PCA). In retaining the characteristics of the dataset that contribute most to its variance, PCA can be used for dimensionality reduction in a dataset. In the functional framework, such a technique is interesting because, it permits to focus on important components, ignoring the more "noisy components". Like in the classical framework, in fda, PCA try to find the projection of a given set of data $\{\mathcal{X}_1, \ldots, \mathcal{X}_N\}$ in the subspace which maximizes the portion of explained variance (for a chosen K):

$$\widehat{\mathcal{X}_i}(t) = \sum_{k=1}^{K} f_{ik}\xi_k(t) \tag{2.9}$$

where the basis functions $\{\xi_1, \ldots, \xi_K\}$ are orthogonal.

In data analysis, after a first exploratory step, often, there is a need to find a global model which permits to understand and predict the underlying phenomenon. And a popular tool for this the regression. In the functional case the regression is given by:

$$\mathcal{Y} = r(\mathcal{X}) + \epsilon \tag{2.10}$$

where $\mathcal{X}$ is the functional variable, r the regression function and $\mathcal{Y}$ the response variable. This response can be scalar or functional. As in the multivariate case the linear model is the most popular, and can be fitted using a sum of square criterion.

After regression tools, the probabilistic models are often used in data analysis. The studies of stochastic process has developed probability models in the case of infinite dimensional spaces, but rather with strong assumptions. In the fda framework, the infinite dimensional nature of functional data makes difficult to define general probabilistic models (i.e. with very few assumptions). Some results on probabilities exist in the non linear regression problem, where the regression can be defined using probabilities:

$$r(\mathcal{X}) = \mathbb{E}[\mathcal{Y}|\mathcal{X} = u] \tag{2.11}$$

and the conditional cdf of $\mathcal{Y}$ given $\mathcal{X}$ is defined by:

$$P\left(\mathcal{Y} \leq y | \mathcal{X} = u\right). \tag{2.12}$$

In this case [Ferraty and Vieu, 2006] developed a nonparametric approach, using kernel methods, for these last two expressions.

To give the rate of convergence of the non parametric method detailed in [Ferraty and Vieu, 2006], the concept of *small ball probability function* is used:

$$\varphi_u(r) = P(\mathcal{X} \in B(u, r)) = P(d(\mathcal{X}, u) \leq r). \tag{2.13}$$

where d is a semi-metric.

The kernel methods are also used in supervised and unsupervised classification tasks (see also chapter 6 for more details), to compute the probability that a given functional data u belonging to a group:

$$P(Y = k | \mathcal{X} = u) \tag{2.14}$$

where $Y = k$ means: "the function belong to the group labeled k".

In the same nonparametric approach, [Dabo-Niang, 2002] and [Dabo-Niang, 2004] propose a first approach to the estimation of the density using also small balls over a sample $X_1, \ldots, X_n$:

$$f_n(u) = \frac{1}{n\mu(B(u, r_n))} \sum_{i=1}^{n} I(X_i \in B(u, r_n)) \tag{2.15}$$

where $r_n \to 0$, $n\mu(B(u, r_n)) \to \infty$, when $n \to \infty$.

An extended review of the existing techniques in *functional data analysis* can be found in the following monographes [Ramsay and Silverman, 2005] and [Ferraty and Vieu, 2006]).

It is a well known fact that probability distributions are fundamental tools to build reliable methods for statistical inference and efficient data analysis techniques. Then to extend classical data analysis techniques to the functional case, the functional data analysis need to develop proper probabilities and statistical methods, as noted J. Ramsay [Ramsay and Silverman, 2005]: *Because of the infinite dimensional nature of functional variation, the whole matter of extending conventional methods of inference whether parametric or nonparametric, Bayesian or frequentist is one that will require considerable thought before being well understood. We consider that there is much to do before functional data analysis will have an inferential basis as developed as that of multivariate statistics.*

Chapter 3

Probability Distributions In Infinite Dimensional Spaces

> Life is a school of probability.
>
> Walter Bagehot

3.1 Introduction

In this chapter, we are going to recall, in the framework of the probability theory, the problem of defining a probability distribution when the dimension of the considered space is infinite, as in the case of functional random variables.

We start with some recalls of classical concepts of probability (σ-algebra, measurable functions, probability function) required for the definition of real random variables and their probability and density functions. We consider firstly the case of one real random variable and then the case of a vector of real random variables. After this, we consider the problem of the distribution of a functional random variable, which is another name for stochastic processes. And the probability distributions have been studied largely in this framework, but with rather strong hypotheses [Cox and Miller, 1965], [Gihman and Skorohod, 1974], [Bartlett, 1978] and [Stirzaker, 2005]. Some processes are very famous like Markov process [Meyn and L, 1993]. Such a process has the property that present is not influenced by all the past but only by the last visited state. A very particular case is the random walk, which has the property that one-step transitions are permitted only to the nearest neighboring states. Such local changes of state may be regarded as the analogue for discrete states of the phenomenon of continuous changes for continuous states. The limiting process is called the Wiener process or Brownian motion. The Wiener process is a diffusion process having the spe-

cial property of independent increments. Some more general Markov chain with only local changes of state are permissible, gives also Markov limiting process for continuous time and continuous states. The density probability is solution of a special case of the Fokker-Planck diffusion equation.

If we do not want to assume with strong hypothesis like Markov, then we have to deal with the problem of define and compute probability distributions in an infinite-dimensional space, and the classical way to solve it uses two simplifying steps, by the means of dimensionality reductions. The first step, the separability property, permits to reduce our attention from a dimensionality with continuum hypothesis to an infinite, but countable, dimensionality. The separability is precious for stochastic process since [Doob, 1953] shown that when some conditions are fulfilled, a stochastic process is stochastically equivalent to a separable process.

The second "reduction" step is given by the Kolmogorov's extension theorem, which permits to define a probability on $\mathbb{R}^\infty$ built upon a family of finite dimensional joint distributions. The practical consequence of this theorem is that it is usual to work with finite dimensional distributions when they exist, these latter being easier and well known. But even if this classical solution allows us to solve a lot of classical problems we recall that this remains an approximation and the need of a well defined solution in the infinite dimensional space of functional random variables is still intact.

3.2 Real Random Variables

Even if we do not focus our attention on real random variables, it is interesting to recall some basics of the probability theory [Neveu, 1964].

Definition 3.2.1. *Let Ω an nonempty set, called the* sample space*, a subset $A \subset \Omega$ is called an* event*, and an element $\omega \in \Omega$ is an* elementary event*.*
Let $\mathcal{A}$ be a class of events, then $\mathcal{A}$ is a Boolean algebra *if it contains: the* sure event Ω*, the* impossible event $\emptyset$ *and is* closed *under complementation, finite union and intersections, i.e.:*

(i) $\emptyset \subset \mathcal{A}$ *and* $\Omega \subset \mathcal{A}$*,*

(ii) if $A \in \mathcal{A}$*, then* $A^c \in \mathcal{A}$*,*

(iii) $\forall n \in \mathbb{N}_0$*, if* $A_1, \ldots, A_n \in \mathcal{A}$*, then*

$$\left(\bigcup_{i=1}^{n} A_i\right) \in \mathcal{A}, \ \left(\bigcap_{i=1}^{n} A_i\right) \in \mathcal{A}. \tag{3.1}$$

$\mathcal{A}$ is a σ-algebra *if it is also closed for* countable *unions and intersections, i.e.:*

(iv) if $A_1, A_2, \ldots \in \mathcal{A}$, *then*

$$\left(\bigcup_{i=1}^{\infty} A_i\right) \in \mathcal{A}, \ \left(\bigcap_{i=1}^{\infty} A_i\right) \in \mathcal{A}. \tag{3.2}$$

The pair $(\Omega, \mathcal{A})$ *formed by a set* Ω *and a* σ*-algebra* $\mathcal{A}$ *of subset of* Ω *is called a* measurable space.

Now let us recall the classical probability axioms given by Kolmogorov [Kolmogorov, 1950].

Definition 3.2.2. *Let* Ω *be a set of elementary events,* $\mathcal{A}$ *a* σ*-algebra on* Ω *and* P *an application from* $\mathcal{A}$ *into* $[0,1]$ *such that:*

(i) $0 \leq P(A) \leq 1 \ \forall A \in \mathcal{A}$,

(ii) $P(\Omega) = 1$,

(iii) if $\{A_i\}$ *is any countable sequence of disjoint events, then*

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i). \tag{3.3}$$

The triple $(\Omega, \mathcal{A}, P)$ *is called a* probability space, *and* P *is a* probability function.

As our main interest are probability distributions and random variables, let us recall some definitions in the real random variables [Kolmogorov, 1950] [Billingsley, 1968]. Let us start with the definition of a measurable mapping.

Definition 3.2.3. *Let* $(\Omega, \mathcal{A})$ *and* $(\Omega', \mathcal{A}')$ *be measurable spaces. A mapping* $X : \Omega \to \Omega'$ *from* Ω *into* Ω' *is a* measurable mapping *if* $\forall A' \in \mathcal{A}'$

$$X^{-1}(A') = \left\{a \in A : X(a) \in A'\right\} \subset \mathcal{A}. \tag{3.4}$$

If $X^{-1}(\mathcal{A}')$ *denotes the class* $\left\{X^{-1}(A') : A' \in \mathcal{A}'\right\}$, *this condition may be succinctly stated as* $X^{-1}(\mathcal{A}') \subset \mathcal{A}$.

When Ω' is the n-dimensional Euclidean space $\mathbb{R}^n$, we always take $\mathcal{A}'$ to be the class $\mathcal{R}^n$ of n-dimensional Borel sets (the σ-field generated by the open subsets of $\mathbb{R}^n$ with the Euclidean metric), and we say X is measurable if $X^{-1}(\mathcal{R}^n) \subset \mathcal{A}$. Measurable mappings permits to us to use a probability function, defined in a probability space, in another measurable space.

Proposition 3.2.1. *Let* $(\Omega, \mathcal{A}, P)$ *be a probability space and* $(\Omega', \mathcal{A}')$ *a measurable space. If* X *is a* measurable mapping *from* Ω *into* Ω' *then the following application:*

$$P_X(A') = P\left\{X^{-1}(A')\right\} \tag{3.5}$$

defines a probability function from $\mathcal{A}'$ *into* $[0,1]$, *and a probability space* $(\Omega, \mathcal{A}', P_X)$.

Now, using measurable mappings we can define the concept of real random variable.

Definition 3.2.4. *Let $(\Omega, \mathcal{A}, P)$ be a probability space, then the function*

$$X : \Omega \to I\!R : \omega \mapsto X(\omega) \tag{3.6}$$

is a real random variable (rrv) *if X is measurable, i.e. $X^{-1}(\mathcal{R}) \subset \mathcal{A}$. The function $F_X : I\!R \to [0,1]$ given by:*

$$\begin{aligned} F_X(x) &= P_X\{]-\infty, x]\} && (3.7)\\ &= P\{\omega \in \Omega : X(\omega) \leq x\} && (3.8)\\ &= P[X \leq x] && (3.9) \end{aligned}$$

is called the distribution function of the real random variable X *or* cumulative distribution function (cdf).

If the distribution is differentiable, then its derivative with respect to r is called the probability density *of X at the point r or the* probability density function (pdf)*:*

$$f_X(x) = \frac{d}{dx} F_X(x). \tag{3.10}$$

If the probability density exists then we have the following relation:

$$F_X(x) = \int_{-\infty}^{x} f_X(s)ds \tag{3.11}$$

for each r. In this case F_X is called *continuous* and then we can express the probability function $P_X(A)$ for each Borel set $A \in \mathcal{R}$ in term of f_X:

$$P_X(A) = \int_A f_X(s)ds. \tag{3.12}$$

And in the general case, we write, analogously:

$$P_X(A) = \int_A dF_X(s). \tag{3.13}$$

Sometimes the notation X means the function side of the *rrv*, and sometimes it means a possible value $X(\omega)$, for an event ω. The context of the use will always permits to the reader to know the meaning.

We recall here basics properties of *cdf* of a real random variable [Kolmogorov, 1950].

Proposition 3.2.2. *Let X be a* rrv, *F_X its* cdf *and $x, y \in I\!R$, then F_X has the following properties:*

1. F_X *is monotone increasing, i.e.*

$$x < y \Rightarrow F_X(x) \leq F_X(y),$$

2. F_X *is right-continuous, i.e.* $\forall x_0 \in \mathbb{R}$ *we have*

$$\forall \varepsilon > 0,\ \exists \delta > 0 :\ x_0 < x < x_0 + \delta \Rightarrow |F_X(x) - F_X(x_0)| < \varepsilon,$$

3. $\lim_{x \to -\infty} F_X(x) = 0,$

4. $\lim_{x \to +\infty} F_X(x) = 1.$

Definition 3.2.5. *Let X a real random variable with* cdf F_X*. Then for any $p \in]0,1[$ the* quantile function *of X, denoted Q_X is defined by*

$$Q_X(p) = \inf\{x \in \mathbb{R} : p \leq F_X(x)\}. \tag{3.14}$$

If S is an interval where F_X is strictly increasing, then $Q_X(p) = F_X^{-1}(p)$ in the usual sense.

When several *rrv* are considered we talk about multidimensional real random variables.

Definition 3.2.6. *Let $(\Omega, \mathcal{A}, P)$ be a probability space. Let $n \in \mathbb{N}_0$, then if $X_1, \ldots, X_n$ are n real random variables, the vector $\mathbf{X} = (X_1, \ldots, X_n)$ defines a* multidimensional real random variable.
If we denotes $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$, then the function $H_{\mathbf{X}} : \mathbb{R}^n \to [0,1]$ given by:

$$\begin{aligned} H_{\mathbf{X}}(\mathbf{x}) &= H_{X_1,\ldots,X_n}(x_1, \ldots, x_n) && (3.15)\\ &= P_{X_1,\ldots,X_n}\left\{\prod_{i=1}^{n}]-\infty, x_i]\right\} && (3.16)\\ &= P\left\{\omega \in \Omega : X_i(\omega) \leq x_i,\ i = 1, \ldots, n\right\} && (3.17)\\ &= P\left[X_i \leq x_i,\ i = 1, \ldots, n\right] && (3.18) \end{aligned}$$

is called the distribution function of the random variable $\mathbf{X}$ *or* the joint distribution function of the randoms variables X_i.

The distributions

$$F_{X_i}(x_i) = H_{X_1,\ldots,X_n}(+\infty, \ldots, +\infty, x_i, +\infty, \ldots, +\infty) \tag{3.19}$$

are called the margins *or the* marginals *of the joint distribution $H_{X_1,\ldots,X_n}$.*
Moreover, if there exists the derivatives of $H_{\mathbf{X}}$, then

$$h_{\mathbf{X}}(\mathbf{x}) = \frac{\partial^n}{\partial x_1 \ldots \partial x_n} H_{\mathbf{X}}(\mathbf{x}) \tag{3.20}$$

is called the n-dimensional probability density *of* $\mathbf{X}$ *at the point* $\mathbf{x}$.

If the joint density exists for every point $\mathbf{t} = (t_1, \ldots, t_n) \in \mathbb{R}^n$, then we have the following relation:

$$H_{\mathbf{X}}(\mathbf{x}) = \int_{-\infty}^{x_1} \ldots \int_{-\infty}^{x_n} h_{\mathbf{X}}(\mathbf{s}) ds_1 \ldots ds_n. \tag{3.21}$$

In this case $H_{\mathbf{X}}$ is called *continuous*, and then for every Borel set $A \in \mathcal{R}^n$ we have

$$P_{\mathbf{X}}(A) = \iint \ldots \int_A h_{\mathbf{X}}(\mathbf{s}) ds_1 \ldots ds_n \tag{3.22}$$

Let us close this section of recalls with two important properties of the joint distribution functions. These classical properties will play a major role in the Daniell-Kolmogorov extension method in order to define a probability distribution in infinite-dimensional spaces (cf.p.50).

Let σ a permutation of $\{1, \ldots, n\}$, and r_σ the transformation of the space $\mathbb{R}^n$ such:

$$r_\sigma(x_1, \ldots, x_n) = (x_{\sigma(1)}, \ldots, x_{\sigma(n)})$$

then it is obvious that

$$P_{X_{\sigma(1)}, \ldots, X_{\sigma(n)}}(A) = P_{X_1, \ldots, X_n}(r_\sigma^{-1}(A)). \tag{3.23}$$

Now let $\pi_{n,k}$ denotes the natural projection from $\mathbb{R}^n$ into $\mathbb{R}^k$ $(k < n)$ such that

$$\pi_{n,k}(x_1, \ldots, x_n) = (x_1, \ldots, x_k).$$

The inverse function is defined in the following manner:

$$\pi_{n,k}^{-1}(x_1, \ldots, x_k) = \{(x_1, \ldots, x_k, x_{k+1}, \ldots, x_n), x_{k+1}, \ldots, x_n \in \mathbb{R}\}. \tag{3.24}$$

The application $\pi_{n,k}^{-1}$ is obviously measurable, then using (3.5), p.37, we have the following property:

$$P_{X_1, \ldots, X_k}(A) = P_{X_1, \ldots, X_n}(\pi_{n,k}^{-1}(A)) \tag{3.25}$$

where $A \subset \mathbb{R}^k$.

For the corresponding distribution functions, (3.23) and (3.25) become

$$H_{X_{\sigma(1)}, \ldots, X_{\sigma(n)}}(x_{\sigma(1)}, \ldots, x_{\sigma(n)}) = H_{X_1, \ldots, X_n}(x_1, \ldots, x_n) \tag{3.26}$$

$$H_{X_1, \ldots, X_k}(x_1, \ldots, x_k) = H_{X_1, \ldots, X_n}(x_1, \ldots, x_k, +\infty, \ldots, +\infty). \tag{3.27}$$

Expression (3.27) provides the mean to define the k-multidimensional margins of a joint *cdf*.

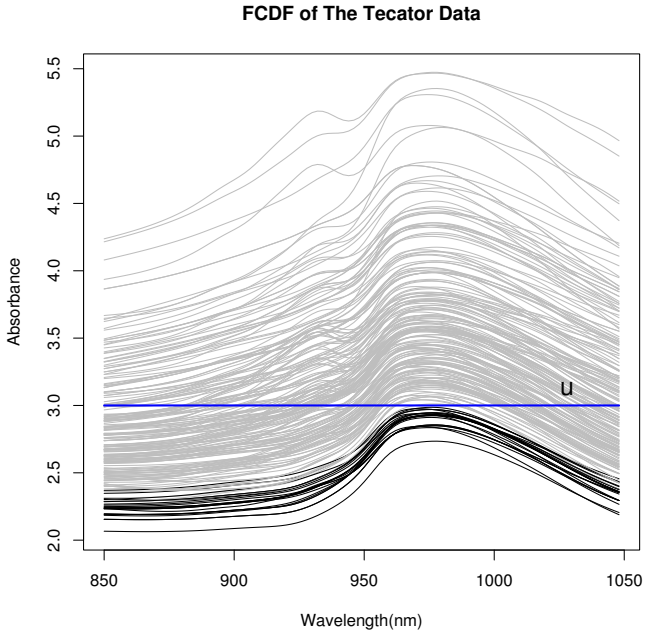


Figure 3.1: Pointwise ordrer illustration (1)

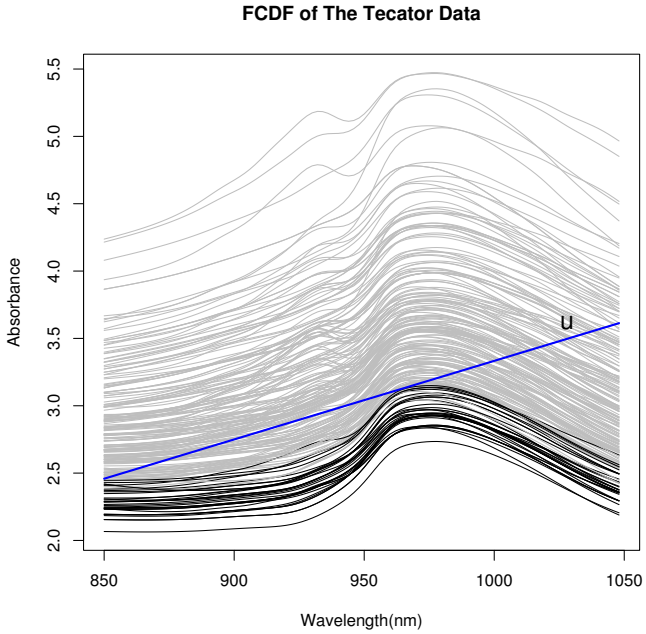


Figure 3.2: Pointwise ordrer illustration (2)

3.3 Functional Random Variables

In the functional case, a random variable becomes a function, thus we are in the field of stochastic processes. We give, here the complete definition used in [Neveu, 1964] (see also [Lamperti, 1977], p2).

Definition 3.3.1. *Let $(\Omega, \mathcal{A}, P)$ a probability space and $\mathcal{D}$ a real interval. A* functional random variable (frv) *or* real random function *on $\mathcal{D}$ is any random variable $\mathcal{X}$ that gives a function from $\mathcal{D} \times \Omega$ into $\mathbb{R}$ and such for any $t \in \mathcal{D}$, $\mathcal{X}(t, .)$, is a real random variable on $(\Omega, \mathcal{A}, P)$.*
Each function $\mathcal{X}(., \omega)$ is called a realization (or a sample path). In the following we will write $\mathcal{X}_t$ for the real random variable resulting from the computations in t of the functional realizations: $\mathcal{X}(t, .)$.

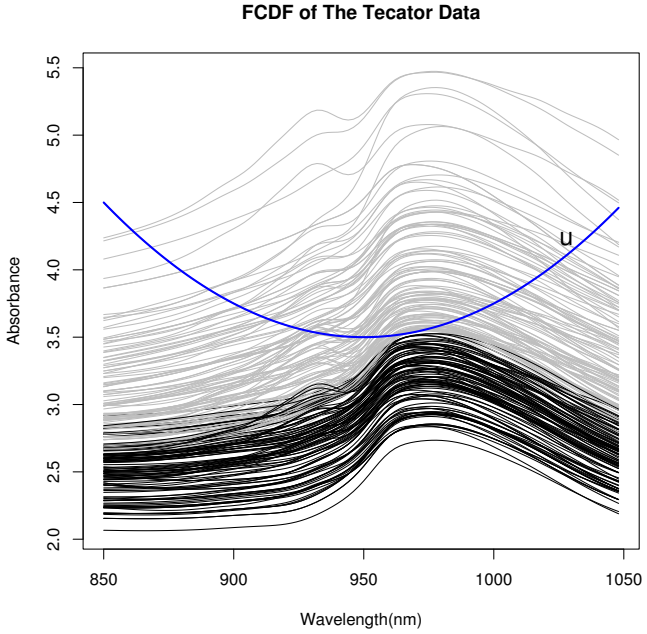


Figure 3.3: Pointwise ordrer illustration (3)

In the case of a real random variable, the distribution is simply defined in expression (3.9) as the probability that random variable is less or equal to the chosen value. In the functional case the same definition can be used if we use an order on a set of functions.

Definition 3.3.2. *Let u, v two functions defined on $\mathcal{D}$. The pointwise orders $\leq_{\mathcal{D}}$ and $<_{\mathcal{D}}$ between two functions u and v defined on $\mathcal{D}$ are given respectively by:*

$$u \leq_{\mathcal{D}} v \iff \forall t \in \mathcal{D}, u(t) \leq v(t), \tag{3.28}$$

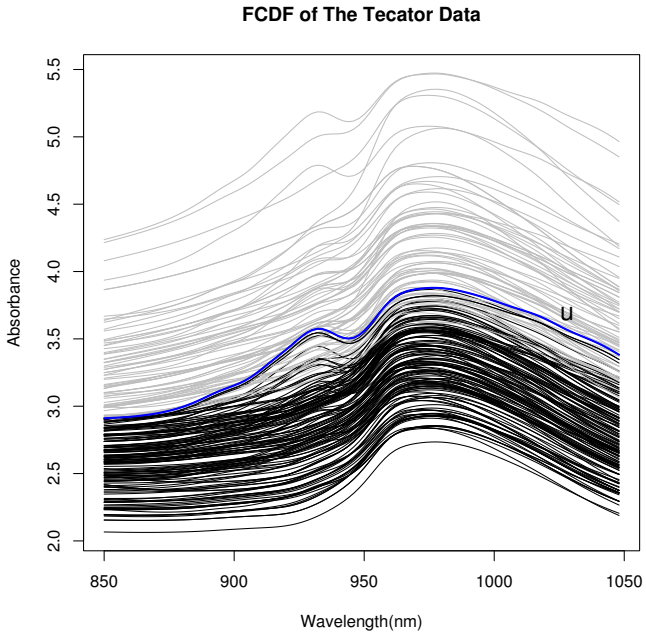


Figure 3.4: Pointwise ordrer illustration (4)

$$u <_{\mathcal{D}} v \iff \forall t \in \mathcal{D}, u(t) < v(t). \tag{3.29}$$

If $\forall t \in \mathcal{D}\ u(x) \geq 0$*, then we will write* $u \geq_{\mathcal{D}} 0$*, and* $u >_{\mathcal{D}} 0$ *in the strict case.*

Figures 3.1, 3.2, 3.3 and 3.4 show four subsets of the Tecator data, less or equal to a chosen function u. In the figure 3.1, u is a constant function; in the figure 3.2, u is straight line with a positive slope. In the figure 3.3, u is a parabola; and finally in the figure 3.4, u is a function from the dataset.

Let us recall the uniform convergence [Lusternik and Sobolev, 1974].

Definition 3.3.3. *Let a sequence* $\{u_n, n \in I\!N\}$ *of functions of* $\mathcal{I}^{\mathcal{D}}$*.*

1. *We say that the sequence* $\{u_n\}$ *is* uniformly convergent *to* $u \in \mathcal{I}^{\mathcal{D}}$ *if,* $\forall \varepsilon > 0, \exists n \in I\!N$ *such,*

$$m \geq n \Rightarrow |u_m(t) - u(t)| < \varepsilon,\ \forall t \in \mathcal{D}$$

or equivalently

$$m \geq n \Rightarrow \sup_{t \in \mathcal{D}} |u_m(t) - u(t)| < \varepsilon.$$

We say that u *is the limit of the sequence* $\{u_n\}$*, and write* $u_n \to u$ *as* $n \to \infty$*, or* $\lim_{n \to \infty} u_n = u$*.*

2. *We say that the sequence* $\{u_n\}$ *is uniformly divergent to* $+\infty$ *if,* $\forall \varepsilon > 0, \exists n \in I\!N$ *such,*

$$m \geq n \Rightarrow u_m(t) > \varepsilon, \ \forall t \in \mathcal{D}$$

or equivalently

$$m \geq n \Rightarrow \sup_{t \in \mathcal{D}} u_m(t) > \varepsilon.$$

We say that $+\infty$ *is the limit of the sequence* $\{u_n\}$*, and write* $u_n \to +\infty$ *as* $n \to \infty$*, or* $\lim_{n\to\infty} u_n = +\infty$*.*

3. *We say that the sequence* $\{u_n\}$ *is uniformly divergent to* $-\infty$ *if, the sequence* $\{-u_n\}$ *is uniformly divergent to* $+\infty$*. We say that* $-\infty$ *is the limit of the sequence* $\{u_n\}$*, and write* $u_n \to -\infty$ *as* $n \to \infty$*, or* $\lim_{n\to\infty} u_n = -\infty$*.*

Remark 3.3.1. *Let us remark that, the uniform convergence can be defined using the pointwise order:*

$$\begin{aligned} \sup_{t\in\mathcal{D}} |u_m(t) - u(t)| < \varepsilon \quad &\Leftrightarrow \quad |u_m(t) - u(t)| < \varepsilon, \forall t \in \mathcal{D} \\ &\Leftrightarrow \quad u(t) - \varepsilon < u_m(t) < u(t) + \varepsilon, \forall t \in \mathcal{D} \\ &\Leftrightarrow \quad u(t) - \varepsilon <_{\mathcal{D}} u_m(t) <_{\mathcal{D}} u(t) + \varepsilon. \end{aligned}$$

Define and compute the probability of such realizations of a functional random variable is the main goal of this work, and using the pointwise order between two functions, we can define this probability distribution as easily as for a real random variable.

Definition 3.3.4. *The* functional cumulative distribution function (fcdf) *of an* frv $\mathcal{X}$ *on* $\mathcal{D}$ *computed at* u *defined on* $\mathcal{D}$ *is given by:*

$$\begin{aligned} \mathcal{F}_{\mathcal{X},\mathcal{D}}(u) &= P[\mathcal{X} \leq_{\mathcal{D}} u] && (3.30) \\ &= P[\mathcal{X}(t) \leq u(t), \forall t \in \mathcal{D}] && (3.31) \\ &= P\{\omega \in \Omega : \mathcal{X}(t,\omega) \leq u(t), \forall t \in \mathcal{D}\} && (3.32) \\ &= P\left(\mathcal{A}_{\mathcal{X},\mathcal{D}}(u)\right) && (3.33) \end{aligned}$$

where

$$\mathcal{A}_{\mathcal{X},\mathcal{D}}(u) = \{\omega \in \Omega : \mathcal{X} \leq_{\mathcal{D}} u\}. \quad (3.34)$$

This definition is the continuous extension of the notion of distributions of distributions (i.e. when $\mathcal{X}_t$ and $u(t)$ are themselves univariate *cdfs* for any t) proposed by [Diday, 2002] in the symbolic data analysis framework.

More precisely distributions of distributions are finite dimensional distributions (see below section 3.5), while the above definition of *fcdf* is not finite dimensional.

It is easy to see that the pointwise order is a partial order over the function defined on $\mathcal{D}$, and not a total order. This characteristic makes difficult to compute expression (3.30) in an empirical way. Let us start

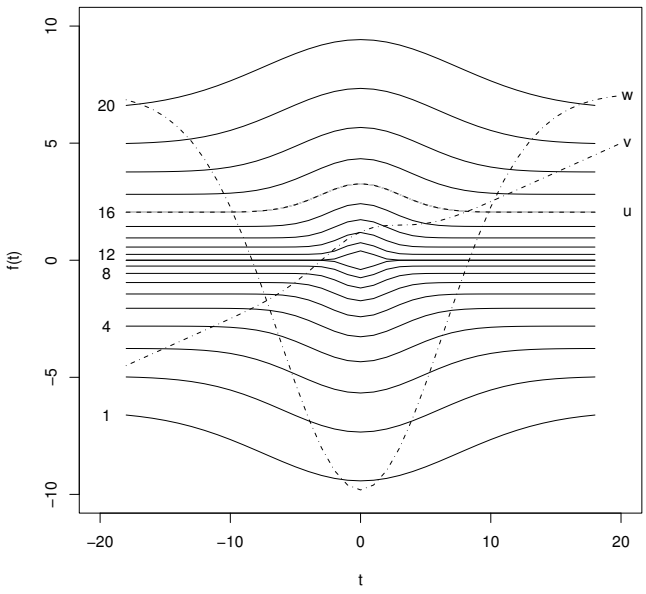


Figure 3.5: A example with 20 sample functions

intuitively with a simple and empirical example on synthetic data (Fig. 3.5), where $\mathcal{D} = [-20, 20]$, and suppose that the drawn functions with solid lines (included u) form a representative sample A of a functional random variable $\mathcal{X}$. We can try to empirically estimate the distribution of $\mathcal{X}$ at u, this later being a function of the sample:

$$\widehat{F}_{\mathcal{X},\mathcal{D}}(u) = \frac{\#\{f \in A : f \leq_{\mathcal{D}} u\}}{\#A} = \frac{16}{20} = \frac{4}{5}.$$

In the same manner we can calculate the distribution in v, which is not a part of the sample:

$$\widehat{F}_{\mathcal{X},\mathcal{D}}(v) = \frac{\#\{f \in A : f \leq_{\mathcal{D}} v\}}{\#A} = \frac{2}{20} = \frac{1}{10}.$$

But for w, as w is not comparable with any function of the sample:

$$\{f \in A : f \leq_{\mathcal{D}} w\} = \emptyset$$

the estimation gives 0, and this in spite of the fact that w is greater than $1/20$ of the functions of A for most of the values of $\mathcal{D}$.

Another problem to compute the expression (3.30) is the infinite nature of the functional data, and more precisely, the uncountable infinity. Thus, following set:

$$\mathcal{A}_{\mathcal{X},\mathcal{D}}(u) = \bigcap_{t\in\mathcal{D}} \{\omega \in \Omega : \mathcal{X}(t,\omega) \leq u(t)\} \tag{3.35}$$

is not necessarily measurable since $\mathcal{D}$ is not countable.

The classical way to find a solution to this problem could be seen as a "simplification" process in two steps: first the separability condition permits us to consider only the infinitely countable case, and then the Kolmogorov's extension theorem leads us to consider only finite dimensional distributions of the process.

Original contribution(s) 1. *In this section we give a formal definition of probability distributions for* frvs *using the pointwise order between two functions.*

3.4 Separable Functions

The separability property is interesting for *frv* defined in metric spaces with images in a compact space. Thus we start this section with the recall of these notions [Lusternik and Sobolev, 1974].

Definition 3.4.1. *A set V is called a* metric space *if for every pair (x, y) of this set is associated a non-negative real number $\rho_V(x, y)$, with the following properties:*

1. $\rho_V(x, y) = 0$ *if and only if* $x = y$ *(*identity*),*
2. $\rho_V(x, y) = \rho_V(y, x)$ *(*symmetry*),*
3. $\rho_V(x, y) + \rho_V(y, z) \geq \rho_V(x, z)$ *(*triangle inequality*).*

The number $\rho_V(x, y)$ is called the distance *between the elements x and y.*

Definition 3.4.2. *A space V is by definition* separable *if it contains a countable, dense subset.*

In other words, if V is separable, then there exists in V a sequence $\mathbf{x} = (x_1, x_2, \ldots)$, belonging to $\mathbb{R}^\infty$ the set of sequences of real numbers, such that for any $x \in V$ we find a subsequence $\{x_{n_1}, x_{n_2}, \ldots\}$ of $\mathbf{x}$ which converges to x.

And if V is a metric space, then for every $x \in V$ and every $\varepsilon > 0$ there exists an element x_{n_0} of $\mathbf{x}$, verifying $\rho_V(x, x_{n_0}) < \varepsilon$.

An important example is $\mathbb{R}$: $\mathbb{Q}$ is dense in the real line, and $\mathcal{Q}$ is countable, then $\mathbb{R}$ is separable.

We give below two equivalent definitions of a compact space in the general case of metric spaces [Atkinson and Han, 2001].

Definition 3.4.3. *Let S be a subset of a metric space V.*

1. *S is* compact *if every sequence $\{x_j\} \subseteq S$ contains a convergent subsequence $\{x_{j_k}\}$ that converge to a point $x \in S$.*

2. *We say S has an* open covering *if there exists a collection of open sets C such that S is a subset of the union other open sets indexed in C, i.e. if $C = \{U_\alpha | \alpha \in \Lambda\}$ (where Λ is an index set), then*

$$S \subseteq \bigcup_{\alpha \in \Lambda} U_\alpha.$$

   We say S is compact *if for every open covering of S, there is a finite subcover $\{U_{\alpha_j} | j = 1, \ldots, m\} \subseteq \{U_\alpha | \alpha \in \Lambda\}$ which also covers S.*

Now we give the separability property for *frv* [Gihman and Skorohod, 1974].

Definition 3.4.4. *An* frv $\mathcal{X}$ *is called* separable *if there exists in $\mathcal{D}$ an everywhere countable set S of points $\{t_i\}$ and a set N of Ω of probability 0 such that for an arbitrary open set $G \subset \mathcal{D}$ and an arbitrary closed set $F \subset I\!R$ the two sets*

$$\{\omega : \mathcal{X}(t, \omega) \in F, \ \forall t \in G\}$$

$$\{\omega : \mathcal{X}(t, \omega) \in F, \ \forall t \in G \cap S\}$$

differ from each other only on the subset N. The set S is called the separability set.

This property is interesting when, for an *frv* $\mathcal{X}$, we can find a separable *frv* "equivalent" to $\mathcal{X}$.

Definition 3.4.5. *Two* frv $\mathcal{X}_1(t, \omega)$ *and* $\mathcal{X}_2(t, \omega)$ *($t \in \mathcal{D}, \omega \in \Omega$) are called stochastically equivalent if for any $t \in \mathcal{D}$*

$$P\{\mathcal{X}_1(t, \omega) \neq \mathcal{X}_2(t, \omega)\} = 0 \tag{3.36}$$

The interest of separability comes from the following theorem [Doob, 1953].

Theorem 3.4.1 (J.L. Doob)**.** *Let U and V be metric spaces, U be separable, V be compact. An arbitrary random function $\mathcal{X}(t, \omega)$, $t \in U$ with values in V is stochastically equivalent to a certain separable random function.*

Doob's theorem allows us, for a given *frv*, to work with a stochastic equivalent separable *frv*. The distribution of the former needs to compute the probability of an infinite uncountable intersection, whereas the latter needs "only" an infinite countable intersection. In other words the *fcdf* of a separable *frv* is evaluated with an intersection of $\aleph_0$[1] sets, while the equivalent *frv* is evaluated in an intersection of $2^{\aleph_0}$ sets.

Knowing that any interval $\mathcal{D} \subseteq \mathbb{R}$, with the classical Euclidean distance, is a metric space, and any closed real interval $\mathcal{I}$ is compact, in the following, we always work with *frv* belonging to the set $\mathcal{I}^{\mathcal{D}}$ of all functions from $\mathcal{D}$ into $\mathcal{I}$.

If, for a given separability set $S = \{t_1, t_2, \ldots\}$, we define a projection operator $\mathcal{P}_{\mathcal{I}^{\mathcal{D}},\mathbb{R}^\infty}$ from $\mathcal{I}^{\mathcal{D}}$ into $\mathbb{R}^\infty$ by

$$\mathcal{P}_{\mathcal{I}^{\mathcal{D}},\mathbb{R}^\infty}(u) = (u(t_1), u(t_2), \ldots) = \mathbf{u}, \tag{3.37}$$

then Doob's theorem allows us, for a given *frv* $\mathcal{X}$, to work with the following *frv*

$$\mathcal{X}' = \mathcal{P}_{\mathcal{I}^{\mathcal{D}},\mathbb{R}^\infty}(\mathcal{X}) = (\mathcal{X}_{t_1}, \mathcal{X}_{t_2}, \ldots)$$

defined on the S.

Now the *fcdf* can be evaluated with the following expression

$$P\left\{\bigcap_{i\in\mathbb{N}} \{\omega \in \Omega : \mathcal{X}_{t_i}(\omega) \leq u(t)\}\right\}. \tag{3.38}$$

This expression in $\mathbb{R}^\infty$, is not obvious to compute, so the second "simplification" step is to work only in k-dimensional space using the Kolmogorov's extension theorem.

3.5 Kolmogorov's extension

In this section we present the Daniell-Kolmogorov's extension theorem. This result was published by [Daniell, 1919a], but not in a probabilistic context, and later in [Kolmogorov, 1933] associated to the distributions functions. We present this result following the presentation found in [Kolmogorov, 1950] but with some notations used in [Billingsley, 1968] and [Gihman and Skorohod, 1974].

This theorem link the probability in $\mathbb{R}^\infty$ with the probabilities in $\mathbb{R}^n$, $n \in \mathbb{N}_0$. In previous sections we have defined $\pi_{n,k}$, the projection from $\mathbb{R}^n$ into $\mathbb{R}^k$, and $\mathcal{P}_{\mathcal{I}^{\mathcal{D}},\mathbb{R}^\infty}$ the projection from[2] $\mathcal{I}^{\mathcal{D}}$ into $\mathbb{R}^\infty$. Now, let us define a third, and intermediate, projection from $\mathbb{R}^\infty$ into $\mathbb{R}^n$.

[1] $\aleph_0$ is the cardinal of $\mathbb{N}$, and $2^{\aleph_0}$ is the cardinal of $\mathbb{R}$.

[2] More precisely we should consider the projection from $\mathcal{I}^{\mathcal{D}}$ into $\mathcal{I}^\infty$, but $\mathcal{P}_{\mathcal{I}^{\mathcal{D}},\mathbb{R}^\infty}$ is more general since $\mathcal{I}^\infty \subset \mathbb{R}^\infty$.

Definition 3.5.1. *Let $n \in \mathbb{N}_0$ and, $\{i_1, \ldots, i_n\}$ n be different natural numbers, we define $\mathbb{P}_{\mathbb{N},\{i_1,\ldots,i_n\}}$ the natural projection from $I\!R^\infty$ to $I\!R^n$, by*

$$\mathbb{P}_{\mathbb{N},\{i_1,\ldots,i_n\}}(\mathbf{x}) = (x_{i_1}, \ldots, x_{i_n}). \tag{3.39}$$

Of course, for a given vector $(x_{i_1}, \ldots, x_{i_n})$, there exists an infinity of sequence of $\mathbb{R}^\infty$ which have this vector as result of the projection $\mathbb{P}_{\mathbb{N},\{i_1,\ldots,i_n\}}$.

Definition 3.5.2. *For $n \geq 1$, and $A^{(n)} \subset I\!R^n$ a* finite-dimensional set, cylindrical set *or* cylinder*, is a set of the folowing form:*

$$\begin{aligned} \mathcal{C}_{i_1,\ldots,i_n}(A^{(n)}) &= \mathbb{P}^{-1}_{\mathbb{N},\{i_1,\ldots,i_n\}}(A^{(n)}) && (3.40)\\ &= \left\{\mathbf{x} \in I\!R^\infty : \mathbb{P}_{\mathbb{N},\{i_1,\ldots,i_n\}}(\mathbf{x}) \in A^{(n)}\right\}. && (3.41) \end{aligned}$$

A cylindrical set is a Borel cylindrical set *if the corresponding set $A^{(n)}$ is in $\mathcal{R}^n$. All the Borel cylindrical set of the space $I\!R^\infty$ form a Boolean algebra denoted $\mathcal{F}$, and $\mathcal{BF}$ denotes the smallest $\sigma-$algebra containing $\mathcal{F}$.*

In a probabilistic point of view, for a separable *frv* with separability set $S = \{t_1, t_2, \ldots\}$, the cylinder corresponding to $A^{(n)}$ is also given by:

$$\begin{aligned} \mathcal{C}_{i_1,\ldots,i_n}(A^{(n)}) &= \left\{\omega \in \Omega : \mathbb{P}_{\mathbb{N},\{i_1,\ldots,i_n\}} \circ \mathcal{P}_{\mathcal{I}^{\mathcal{D}},\mathbb{R}^\infty}(\mathcal{X}) \in A^{(n)}\right\} && (3.42)\\ &= \left\{\omega \in \Omega : \left(\mathcal{X}_{t_{i_1}}, \ldots, \mathcal{X}_{t_{i_n}}\right) \in A^{(n)}\right\}. && (3.43) \end{aligned}$$

A cylindrical set can be defined with different choices of coordinates. It is obvious that:

$$\mathcal{C}_{i_1,\ldots,i_n}(A^{(n)}) = \mathcal{C}_{i_1,\ldots,i_n,i_{n+1},\ldots,i_{n+m}}(A^{(n)} \times \mathbb{R}^m). \tag{3.44}$$

Then any cylinder defined with the index set $I = \{i_1, \ldots, i_n\}$ can be defined on any index set containing I.

Now, let us suppose that we have a probability function P defined on $(\mathbb{R}^\infty, \mathcal{R}^\infty)$ (where $\mathcal{R}^\infty$ is the σ-field of Borel sets in $\mathbb{R}^\infty$). Then using $\mathbb{P}^{-1}_{\mathbb{N},\{i_1,\ldots,i_n\}}$ and (3.5), it is easy to define joint probability distributions on $(\mathbb{R}^n, \mathcal{R}^n)$. For any Borel cylinder generated by a Borel set $A^{(n)} \in \mathcal{R}^n$, the projection $\mathbb{P}^{-1}_{\mathbb{N},\{i_1,\ldots,i_n\}}$ defines the following probability on $(\mathbb{R}^n, \mathcal{R}^n)$:

$$P_{i_1,\ldots,i_n}(A^{(n)}) = P\{\mathbb{P}^{-1}_{\mathbb{N},\{i_1,\ldots,i_n\}}(A^{(n)})\}. \tag{3.45}$$

This latter defines also the joint distribution functions for the vector of random real variables $(\mathcal{X}_{t_{i_1}}, \ldots, \mathcal{X}_{t_{i_n}})$:

$$H_{i_1,\ldots,i_n}(x_{i_1}, \ldots, x_{i_n}) = P_{i_1,\ldots,i_n}\left\{\prod_{j=1}^{n}]-\infty, x_{i_j}]\right\}. \tag{3.46}$$

But we are, here, more interested by the converse. In other words, under what conditions, a system of distribution functions $H_{i_1,\ldots,i_n}$ given *a priori* define a probability on $(\mathbb{R}^\infty, \mathcal{R}^\infty)$?

Theorem 3.5.1 (Kolmogorov's Extension)**.** *Every system of distribution functions $H_{i_1,\ldots,i_n}$, satisfying the following compatibility conditions:*

$$H_{\sigma(i_1),\ldots,\sigma(i_n)}(x_{\sigma(i_1)},\ldots,x_{\sigma(i_n)}) = H_{i_1,\ldots,i_n}(x_{i_1},\ldots,x_{i_n}), \tag{3.47}$$

$$H_{i_1,\ldots,i_n}(x_{i_1},\ldots,x_{i_n}) = H_{i_1,\ldots,i_n}(x_{i_1},\ldots,x_{i_n},+\infty,\ldots,+\infty), \tag{3.48}$$

defines a probability function P on $\mathcal{BF}$ satisfying (3.45)*.*

We give here the sketch of the proof, the whole proof can be found in [Kolmogorov, 1930], [Lamperti, 1966] and [Billingsley, 1968].

First, let us remark that every distribution function $H_{i_1,\ldots,i_n}$ defines uniquely a corresponding probability function $P_{i_1,\ldots,i_n}$ for all Borel sets of $\mathbb{R}^n$. Then for every cylinder we set the probability by

$$P\left(\mathcal{C}_{i_1,\ldots,i_n}\left(A^{(n)}\right)\right) = P_{i_1,\ldots,i_n}\left(A^{(n)}\right). \tag{3.49}$$

The first thing to verify is the non-ambiguity of this definition, since the same cylinder can be defined by several Borel sets of $\mathbb{R}^n$.

Suppose that the cylinder $\mathcal{C}$ is defined by two Borel sets $A^{(n)}$ and $B^{(m)}$:

$$\mathcal{C} = \mathcal{C}_{i_1,\ldots,i_n}\left(A^{(n)}\right) = \mathcal{C}_{i_1,\ldots,j_m}\left(B^{(m)}\right),$$

Thus, we have two different projections for $\mathcal{C}$:

$$\mathbb{P}_{\mathbb{N},\{i_1,\ldots,i_n\}}(\mathcal{C}) = A^{(n)} \subset \mathbb{R}^n,$$

$$\mathbb{P}_{\mathbb{N},\{j_1,\ldots,i_m\}}(\mathcal{C}) = B^{(m)} \subset \mathbb{R}^m.$$

If we set $\{k_1,\ldots,k_l\} = \{i_1,\ldots,i_m\} \cup \{j_1,\ldots,j_n\}$, then this cylinder can be rewritten as folow:

$$\mathcal{C} = \mathcal{C}_{i_1,\ldots,k_l}\left(A^{(n)} \times \mathbb{R}^{l-n}\right) = \mathcal{C}_{i_1,\ldots,k_l}\left(B^{(m)} \times \mathbb{R}^{l-m}\right).$$

From the above expression it is obvious that the only possible difference between $A^{(n)} \times \mathbb{R}^{l-n}$ and $B^{(m)} \times \mathbb{R}^{l-m}$ is a permutation σ of the copies of $\mathbb{R}$. Then using the two compatibility conditions we have that

$$\begin{aligned} P(\mathcal{C}) &= P_{i_1,\ldots,i_n}\left(A^{(n)}\right) = P_{k_1,\ldots,k_l}\left(A^{(n)} \times \mathbb{R}^{l-n}\right) \\ &= P_{\sigma(k_1),\ldots,\sigma(k_l)}\left(B^{(m)} \times \mathbb{R}^{l-m}\right) = P_{j_1,\ldots,j_m}\left(B^{(m)}\right). \end{aligned}$$

Once its unicity of at is proved, it remains to prove that the probability defined by (3.49) satisfies the axioms of probability (3.2.2), see [Kolmogorov, 1930], [Lamperti, 1966] or [Billingsley, 1968] for details.

The Kolmogorov's extension theorem tells us that, if for any $n \in \mathbb{N}_0$ and any set $\{t_1, \ldots, t_n\}$ $(t_i \in \mathcal{D})$ the probability of this set

$$\mathcal{A}_{\mathcal{X},\mathcal{D},n}(u) = \bigcap_{i=1}^{n} \{\omega \in \Omega : \mathcal{X}(t_i, \omega) \leq u(t_i)\} \tag{3.50}$$

can be computed with a probability P that fulfills the compatibility conditions (3.47) and (3.48), then there exits a probability for the set $\mathcal{A}_{\mathcal{X},\mathcal{D}}(u)$.

But the usual way to proceed is to work with the multivariate distributions of the real random variables $\mathcal{X}_{t_1}, \ldots, \mathcal{X}_{t_n}$ (with $i \neq j \Rightarrow t_i \neq t_j$), called the *finite dimensional distributions* of the process:

$$P[\mathcal{A}_{\mathcal{X},\mathcal{D},n}(u)] = H_{\mathcal{X}_{t_1},\ldots,\mathcal{X}_{t_n}}(u(t_1), \ldots, u(t_n)) \tag{3.51}$$

where $H_{\mathcal{X}_{t_1},\ldots,\mathcal{X}_{t_n}}$ is a classical joint distribution of dimension n.

These finite dimensional distributions are the basic probability distributions of the process [Doob, 1953] and are the most important aspect of the process [Lamperti, 1977]. The use of the finite dimensional distributions is prevalent to solve the problem of compute a distribution of *frv*, and this is due to the fact that [Gihman and Skorohod, 1974]: "*firstly, in many practical problems, random functions are characterized by their marginal distributions(...). Secondly, in many cases, it is simpler to define marginal distributions than the corresponding probability spaces(...)*"

But, even if the finite dimensional distributions allows us to work with the well known joint distributions, they are still approximations of (3.30). Indeed, the Kolmogorov's extension theorem build a probability based on the cylinders, but for any $t < t_1$, or $t_n < t$, or $t_i < t < t_{i+1}$, it is impossible to decide what happens in a probabilistic way. And this lack of information prevents us to consider interesting problems [Kolmogorov, 1950] [3] : "*We can therefore handle all questions touching upon a denumerable sequence of random variables. But if* $(\mathcal{D})$ *is not denumerable, many simple and interesting subset of* $(\mathcal{I}^{\mathcal{D}})$ *remain outside of* $(\mathcal{BF})$*. For example, the set of all element* $(\mathcal{X})$ *for which* $\mathcal{X}_{t_i}$ *remains smaller than a fixed constant for all indices* i*, does not belong to the system* $(\mathcal{BF})$ *if the set* $(\mathcal{D})$ *is non-denumerable*". In other words, the probability defined by expression (3.45) does not allow directly to calculate the *fcdf* defined by the expression (3.30).

So, the need of a probability distribution directly defined in the infinite dimensional space of functional data, still acute, and thus, for several reasons: firstly for avoid the usual approximation of the use of finite dimensional distribution, secondly, because some important problems are out of reach in the actual framework, and finally, because a well defined probability distribution could be an opportunity to develop new tools and methods for functional data analysis.

[3] In this quotation, we have slightly changed some notations with respect to our conventions. Changed parts of the quotation are in brackets.

3.6 Conclusion

In this chapter, after some recalls on probability in the one dimensional case and in the multidimensional case, we have explained the classical way to deal with the problem of define and compute a probability distribution for functional data. This classical methodology can be seen as a two step projection. Firstly the separability property allows us two work in $\mathbb{R}^{\infty}$ (or $\mathcal{I}^{\infty}$) instead of $\mathbb{R}^{\mathcal{D}}$ (or $\mathcal{I}^{\mathcal{D}}$). This first projection step permits to reduce the dimensionality of the problem from $2^{\aleph_0}$ to $\aleph_0$. The second projection step, using the Kolmorogorov's extension theorem, is a projection from $\mathbb{R}^{\infty}$ into the spaces $\mathbb{R}^n$. The idea is to define the probability on $\mathbb{R}^{\infty}$ upon a family of probability of the spaces $\mathbb{R}^n$ using the concept of cylindrical sets. But, even if this solution permits us to work on many problems, it remains very important to define a probability distribution directly defined in a non-denumerable set $\mathcal{D}$, firstly, because the use of finite dimensional distributions involve an approximation of the wanted probability, and secondly, because all problems that require to compute a probability in a non-denumerable domain $\mathcal{D}$, remains out of reach in this framework. For example many probabilistic "questions" about calculus (boundedness, continuity, differentiability) are in this case.

Chapter 4

Building FCDF Using Copulas

A man's errors are his portals of discovery.

James Joyce

Si vous fermez la porte à toutes les erreurs, la vérité restera dehors.

Rabindranàth Tagore

Nul ne peut atteindre l'aube sans passer par le chemin de la nuit.

Khalil Gibran

4.1 Introduction

The main idea of this chapter is a try to compute the *fcdf* of an *frv* $\mathcal{X}$ defined on $\mathcal{D}$ as the limit of finite dimensional distributions (3.51) of dimension n:

$$\mathcal{F}_{\mathcal{X},\mathcal{D}}(u) = \lim_{n\to\infty} P[\mathcal{A}_{\mathcal{X},\mathcal{D},n}(u)] = \lim_{n\to\infty} H_{\mathcal{X}_{t_1},\ldots,\mathcal{X}_{t_n}}\left(u(t_1),\ldots,u(t_n)\right). \quad (4.1)$$

To carry out this try, we will show, firstly, how it is easy to compute the margins of the finite dimensional distributions, and we will see how it is possible to build a joint distribution using copulas and the marginals.

The copulas are the functions that join or "couples" multivariate distributions to their one dimensional distribution functions. Their are also

multivariate distribution functions with uniform one-dimensional margins defined on the interval $[0, 1]$. The name *copula* was given to these functions by Abe Sklar [Sklar, 1959], and is a Latin noun for "a link, tie, bound". This word was first used by Sklar in a statistical sense, but was used, previously and nowadays, in linguistic for a word whose links the subject of a sentence to a predicate. Although the Sklar's paper of 1959 is often seen as the beginning of the study of copulas, but but were already studied by Wassily Hoeffding ([Hoeffding, 1940] and [Hoeffding, 1941])as *standardized distributions* whose support was on $[-1/2, 1/2]^2$ and with uniform margins on $[-1/2, 1/2]$. Maurice Fréchet ([Fréchet, 1951]) gets similar results without knowing the Hoeffding's work, that is why we talk about *Fréchet-Hoeffding* bounds or classes. More recently the same functions were rediscovered by several other authors: Kimmeldorf and Sampson ([Kolmogorov, 1930]) whose called them *uniform representation*, Galambos and Deheuvels ([Galambos, 1978]) whose used *dependence functions*. This rediscovery is probably due to the fact that between 1959 and the early 80' a very few publications was mad on this subject: Sklar's papers [Sklar, 1959] and [Sklar, 1973], an article for Schweizer and Sklar [Schweizer and Sklar, 1974], the work of Schweizer and Wolff [Schweizer and Wolff, 1981] et finally the book of Schweizer and Sklar on *probabilistic metric spaces* [Schweizer and Sklar, 1983]. From the middle of the 80' Christian Genest published, with several co-authors, many papers on the Archimedean copulas. These copulas were known before, because, they derive from the notion of *t-norms* ([Ling, 1965] and [Schweizer and Sklar, 1983]), but he was the first to begin to study largely these functions ([Genest and MacKay, 1986b]). Nowadays copulas has a lot of statistical applications, and the number of papers in his field increases every year.
The Nelsen's book ([Nelsen, 1999]) is the best introduction to the field of copulas. It deals with all mathematical aspects of copulas. We have used several definitions and propositions from this book, but we have not always followed its presentation. We have also found inspiration in the book of Joe ([Joe, 1997]) which is on the more general concept of dependence. Actuaries use the copulas for the risk analysis, and we have also found some interesting examples in a course dedicated to this public [Roncalli, 2002].

In the following we introduce the copulas, and Archimedean copulas, first in the bivariate case and then in the multivariate case. The two cases share most of their characteristics, but have also some differences, however it is more easy to understand all the concepts and properties in 2 dimensions before going in superior dimensions.

We will finish this chapter also showing how this try to use the copulas to build an *fcdf* fails! However this chapter is not only a trace of the progress of our research, because the tool built in the following chapter use the Archimedean copulas, and then, shares all interesting properties of these latter.

Let us conclude this introduction, with our main contribution in the field of copulas: to have found general formulas for the density of the following Archimedean copulas: Frank, Gumbel-Hougaard and Joe.

4.2 Surfaces Of Margins

In the section 3.3, we have recalled that an *frv* $\mathcal{X}$, with value in $\mathcal{I}^{\mathcal{D}}$, is a stochastic process, and a stochastic process is such that, for any $t \in \mathcal{D}$, $\mathcal{X}_t$ is a real random variable. Well then, the study of real random variables is well known, thus we are able to define a function that gives to us the probability distribution of $\mathcal{X}_t$ for any $t \in \mathcal{D}$.

Definition 4.2.1. *Let* $\mathcal{X}$ *an* frv *with value in* $\mathcal{I}^{\mathcal{D}}$*. We define respectively the* surface of margins *and the* surface of densities *associated to* $\mathcal{X}$ *in* $\mathcal{D}$ *as follow:*

$$G_{\mathcal{X},\mathcal{D}} : \mathcal{D} \times I\!R \to [0,1] : (t,y) \mapsto P[\mathcal{X}_t \leq y] = F_{\mathcal{X}_t}(y) \tag{4.2}$$

$$g_{\mathcal{X},\mathcal{D}} : \mathcal{D} \times I\!R \to [0,1] : (t,y) \mapsto \frac{\partial}{\partial y} G_{\mathcal{X},\mathcal{D}}\,(t,y) = f_{\mathcal{X}_t}(y). \tag{4.3}$$

If θ *is the parameter (number or vector) for* $G_{\mathcal{X},\mathcal{D}}$ *and* $g_{\mathcal{X},\mathcal{D}}$*, then we will use the following notation when it is necessary for disambiguation:* $G_{\mathcal{X},\mathcal{D},\theta}\,(t,y)$ *and* $g_{\mathcal{X},\mathcal{D},\theta}\,(t,y)$*.*

Surface of margins and surface of densities are generalizations of the concepts of surface of distributions of distributions and surface of densities of distributions, introduced in [Vrac, 2002], in the symbolic data analysis framework (i.e. when $\mathcal{X}_t$ are themselves univariate *cdfs* for any t).

We can use various methods for determining suitable $g_{\mathcal{X},\mathcal{D}}$ and $G_{\mathcal{X},\mathcal{D}}$ for a chosen of $\mathcal{X}$. Thus, for example, if $\mathcal{X}$ is a Gaussian process, then it is easy to compute (or estimate) the mean value $\mu(t)$ and the standard deviation $\sigma(t)$ for all t in $\mathcal{D}$. Then, the surface of margins and, the surface of densities, are given by:

$$G_{\mathcal{X},\mathcal{D}}\,(t,y) = F_{\mathcal{N}(\mu(t),\sigma(t))}(y) \tag{4.4}$$

$$g_{\mathcal{X},\mathcal{D}}\,(t,y) = f_{\mathcal{N}(\mu(t),\sigma(t))}(y) \tag{4.5}$$

where $F_{\mathcal{N}(\mu(t),\sigma(t))}$ is the *cdf* of a normal random variable with mean $\mu(t)$ and standard deviation $\sigma(t)$, and $f_{\mathcal{N}(\mu(t),\sigma(t))}$ is the corresponding *pdf*.

In other cases we can use estimation techniques, as the empirical cumulative distribution function and the kernel density estimation [Parzen, 1962] to estimate $\widehat{G}_{\mathcal{X},\mathcal{D}}$ and $\widehat{g}_{\mathcal{X},\mathcal{D}}$:

$$\widehat{G}_{\mathcal{X},\mathcal{D}}\,(t,y) = \frac{\#\,\{u_i(t) \leq y\}}{N} \tag{4.6}$$

$$\widehat{g}_{\mathcal{X},\mathcal{D}}\,(t,y) = \frac{1}{N \cdot h(t)} \sum_{i=1}^{N} k\left(\frac{y - u_i(t)}{h(t)}\right) \tag{4.7}$$

Name	$k(t)$
Uniform	$\frac{1}{2}\ \mathbf{1}_{(\lvert t\rvert\leq 1)}$
Triangle	$(1-\lvert t\rvert)\ \mathbf{1}_{(\lvert t\rvert\leq 1)}$
Epanechnikov	$\frac{3}{4}(1-t^2)\ \mathbf{1}_{(\lvert t\rvert\leq 1)}$
Quartic	$\frac{15}{16}(1-t^2)^2\ \mathbf{1}_{(\lvert t\rvert\leq 1)}$
Triweight	$\frac{35}{32}(1-t^2)^3\ \mathbf{1}_{(\lvert t\rvert\leq 1)}$
Cosine	$\frac{\pi}{4}\cos\left(\frac{\pi}{2}t\right)\mathbf{1}_{(\lvert t\rvert\leq 1)}$
Gaussian	$\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}t^2}$

Table 4.1: Non exhaustive list of the most common kernels for the density estimation.

where

- $(u_1(t), ..., u_N(t))$, is the sample over which the estimation is made,
- k is the kernel density function,
- $h(t)$ is the function that gives the window width, and can be automatically estimated by Mean Integrated Square Error (MISE) formula (see [Silverman, 1986] and [Wand and Jones, 1995])

$$h_{optimal}(t) = 1.06 \cdot \widehat{\sigma}(t) N^{-1/5} \tag{4.8}$$

  where $\widehat{\sigma}$ is the standard deviation of the sample in a fixed value t. The table 4.1 shows a list of the most common kernels.

Figures from 4.1 to 4.3 show the surfaces of margins $G_{\mathcal{X},\mathcal{D}}$ for Tecator data using the empirical distribution. Firstly for original data (fig. 4.1),

then for the high fat content group (fig. 4.2), and then for the low fat content group (fig. 4.3). Figures from 4.4 to 4.6 show the same for the first derivatives of the data, and figures from 4.7 to 4.9 show the same for the second derivatives of the data.

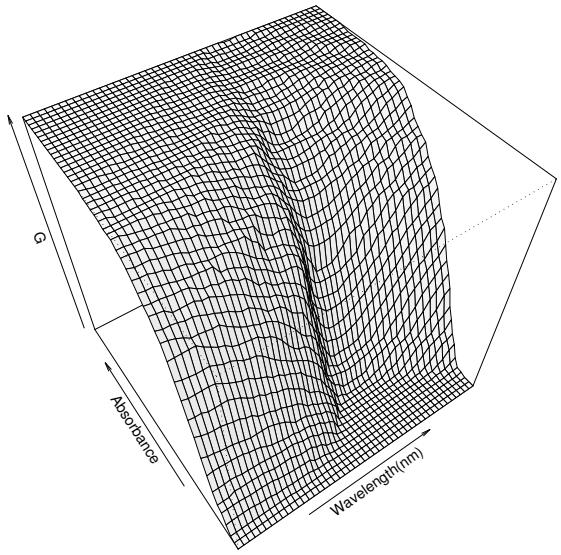


Figure 4.1: The estimated surface of margins $G_{\mathcal{X},\mathcal{D}}(t, y)$ for the Tecator data (see fig. 2.11, p.21).

Figures from 4.10 to 4.18 show the surfaces of densities $g_{\mathcal{X},\mathcal{D}}$ for Tecator data in the same order that above, using the kernel density estimation.

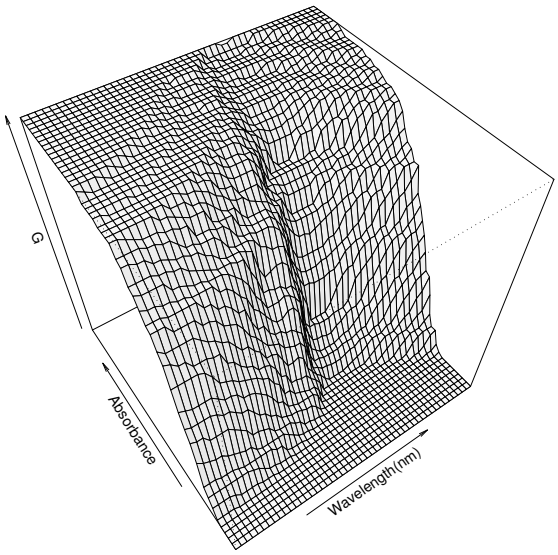


Figure 4.2: The estimated surface of margins $G_{\mathcal{X},\mathcal{D}}(t, y)$ for the Tecator data with high fat content (see fig. 2.22, p.29).

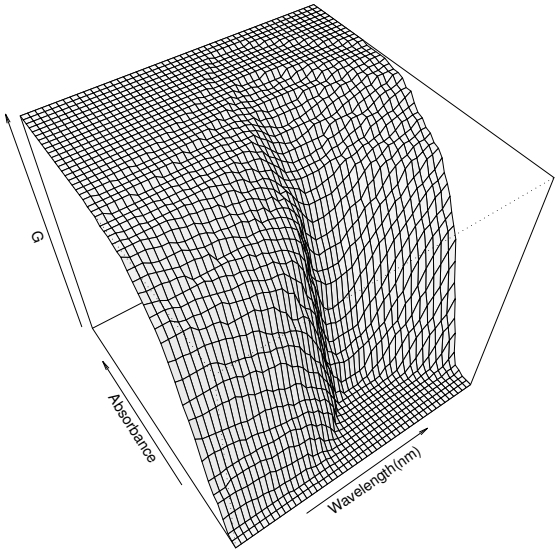


Figure 4.3: The estimated surface of margins $G_{\mathcal{X},\mathcal{D}}(t, y)$ for the Tecator data with low fat content (see fig. 2.23, p.30).

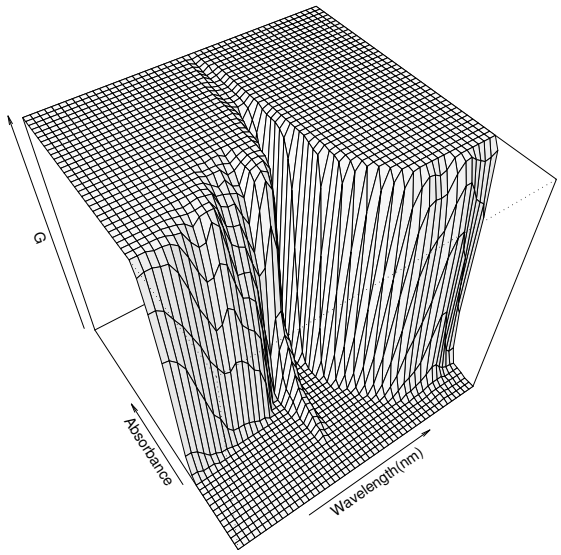


Figure 4.4: The estimated surface of margins $G_{\mathcal{X},\mathcal{D}}(t, y)$ for the first derivative of the Tecator data (see fig. 2.20, p.28).

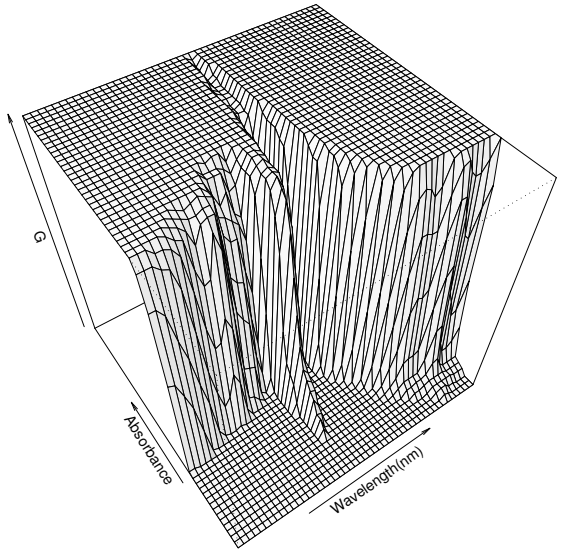


Figure 4.5: The estimated surface of margins $G_{\mathcal{X},\mathcal{D}}(t, y)$ for the first derivative of the Tecator data with high fat content (see fig. 2.24, p.30).

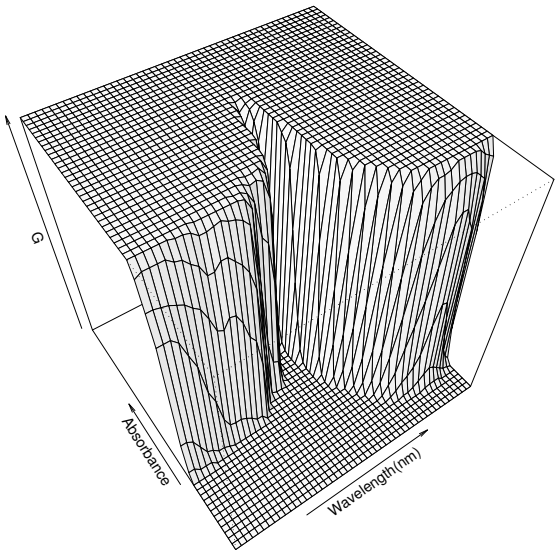


Figure 4.6: The estimated surface of margins $G_{\mathcal{X},\mathcal{D}}(t, y)$ for the first derivative of the Tecator data with low fat content (see fig. 2.25, p.31).

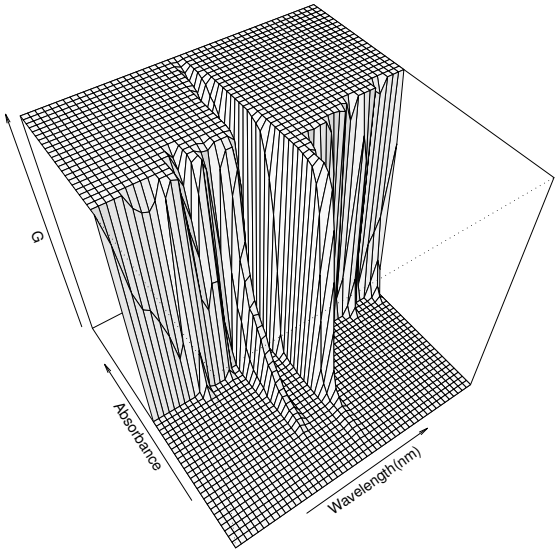


Figure 4.7: The estimated surface of margins $G_{\mathcal{X},\mathcal{D}}(t, y)$ for the first derivative of the Tecator data (see fig. 2.21, p.28).

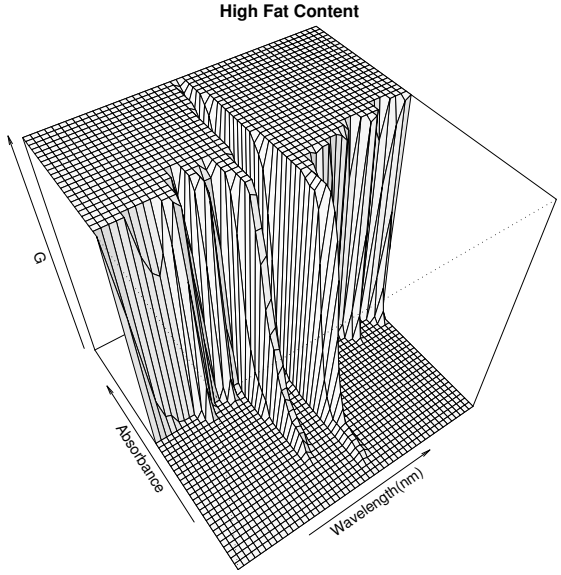


Figure 4.8: The estimated surface of margins $G_{\mathcal{X},\mathcal{D}}(t, y)$ for the first derivative of the Tecator data with high fat content (see fig. 2.21, p.28).

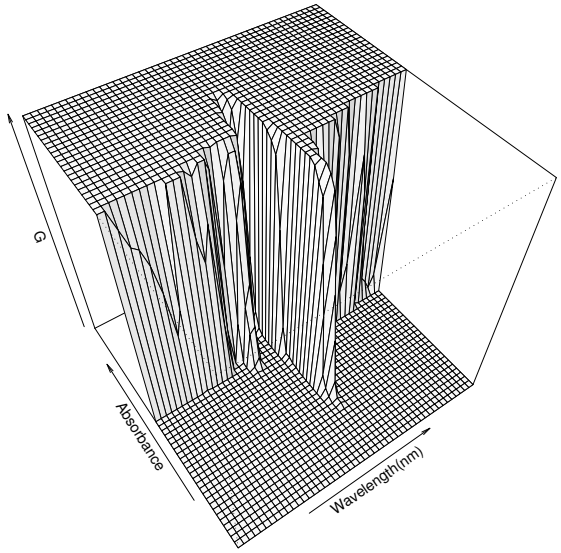


Figure 4.9: The estimated surface of margins $G_{\mathcal{X},\mathcal{D}}(t, y)$ for the first derivative of the Tecator data with low fat content (see fig. 2.21, p.28).

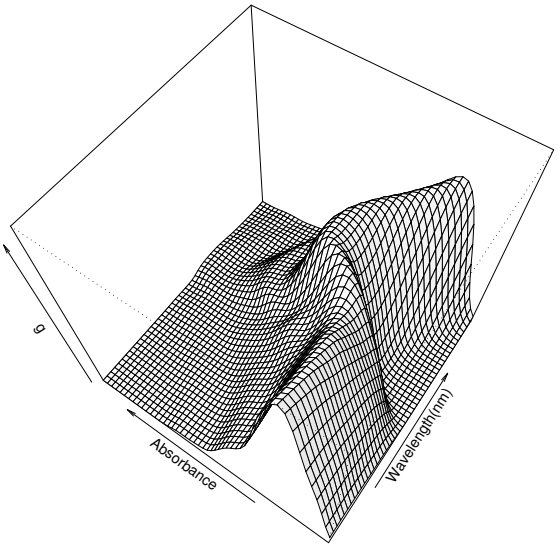


Figure 4.10: The estimated surface of densities $g_{\mathcal{X},\mathcal{D}}(t, y)$ for the Tecator data (see fig. 2.11, p.21).

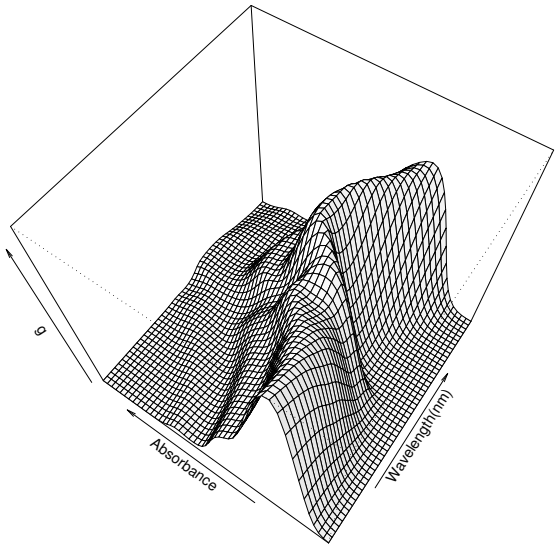


Figure 4.11: The estimated surface of densities $g_{\mathcal{X},\mathcal{D}}(t, y)$ for the Tecator data with high fat content (see fig. 2.22, p.29).

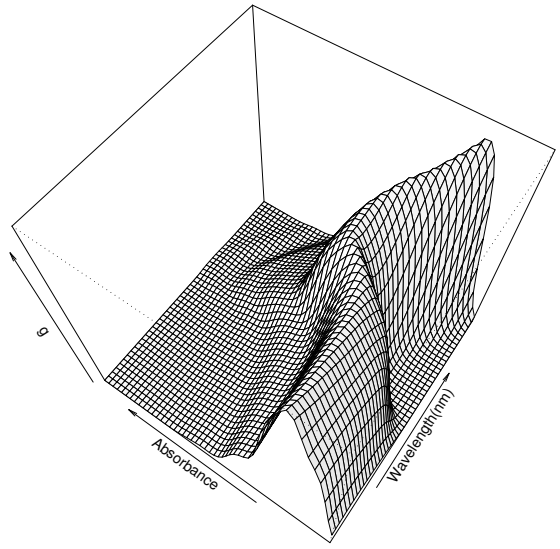


Figure 4.12: The estimated surface of densities $g_{\mathcal{X},\mathcal{D}}(t, y)$ for the Tecator data with low fat content (see fig. 2.23, p.30).

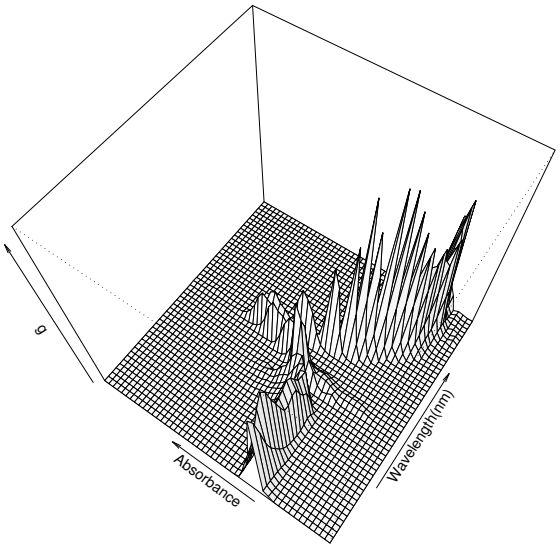


Figure 4.13: The estimated surface of densities $g_{\mathcal{X},\mathcal{D}}(t, y)$ for the first derivative of the Tecator data (see fig. 2.20, p.28).

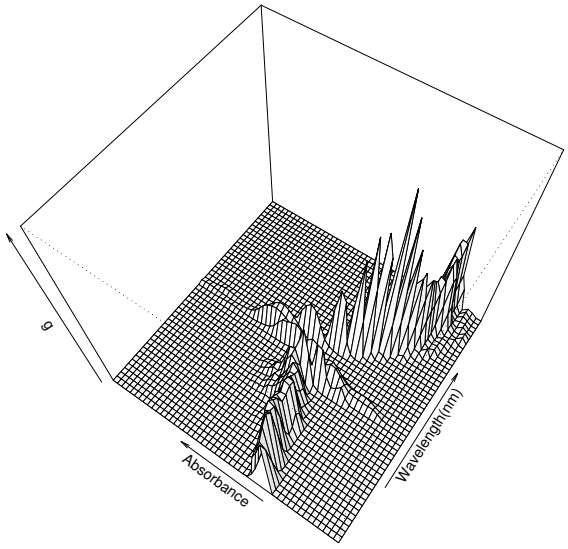


Figure 4.14: The estimated surface of densities $g_{\mathcal{X},\mathcal{D}}(t, y)$ for the first derivative of the Tecator data with high fat content (see fig. 2.24, p.30).

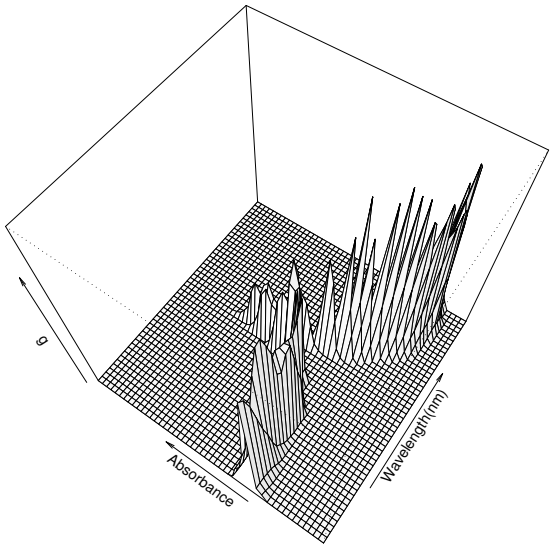


Figure 4.15: The estimated surface of densities $g_{\mathcal{X},\mathcal{D}}(t, y)$ for the first derivative of the Tecator data with low fat content (see fig. 2.25, p.31).

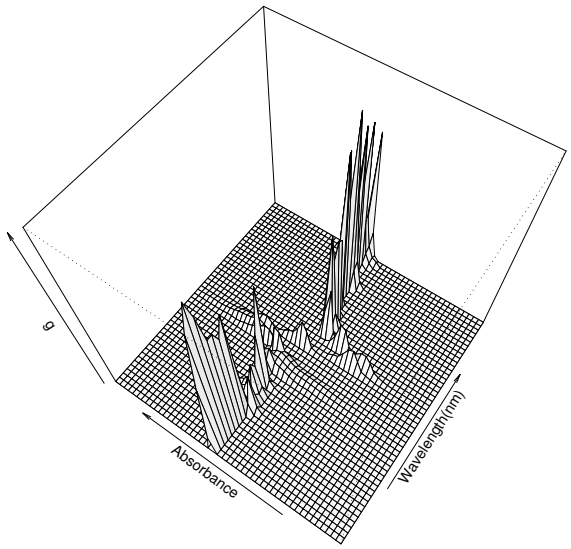


Figure 4.16: The estimated surface of densities $g_{\mathcal{X},\mathcal{D}}(t, y)$ for the first derivative of the Tecator data (see fig. 2.21, p.28).

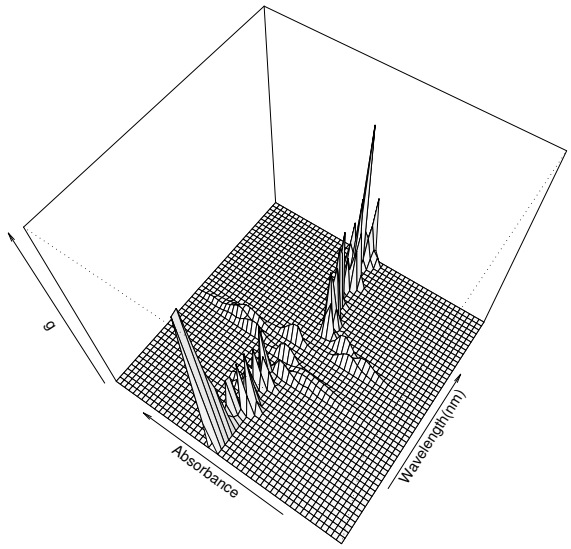


Figure 4.17: The estimated surface of densities $g_{\mathcal{X},\mathcal{D}}(t, y)$ for the first derivative of the Tecator data with high fat content (see fig. 2.21, p.28).

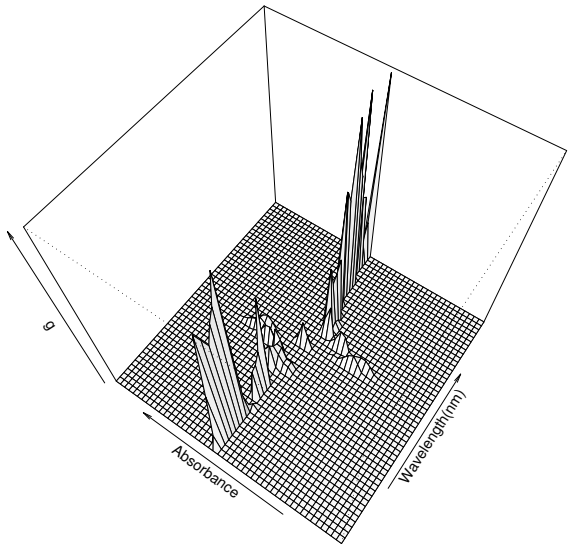


Figure 4.18: The estimated surface of densities $g_{\mathcal{X},\mathcal{D}}(t, y)$ for the first derivative of the Tecator data with low fat content (see fig. 2.21, p.28).

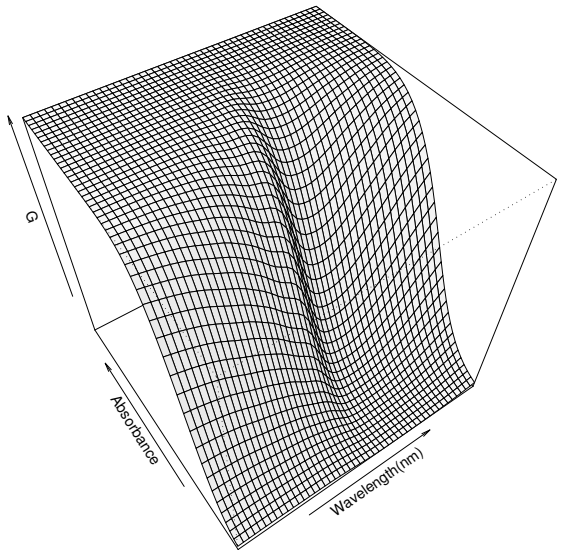


Figure 4.19: The normal surface of margins $G_{\mathcal{X},\mathcal{D}}(t, y)$ for the Tecator data (see fig. 2.11, p.21).

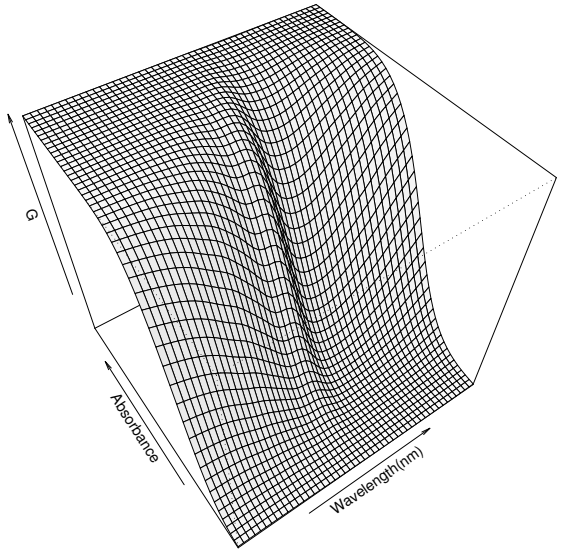


Figure 4.20: The normal surface of margins $G_{\mathcal{X},\mathcal{D}}(t, y)$ for the Tecator data with high fat content (see fig. 2.22, p.29).

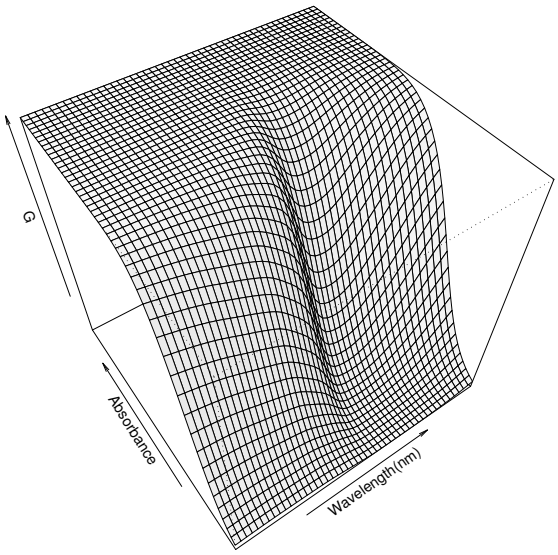


Figure 4.21: The normal surface of margins $G_{\mathcal{X},\mathcal{D}}(t, y)$ for the Tecator data with low fat content (see fig. 2.23, p.30).

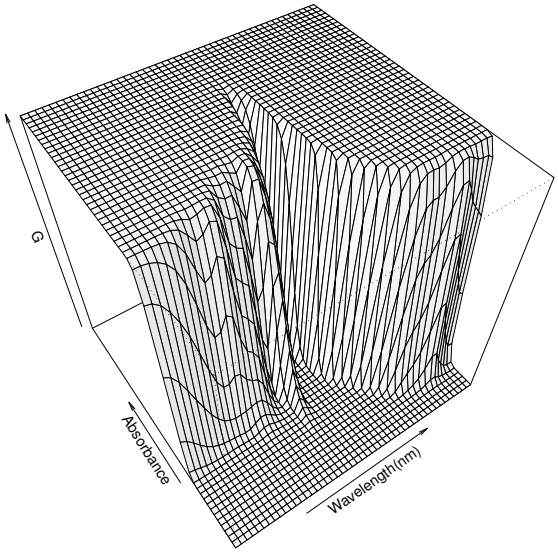


Figure 4.22: The normal surface of margins $G_{\mathcal{X},\mathcal{D}}(t, y)$ for the first derivative of the Tecator data (see fig. 2.20, p.28).

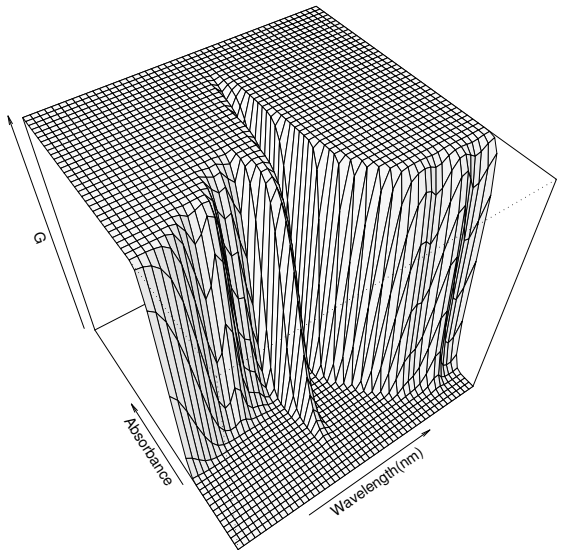


Figure 4.23: The normal surface of margins $G_{\mathcal{X},\mathcal{D}}(t, y)$ for the first derivative of the Tecator data with high fat content (see fig. 2.24, p.30).

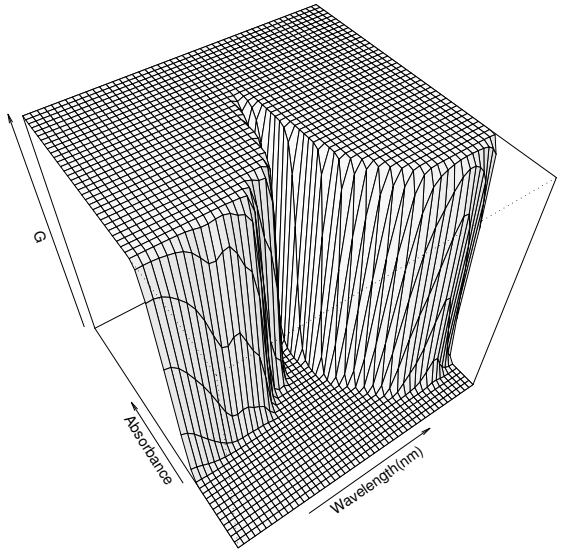


Figure 4.24: The normal surface of margins $G_{\mathcal{X},\mathcal{D}}(t, y)$ for the first derivative of the Tecator data with low fat content (see fig. 2.25, p.31).

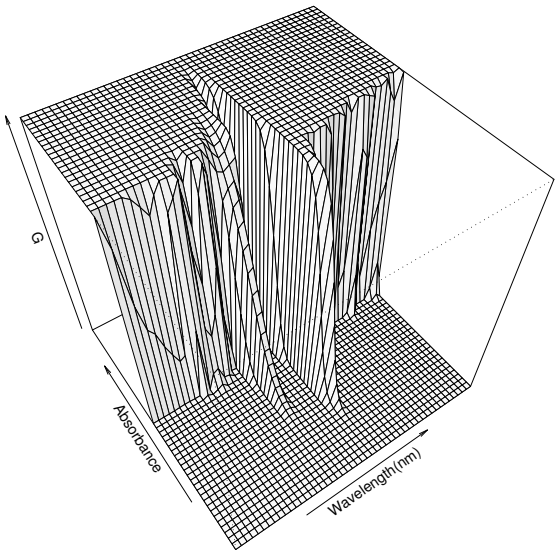


Figure 4.25: The normal surface of margins $G_{\mathcal{X},\mathcal{D}}(t, y)$ for the first derivative of the Tecator data (see fig. 2.21, p.28).

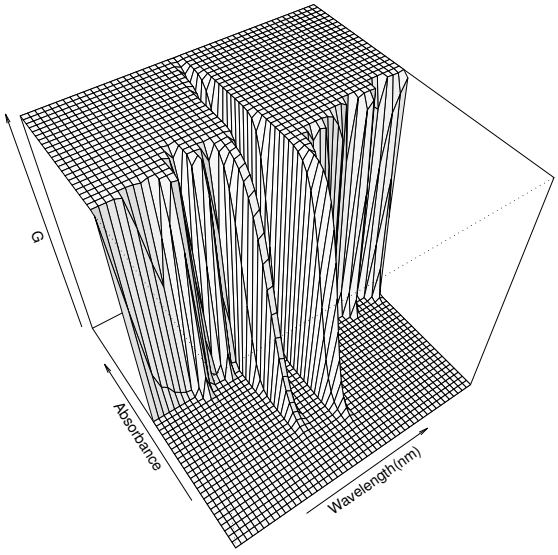


Figure 4.26: The normal surface of margins $G_{\mathcal{X},\mathcal{D}}(t, y)$ for the first derivative of the Tecator data with high fat content (see fig. 2.21, p.28).

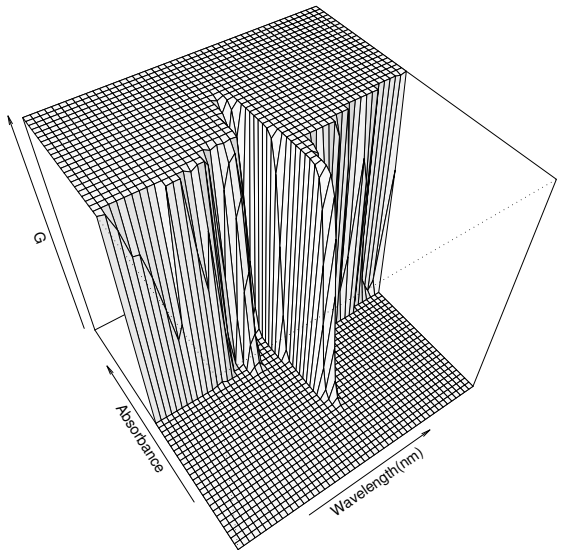


Figure 4.27: The normal surface of margins $G_{\mathcal{X},\mathcal{D}}(t, y)$ for the first derivative of the Tecator data with low fat content (see fig. 2.21, p.28).

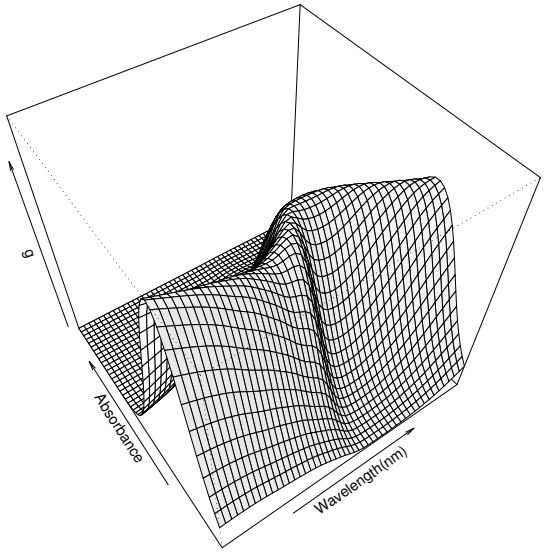


Figure 4.28: The normal surface of densities $g_{\mathcal{X},\mathcal{D}}(t, y)$ for the Tecator data (see fig. 2.11, p.21).

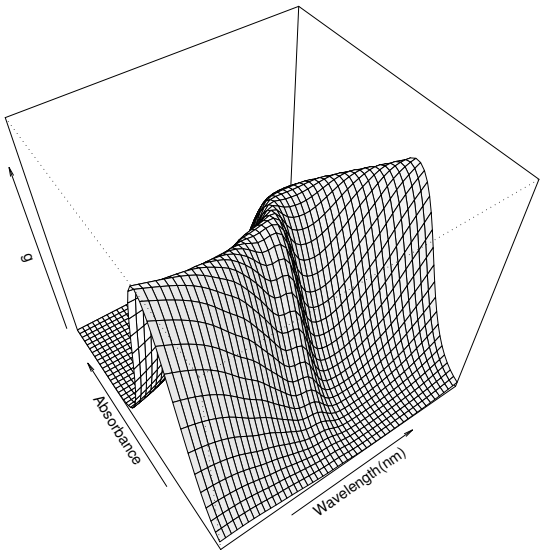


Figure 4.29: The normal surface of densities $g_{\mathcal{X},\mathcal{D}}(t, y)$ for the Tecator data with high fat content (see fig. 2.22, p.29).

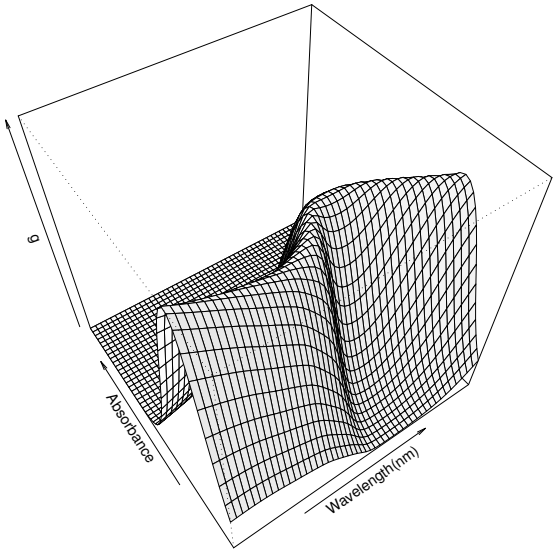


Figure 4.30: The normal surface of densities $g_{\mathcal{X},\mathcal{D}}(t, y)$ for the Tecator data with low fat content (see fig. 2.23, p.30).

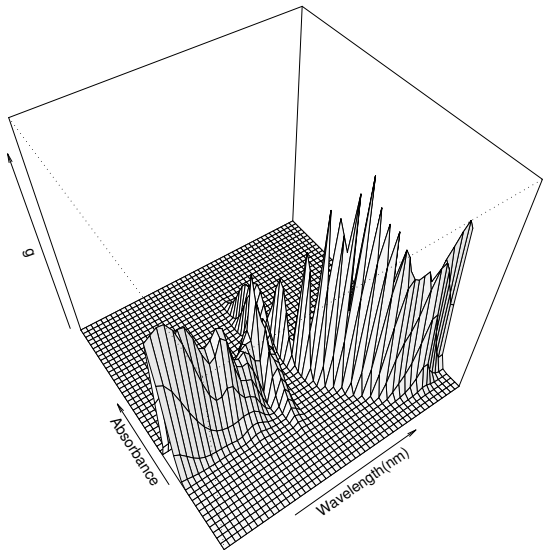


Figure 4.31: The normal surface of densities $g_{\mathcal{X},\mathcal{D}}(t, y)$ for the first derivative of the Tecator data (see fig. 2.20, p.28).

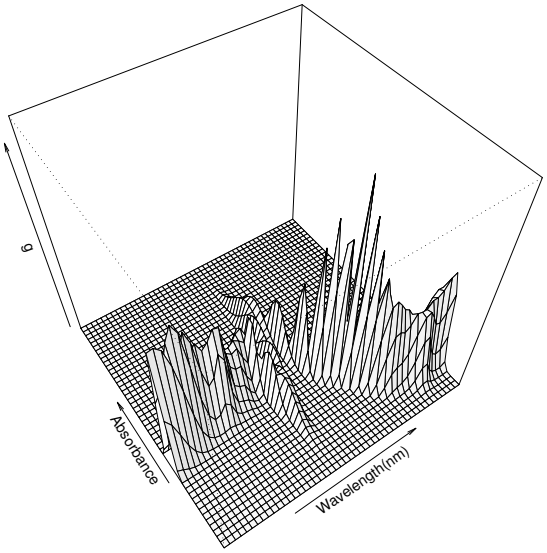


Figure 4.32: The normal surface of densities $g_{\mathcal{X},\mathcal{D}}(t, y)$ for the first derivative of the Tecator data with high fat content (see fig. 2.24, p.30).

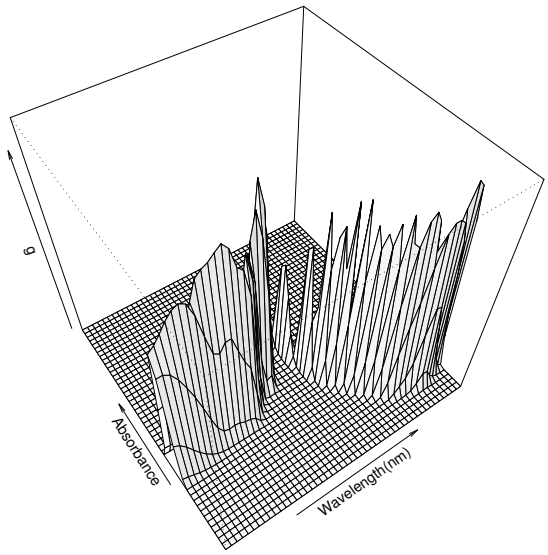


Figure 4.33: The normal surface of densities $g_{\mathcal{X},\mathcal{D}}(t, y)$ for the first derivative of the Tecator data with low fat content (see fig. 2.25, p.31).

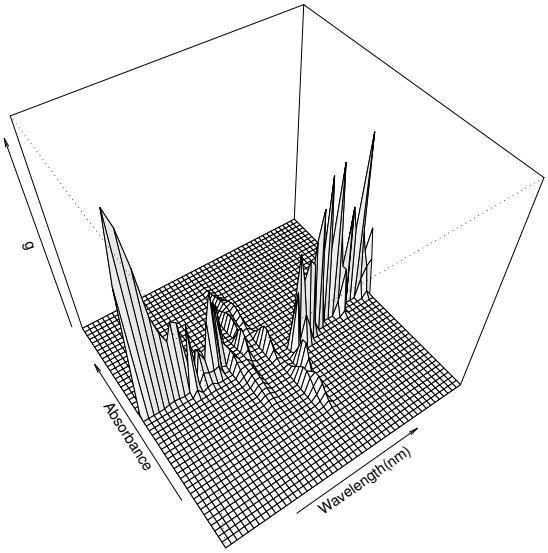


Figure 4.34: The normal surface of densities $g_{\mathcal{X},\mathcal{D}}(t, y)$ for the first derivative of the Tecator data (see fig. 2.21, p.28).

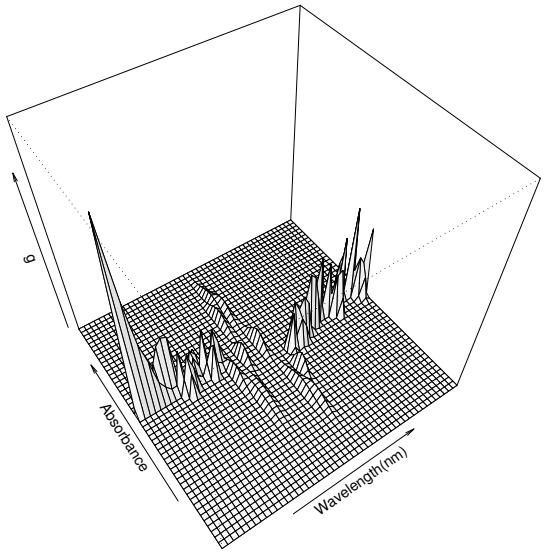


Figure 4.35: The normal surface of densities $g_{\mathcal{X},\mathcal{D}}(t, y)$ for the first derivative of the Tecator data with high fat content (see fig. 2.21, p.28).

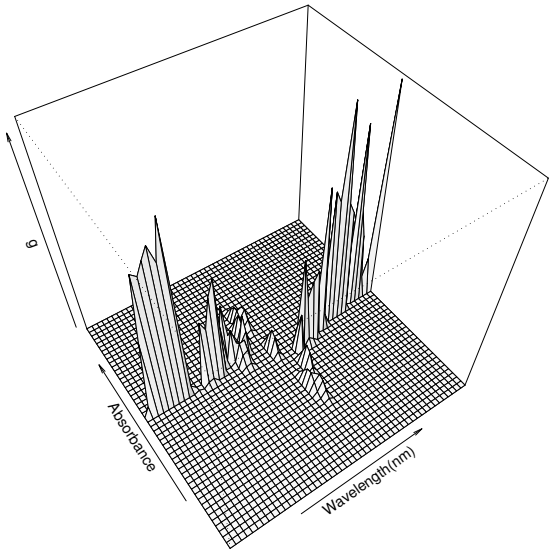


Figure 4.36: The normal surface of densities $g_{\mathcal{X},\mathcal{D}}(t, y)$ for the first derivative of the Tecator data with low fat content (see fig. 2.21, p.28).

Figures from 4.19 to 4.21 show the surfaces of margins $G_{\mathcal{X},\mathcal{D}}$ and the surfaces of densities for Tecator data using the normal distribution. The same order is use as in the case of the estimation case In the following we will alway use this function $G_{\mathcal{X},\mathcal{D}}$ in conjunction with a function u of $\mathcal{I}^{\mathcal{D}}$: $G_{\mathcal{X},\mathcal{D}}(t, u(t))$, so, for ease the notations, we define the following.

Definition 4.2.2. *Let $\mathcal{X}$ an* frv *with values in $\mathcal{I}^{\mathcal{D}}$. We define respectively the* functional surface of margins $\mathbf{G}_{\mathcal{X},\mathcal{D}}$ *and the* functional surface of densities $\mathbf{g}_{\mathcal{X},\mathcal{D}}$ *as follow:*

$$\mathbf{G}_{\mathcal{X},\mathcal{D}} : \mathcal{D} \times \mathcal{I}^{\mathcal{D}} \to [0,1] : (t,u) \mapsto \mathbf{G}_{\mathcal{X},\mathcal{D}}[t;u] = G_{\mathcal{X},\mathcal{D}}(t, u(t)), \tag{4.9}$$

$$\mathbf{g}_{\mathcal{X},\mathcal{D}} : \mathcal{D} \times \mathcal{I}^{\mathcal{D}} \to [0,1] : (t,u) \mapsto \mathbf{g}_{\mathcal{X},\mathcal{D}}[t;u] = g_{\mathcal{X},\mathcal{D}}(t, u(t)). \tag{4.10}$$

In the following we will write $\mathbf{G}_{\mathcal{X},\mathcal{D}}[t;u]$ for $G_{\mathcal{X},\mathcal{D}}(t, u(t))$ and $\mathbf{g}_{\mathcal{X},\mathcal{D}}[t;u]$ for $g_{\mathcal{X},\mathcal{D}}(t, u(t))$.

If θ is the parameter (number or vector) for $G_{\mathcal{X},\mathcal{D}}$ and $g_{\mathcal{X},\mathcal{D}}$, then we will use the following notation when it is necessary for disambiguation: $\mathbf{G}_{\mathcal{X},\mathcal{D},\theta}[t;u]$ and $\mathbf{g}_{\mathcal{X},\mathcal{D},\theta}[t;u]$.

For a given a surface of margins $G_{\mathcal{X},\mathcal{D}}$, we can consider the level curves of this surface: they are the "functional" quantiles of $G_{\mathcal{X},\mathcal{D}}$.

Definition 4.2.3. *Let $\mathcal{X}$ an* frv *with values in $\mathcal{I}^{\mathcal{D}}$, and surface of margins $G_{\mathcal{X},\mathcal{D}}$. A* functional quantile *of order $p \in [0,1]$ is any function $u \in \mathcal{I}^{\mathcal{D}}$ such:*

$$\mathbf{G}_{\mathcal{X},\mathcal{D}}[t;u] = p, \ \forall t \in \mathcal{D}. \tag{4.11}$$

We define also the lowest functional quantile $Q_{\mathcal{X},\mathcal{D},p}$ *by*

$$Q_{\mathcal{X},\mathcal{D},p}(t) = Q_{\mathcal{X}_t}(p) = \inf\{y \in \mathcal{I} : p \leq F_{\mathcal{X}_t}(y)\}. \tag{4.12}$$

In the following, a functional quantile will always be denoted using a capital letter with its related informations (*frv*, domain and order) in subscript:

- $U_{\mathcal{X},\mathcal{D},p}$ will be a functional quantile of order $p \in [0,1]$ for the *frv* $\mathcal{X}$ defined on $\mathcal{D}$,
- $V_{Y,[0,1],0.5}$ will be a functional quantile of order 0.5 for the *frv* $\mathcal{Y}$ defined on $[0,1]$,
- ...

However the letter Q will be used only for the concept of lowest functional quantile, in reference to the classical concept of quantile function (see definition 3.2.5, page 39).

Remark 4.2.1. *Of course we have the following equality*

$$\mathbf{G}_{\mathcal{X},\mathcal{D}}[t; U_{\mathcal{X},\mathcal{D},p}] = \mathbf{G}_{\mathcal{X},\mathcal{D}}[t; V_{\mathcal{X},\mathcal{D},p}] = p, \ \forall t \in \mathcal{D}, \tag{4.13}$$

but, this latter imply the equality between $U_{\mathcal{X},\mathcal{D},p}$ and $V_{\mathcal{X},\mathcal{D},p}$, if and only if for any $t \in \mathcal{D}$ the surface of margins $G_{\mathcal{X},\mathcal{D}}(t,.) = F_{\mathcal{X}_t}(.)$ is defined by a strictly increasing distribution (see definition 3.2.5, page 39). In this case the *lowest functional quantile* $Q_{\mathcal{X},\mathcal{D}}$ is the only *functional quantile* of order p, and then we only use the letter Q.

Figures from 4.37 to 4.45 show estimated quantiles for Tecator data in the same order that above, using the Normal distribution for the surface of margins $G_{\mathcal{X},\mathcal{D}}$. The chosen probability levels for the functional quantiles are the following: 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9 and 1.

It is important to understand that these functions do not belong necessarily to the original data set shown in figure 2.11, but are built using, for all $t \in \mathcal{D}$, $Q_{\mathcal{X}_t}(p)$.

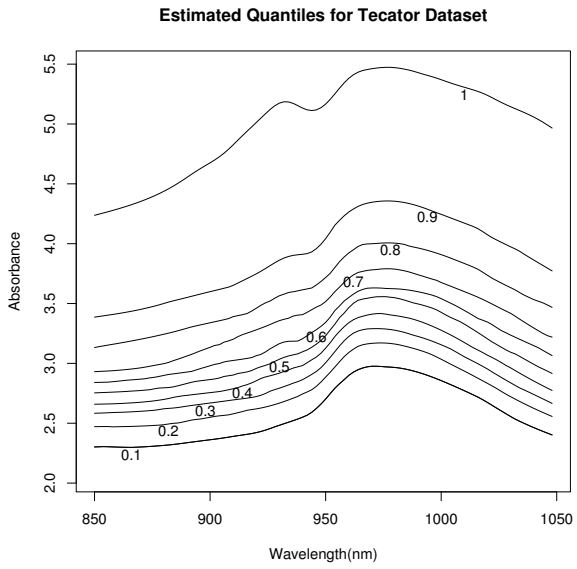


Figure 4.37: Estimated functional quantiles $Q_{\mathcal{X},\mathcal{D},p}(t)$ for the Tecator data (see fig. 2.11, p.21).

It is interesting to notice that these kind of quantiles functions are not completely new, figures 4.46 and 4.47 show two well known examples of percentiles curves for length and weight of children[1]. These curves are exactly functional quantiles.

[1]Data available on the website of the Centers for Disease Control and Prevention: `http://www.cdc.gov/nchs/about/major/nhanes/growthcharts/clinical_charts.htm`

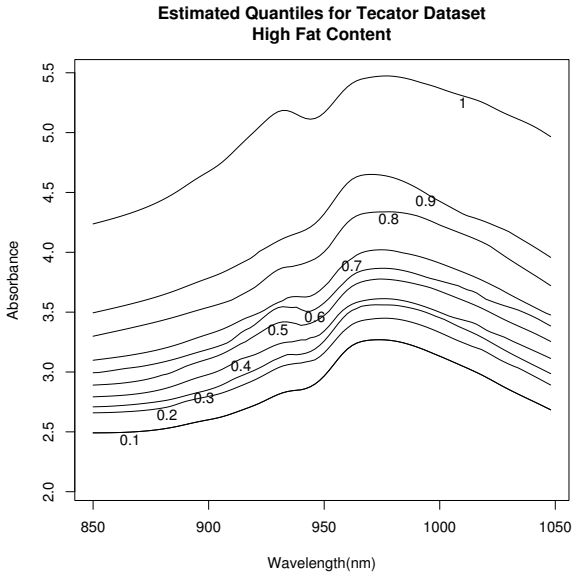


Figure 4.38: Estimated functional quantiles $Q_{\mathcal{X},\mathcal{D},p}(t)$ for the Tecator data with high fat content (see fig. 2.22, p.29).

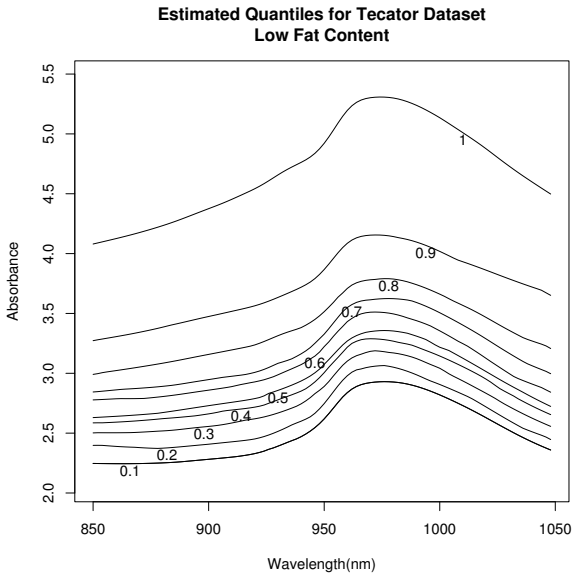


Figure 4.39: Estimated functional quantiles $Q_{\mathcal{X},\mathcal{D},p}(t)$ for the Tecator data with low fat content (see fig. 2.23, p.30).

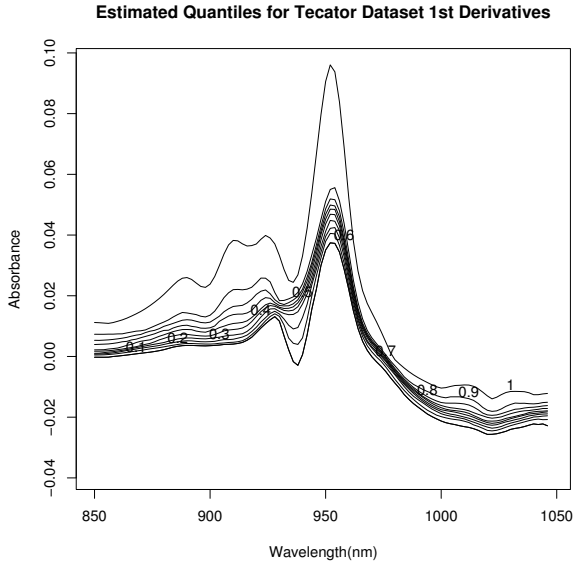


Figure 4.40: Estimated functional quantiles $Q_{\mathcal{X},\mathcal{D},p}(t)$ for the first derivative of the Tecator data (see fig. 2.20, p.28).

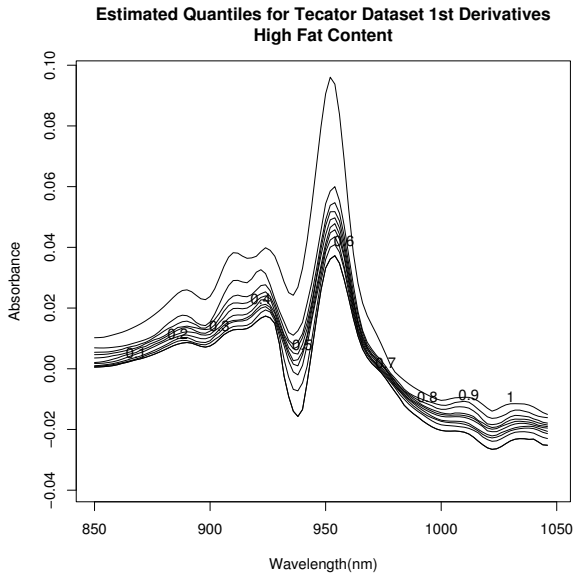


Figure 4.41: Estimated functional quantiles $Q_{\mathcal{X},\mathcal{D},p}(t)$ for the first derivative of the Tecator data with high fat content (see fig. 2.24, p.30).

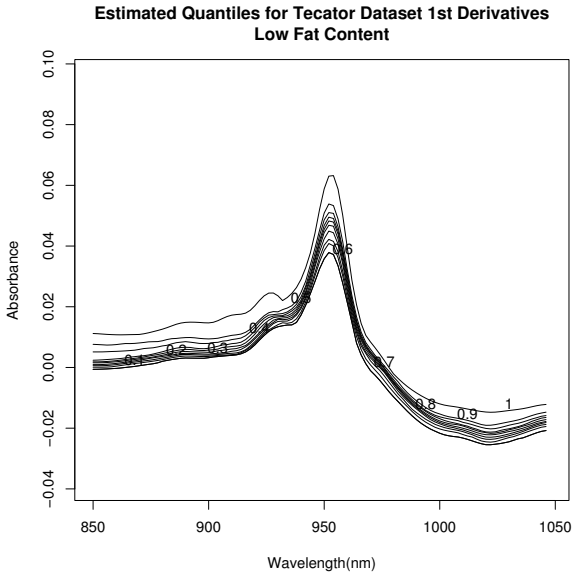


Figure 4.42: Estimated functional quantiles $Q_{\mathcal{X},\mathcal{D},p}(t)$ for the first derivative of the Tecator data with low fat content (see fig. 2.25, p.31).

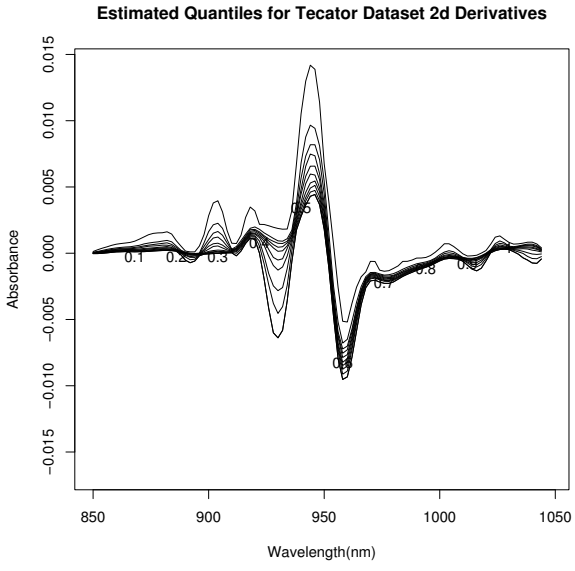


Figure 4.43: Estimated functional quantiles $Q_{\mathcal{X},\mathcal{D},p}(t)$ for the second derivative of the Tecator data (see fig. 2.21, p.28).

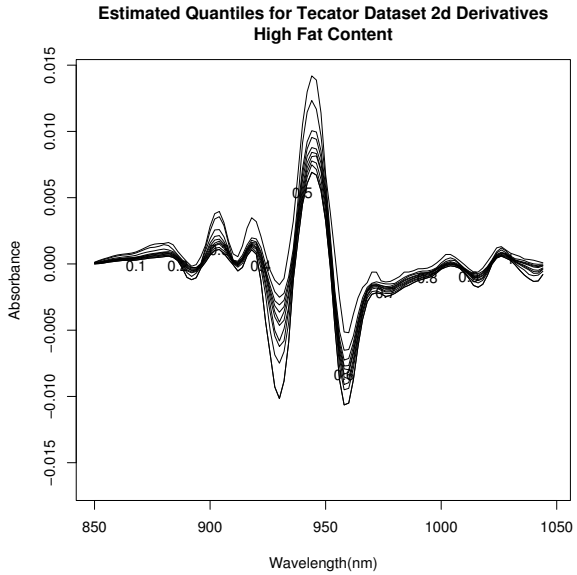


Figure 4.44: Estimated functional quantiles $Q_{\mathcal{X},\mathcal{D},p}(t)$ for the second derivative of the Tecator data with high fat content (see fig. 2.21, p.28).

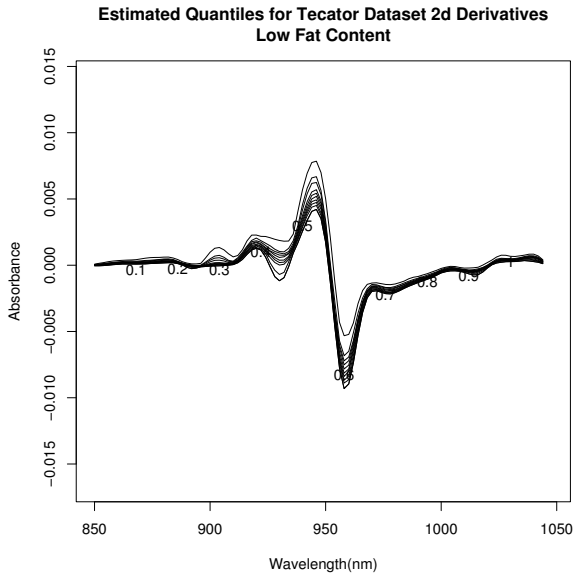


Figure 4.45: Estimated functional quantiles $Q_{\mathcal{X},\mathcal{D},p}(t)$ for the second derivative of the Tecator data with low fat content (see fig. 2.21, p.28).

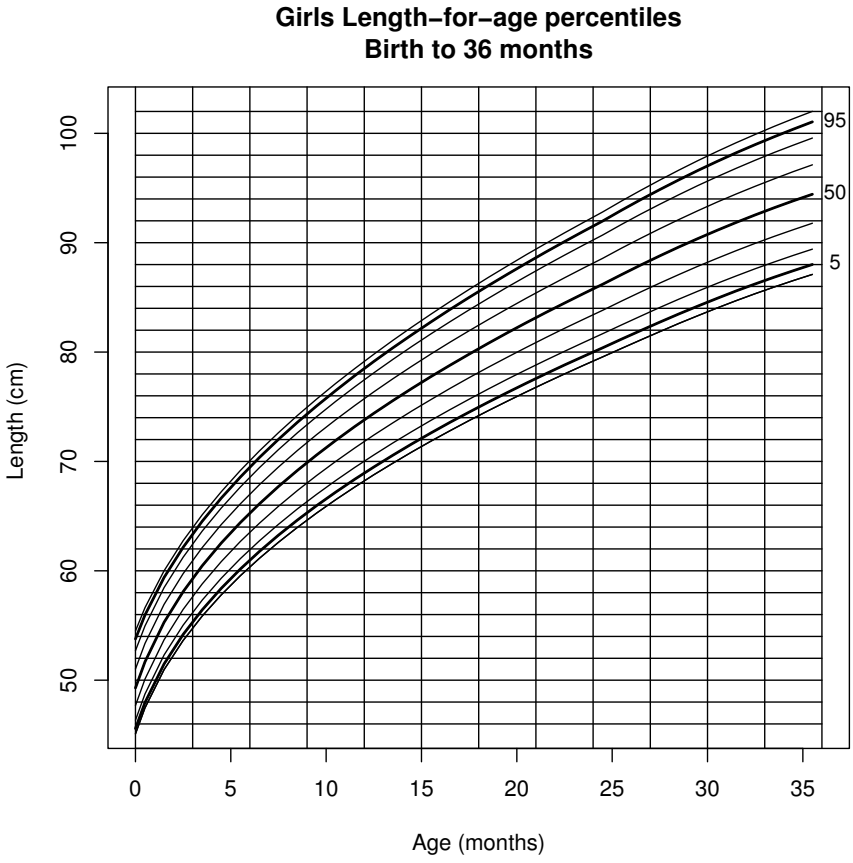


Figure 4.46: Birth to 36 months: Girls Length-for-age and percentiles.

Figure 4.47: 2 to 20 years: Boys Stature Weight-for-age and percentiles.

The *functional quantiles* are increasing following p.

Proposition 4.2.1. *Let $\mathcal{X}$ an* frv *with values in $\mathcal{I}^{\mathcal{D}}$, and surface of margins $G_{\mathcal{X},\mathcal{D}}$. If $p, q \in]0,1[$ then*

$$0 \leq p < q \leq 1 \Rightarrow U_{\mathcal{X},\mathcal{D},p}(t) < V_{\mathcal{X},\mathcal{D},q}(t) \quad \forall t \in \mathcal{D}, \tag{4.14}$$

where $U_{\mathcal{X},\mathcal{D},p}(t)$ is a functional quantile of order p and, $U_{\mathcal{X},\mathcal{D},p}(t)$ a functional quantile of order q.

The proof of this proposition is obvious. Using the surface of margins we can compute for any $t \in \mathcal{D}$ the margin of the joint distribution $H_{\mathcal{X}_{t_1},\ldots,\mathcal{X}_{t_n}}$, and thus we have almost our distribution: we need only to know the relation between the margins. And that what the copulas permit: capture the dependence relation between margins.

Original contribution(s) 2. *In this section we have defined the surface of margins and the surface of densities for a given* frv *$\mathcal{X}$. The definition of the surface of margins leads us to define functional quantiles. Functional quantiles are functions which are level curves of the surface of margins.*

4.3 Bivariate copulas

Even if copulas deal with real random variables they definition do not need immediately this concept, but some "increasing" properties [Nelsen, 1999].

Definition 4.3.1. *Let D_1 and D_2 two nonempty real subset, and let H a function of domain $DomH = D_1 \times D_2$. Let $B = [x_1, x_2] \times [y_1, y_2]$ be a rectangle included in $DomH$. The* H-volume of B *is given by*

$$V_H(B) = H(x_2, y_2) - H(x_2, y_1) - H(x_1, y_2) + H(x_1, y_1). \tag{4.15}$$

If we define the first order differences of H on B as

$$\Delta_{x_1}^{x_2} H(x, y) = H(x_2, y) - H(x_1, y)$$

$$\Delta_{y_1}^{y_2} H(x, y) = H(x, y_2) - H(x, y_1)$$

then the H-volume of B is the second order difference of H on B,

$$V_H(B) = \Delta_{y_1}^{y_2} \Delta_{x_1}^{x_2} H(x, y).$$

Definition 4.3.2. *A two dimensional real function H is* 2-increasing *if $V_H(B) \geq 0$ for all rectangles included in the domain of H.*

If H is 2-increasing, then its "density[2]" h will be positive or equal to zero.

The property "being 2-increasing" neither implies nor is implied by the property "being nondecreasing in each argument", but a link exist between these two properties [Nelsen, 1999].

Definition 4.3.3. *Let $B_1, B_2 \subset [0,1]$, and suppose that a_1, a_2 are the least elements respectively of B_1 and B_2. A real function H defined on $B_1 \times B_2$ is* grounded *if $H(x, a_2) = 0 = H(a_1, y)$ for all (x, y) in $B_1 \times B_2$.*

Lemma 4.3.1. *Let B_1 and B_2, two nonempty real subsets, and let H a 2-increasing and a grounded function with domain $B_1 \times B_2$. Then H is non decreasing in each argument.*

Now, we are able to define what is a copula function[Nelsen, 1999].

Definition 4.3.4. *A* two-dimensional copula *is a grounded and 2-increasing function $C : [0,1]^2 \mapsto [0,1]$ with the following additional property:*

$$C(u, 1) = u \quad \& \quad C(1, v) = v. \tag{4.16}$$

Even if the values taken of any copula are bounded by 0 and 1, we can determine, using the "2-increasing" and "grounded" properties, more precise bounds for a copula C [Nelsen, 1999].

Theorem 4.3.2. *Let C a copula, then $\forall u, v \in [0,1]$ we have the following property:*

$$max(u + v - 1, 0) \leq C(u, v) \leq min(u, v). \tag{4.17}$$

The above inequality is called the *Fréchet-Hoeffding bounds* inequality, and it defines the *Fréchet-Hoeffding bounds*. Another important copula which we will encounter is the *product copula*. Let us define these three important copulas formally:

- the *Fréchet-Hoeffding lower bound copula* is defined by the following expression:
$$W(u, v) = max(u + v - 1, 0), \tag{4.18}$$
- the *Fréchet-Hoeffding upper bound copula* is defined by the following expression:
$$M(u, v) = min(u, v), \tag{4.19}$$
- the *product copula* is defined by the following expression:
$$\Pi(u, v) = uv. \tag{4.20}$$

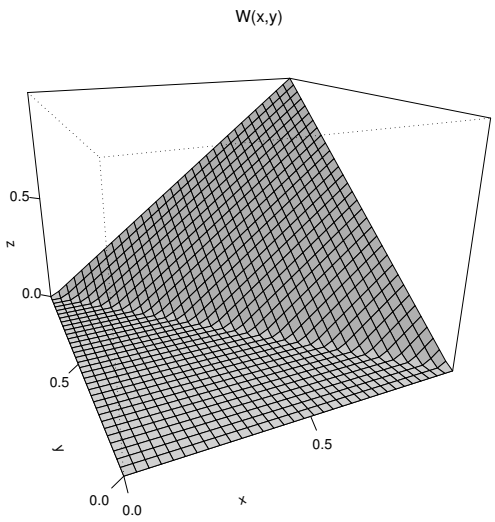


Figure 4.48: Graph of the Fréchet-Hoeffding lower bound copula

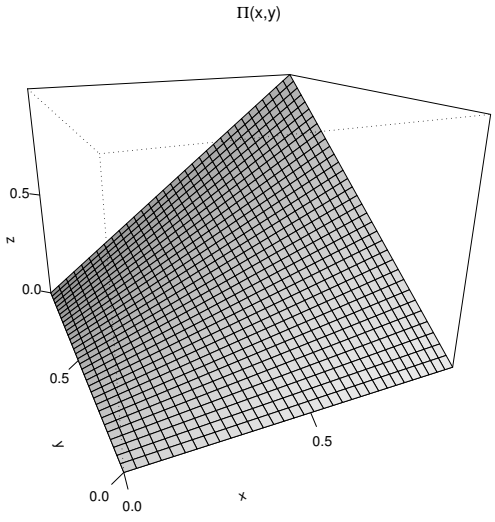


Figure 4.49: Graph of the Π copula

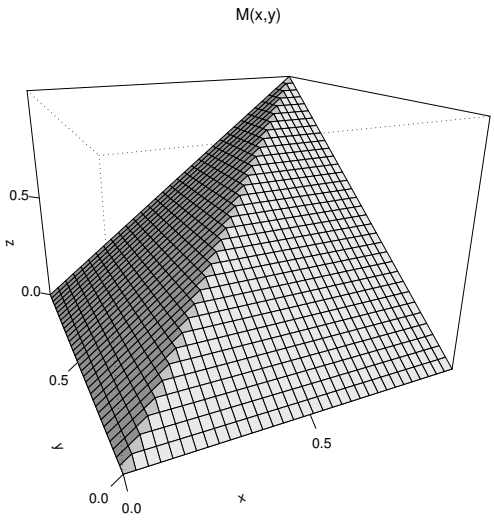


Figure 4.50: Graph of the Fréchet-Hoeffding upper bound copula

Figure 4.48 shows the graph of the Fréchet-Hoeffding lower bound, figure 4.49 shows the graph of the Fréchet-Hoeffding upper bound and figure 4.50 shows the graph of the product copula. These graphs show the surfaces the surface with respect to the equation $z = C(x, y)$ and lie in the unit cube $[0, 1]^3$. Inequality 4.17 implies that the graph of any copula lie also between the surfaces $z = W(x, y)$ and $z = M(x, y)$. It is not easy to see in a two dimensional static representation that the graph of the Π copula is a portion of the hyperbolic paraboloid $z = xy$. And, in general, it is not easy to give a good representation of the shape of a copula using only the 3-D representation. That is why another kind of representation is used, the *contour diagram* [Conway, 1979], where we plot the *level curves* given by $C(x, y) = t$ for some values of $t \in [0, 1]$. Figures 4.51, 4.52 and 4.53 show the contour diagram with 10 levels for the W, M and Π copulas.

It is a classical result that, if F is the cdf of a random variable X, then $U = F(X)$ is an uniform variable defined on $[0, 1]$.

Lemma 4.3.3. *If X is a continuous real random variable, and F its continuous and strictly increasing* cdf*, then the real random variable $U = F(X)$ is continuous and uniform on [0,1].*

[2]The "density" is used in a probabilistic sense, even if the present definition, function H is not necessarily a probability distribution.

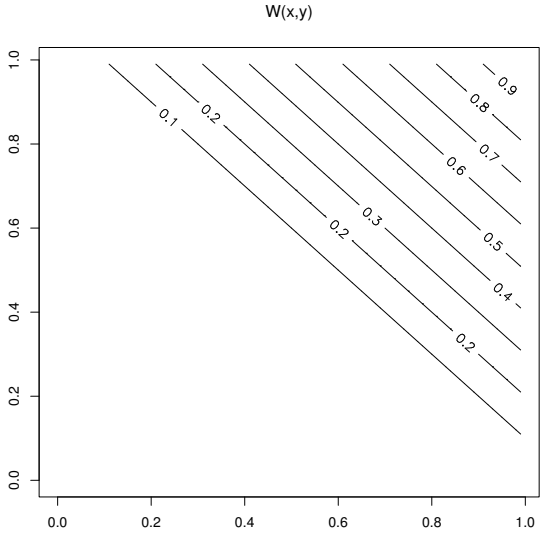


Figure 4.51: Contour diagram of the Fréchet-Hoeffding lower bound copula

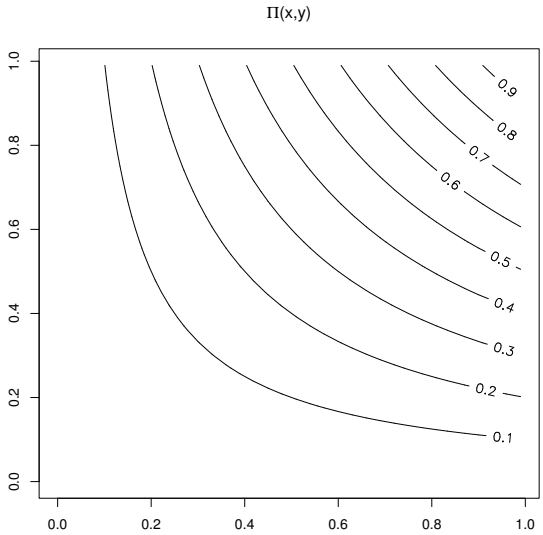


Figure 4.52: Contour diagram of the Π copula

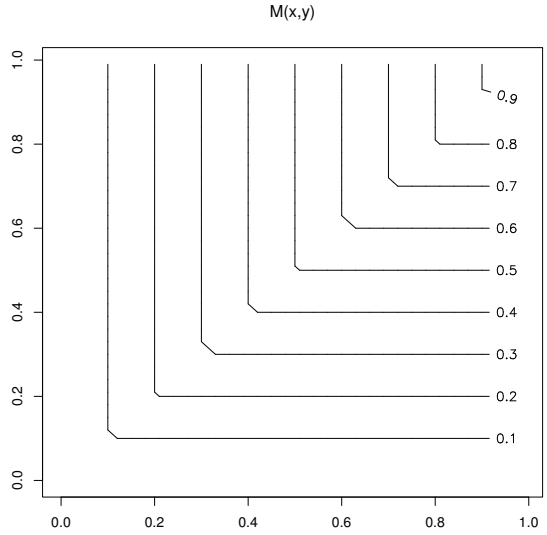


Figure 4.53: Contour diagram of the Fréchet-Hoeffding upper bound copula

Proof. If $Q = F^{-1}$, then

$$\begin{aligned} P[U \leq u] &= P[F(X) \leq u] \\ &= P[X \leq Q(u)] \\ &= F(Q(u)) \\ &= u. \end{aligned}$$

□

Then given two distributions functions F and G it is obvious that $C(F(x), G(y))$ defines a joint distribution, with given margins F and G:

$$\begin{aligned} C(F(x), G(\infty)) &= C(F(x), 1) && (4.21) \\ &= F(x). && (4.22) \end{aligned}$$

Copulas are, then, powerful tools to build joint distributions with given marginals. In this construction, choosing the copula C, is choosing a model for the dependence relation between the margins. Sklar (see [Sklar, 1959] and [Nelsen, 1999]) has shown the conversely: it is always possible to split a joint distribution into the margins in one hand and the dependence relation in the other hand.

Theorem 4.3.4 (Sklar's theorem)**.** *Let H be a joint distribution function with margins F and G. Then there exists a copula C such that for all $x, y \in \mathbb{R}$,*

$$H(x, y) = C(F(x), G(y)). \tag{4.23}$$

If F and G are continuous, then C is unique; otherwise, C is uniquely determined on Range of $F \times$ Range of G.
Conversely, if C is a copula and, F and G are distribution functions, then the function H defined by (4.23) *is a joint distribution function with margins F and G.*

This theorem first appeared in [Sklar, 1959]. The name copula came from the fact that the copula "couples" a joint distribution function to its univariate margins. In dividing a joint distribution in two very different components (margins and copula) it gives flexibility to the construction of joint distribution, and permits to separate the study of the relation between from the margins.

A priori, this theorem does not provide to us a mean to build a copula, but for a given probability distribution H the property given in expression (3.25) links H to its margins F and G in the following manner: $F(x) = \lim_{y \to +\infty} H(x, y)$ and $G(y) = \lim_{x \to +\infty} H(x, y)$. Then, for a given distribution H, we can always find its margins, thus we can (almost) always know its associated copula using the quantiles of the margins. This method is called the "inversion method" [Nelsen, 1999].

Theorem 4.3.5. *Let H a joint distribution, F and G its margins, and C its associated copula. If F and G are continuous and strictly increasing, then for any $u, v \in [0, 1]$*

$$C(u, v) = H\left(F^{-1}(u), G^{-1}(v)\right). \tag{4.24}$$

Thus, it is possible to extract the copula of well known joint distributions, as the normal distribution or the bivariate Gumbel logistic distribution [Roncalli, 2002] (see illustrations on figure 4.54).

Example 4.3.1. *Recall the normal bivariate distribution:*

$$\Phi_\rho(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \int_{-\infty}^{x} \int_{\infty}^{y} \exp\left(-\frac{z}{2(1-\rho^2)}\right) dt\ ds \tag{4.25}$$

where:

- $z = \dfrac{(s-\mu_x)^2}{\sigma_x} - \dfrac{2\rho(s-\mu_x)(t-\mu_y)}{\sigma_x\sigma_y} + \dfrac{(t-\mu_y)^2}{\sigma_y}$,
- $\rho = cor(x, y) = \dfrac{\sigma_{xy}}{\sigma_x\sigma_y}$.

And so, if the both margins are the standard distribution Φ, i.e. $\mu_x = \mu_y = 0$ and $\sigma_x = \sigma_y = 1$, we have the following distribution:

$$\Phi_\rho(x, y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{x} \int_{\infty}^{y} \exp\left(-\frac{s^2 - 2\rho st + t^2}{2(1-\rho^2)}\right) dt\ ds \tag{4.26}$$

and then the normal copula *with correlation ρ is given by*

$$C(x,y;\rho)=\frac{1}{2\pi\sqrt{1-\rho^2}}\int_{-\infty}^{\Phi^{-1}(x)}\int_{\infty}^{\Phi^{-1}(y)}\exp\left(-\frac{s^2-2\rho st+t^2}{2(1-\rho^2)}\right)dt\ ds. \tag{4.27}$$

Example 4.3.2. *The bivariate Gumbel logistic distribution is given by*

$$H(x,y)=\frac{1}{1+\exp(-x)+\exp(-y)}. \tag{4.28}$$

It is easy to find the two margins: $F(x)=H(x,+\infty)=(1+\exp(-x))^{-1}$ and $G(y)=H(+\infty,y)=(1+\exp(-y))^{-1}$. And thus the quantile functions are given by $F^{-1}(u)=\ln u-\ln(1-u)$ and $G^{-1}(v)=\ln v-\ln(1-v)$. We can then find the associated copula:

$$\begin{aligned} C(u,v) &= H(F^{-1}(u),G^{-1}(v)) \\ &= \left(1+\frac{1-u}{u}+\frac{1-v}{v}\right)^{-1} \\ &= \frac{uv}{u+v-uv}. \end{aligned} \tag{4.29}$$

Even if copulas link a joint probability distribution to its probability margins, the definition of copulas and the Sklar's theorem does not need the notion of probability or random variable, only the analytic and algebraic properties of the considered functions are needed. Now if we use the classical framework of random variables, we can restate the Copulas definition and the Sklar's theorem [Nelsen, 1999].

Definition 4.3.5. *A copula is a bivariate cumulative distribution function defined on the unit square $[0,1]^2$ such the two margins distributions are uniform on the interval $[0,1]$:*

$$C:[0,1]^2\to[0,1]:(u,v)\mapsto C(u,v).$$

If U and V denote the uniform univariate random variables of joint distribution C, then

$$C(u,v)=P[U\leq u;V\leq v].$$

Theorem 4.3.6 (Sklar's theorem)**.** *Let X and Y two real random variables with respectively distribution functions F and G, and joint distribution function H. Then there exists a copula C such* (4.23) *holds. If F and G are continuous, C is unique, otherwise C is uniquely determined on range of F $\times$ range of G.*

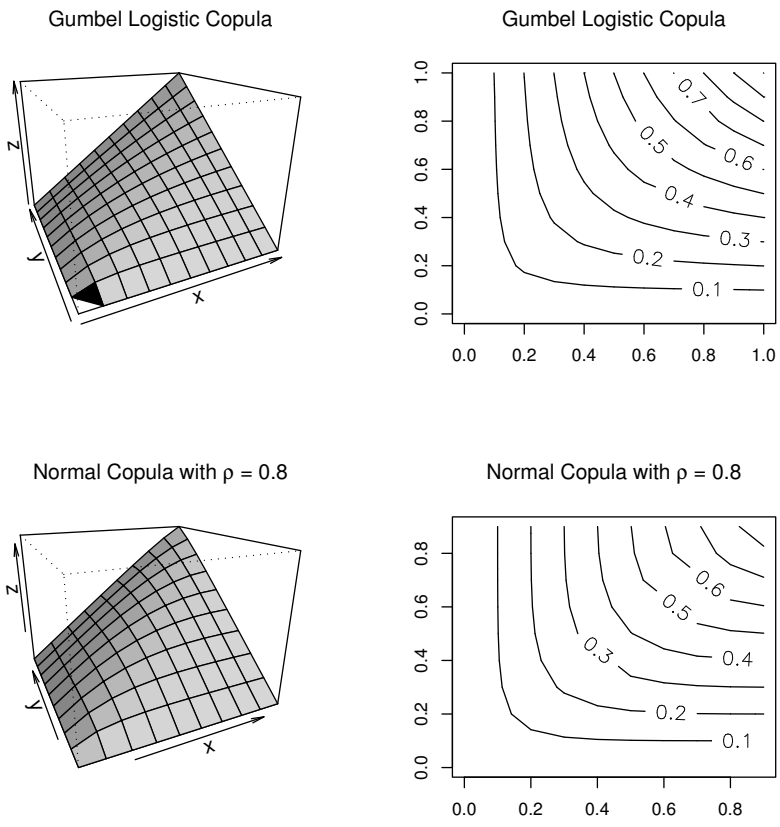


Figure 4.54: Graphs and contour diagrams of the Gumbel and the normal copulas.

Proof. We give here a proof of this theorem, in the continuous and strictly increasing case.

$$\begin{aligned} H(x,y) &= P[X \leq x; Y \leq y] \\ &= P[F(X) \leq F(x); G(Y) \leq G(y)] \\ &= P[U \leq F(x); V \leq G(y)] \\ &= C(F(x), G(y)) \end{aligned}$$

where $U = F(X)$ and $V = G(Y)$ are uniform by the lemma 4.3.3, and C is the joint distribution of U and V. This latter is given by the following expression and is unique

$$C(u,v) = H(F^{-1}(u), G^{-1}(v)).$$

□

The copula C, which link the two variables X and Y, is then denoted C_{XY}. In fact the copula captures the dependence structure of the distribution, and the following theorem shows that model of the dependence holds for strictly increasing transformations of the random variables [Nelsen, 1999].

Theorem 4.3.7. *Let X and Y two continuous random variables with copula C_{XY}. If α and β are two strictly increasing functions defined respectively on range of X and range of Y, then $C_{\alpha(X)\beta(Y)} = C_{XY}$. Thus C_{XY} is invariant under strictly increasing transformations of X and Y.*

For the sake of illustration, let us give the same results if α and β are not both strictly increasing [Nelsen, 1999].

Theorem 4.3.8. *Let X and Y two continuous random variables with copula C_{XY}. Let α and β, two monotone functions defined respectively on range of X and range of Y.*

1. *If α is strictly increasing and β strictly decreasing, then*

$$C_{\alpha(X)\beta(Y)}(u,v) = u - C_{XY}(u, 1-v).$$

2. *If α is strictly decreasing and β strictly increasing, then*

$$C_{\alpha(X)\beta(Y)}(u,v) = v - C_{XY}(1-u, v).$$

3. *If α and β are both strictly decreasing, then*

$$C_{\alpha(X)\beta(Y)}(u,v) = u + v - 1 + C_{XY}(1-u, 1-v).$$

We have previously introduced the three copulas W,Π and M, as important in the copulas theory, let us give three theorem which help us to understand the underlying dependence structures [Nelsen, 1999].

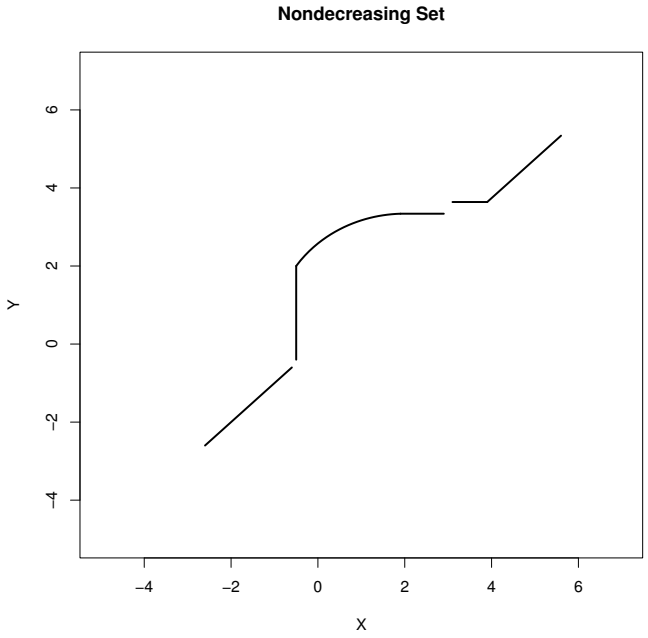


Figure 4.55: The graph of a nondecreasing set

Theorem 4.3.9. *Let X and Y be continuous random variables. Then X and Y are independent if and only if $C_{XY} = \Pi$.*

Although the presentation of the copulas can avoid the measure theory, we need some terminology and results from this field to give the interpretation of *Fréchet Hoeffding* bounds.

A joint distribution H induces a probability measure on $\mathbb{R}^2$: $V_H(]-\infty, x]\times]-\infty, y]) = H(x, y)$. Since a copula C is also a joint distribution with uniform margins on $]0, 1[$, it define also a measure on $[0, 1]^2$ by $V_C([0, u] \times [0, v]) = C(u, v)$ [Nelsen, 1999].

Definition 4.3.6. *Let a copula C. The* C-measure of a set *included in $[0, 1]^2$ is the C-volume V_C.*

Definition 4.3.7. *The support of a copula C is the smallest closed set whose complement with C-measure is zero.*

Definition 4.3.8. *A subset S of $\mathbb{R}^2$ is* non-decreasing *if for any $(u, v) \in S$, $x < u$ implies $y \leq v$. Similarly, a subset S of $\mathbb{R}^2$ is* non-increasing *if for any $(u, v) \in S$, $x < u$ implies $y \geq v$.*

The figure 4.55 gives an example of non-decreasing set. The two following theorem are the key to understand when two *rrv* have W or M as copula [Nelsen, 1999].

Theorem 4.3.10. *Let X and Y two random variables, and H their joint distribution. Then H is identically equal to its Fréchet-Hoeffding upper bound if and only if the support of H is a non decreasing subset of $\mathbb{R}^2$.*

Theorem 4.3.11. *Let X and Y two random variables, and H their joint distribution. Then H is identically equal to its Fréchet-Hoeffding lower bound if and only if the support of H is a non increasing subset of $\mathbb{R}^2$.*

Remark 4.3.1. *If X and Y are two random variables with joint distribution H and margins F and G respectively, then taking into account the Fréchet-Hoeffding inequality* (4.17) *with the Sklar's theorem, we get the following inequality*

$$\max\left(F(x)+G(y)-1,0\right) \leq H(x,y) \leq \min\left(F(x),G(y)\right). \tag{4.30}$$

Then, we are able to interpret the two above theorem. When X and Y are continuous, the support of H can have no horizontal or vertical segments, and in this case it is common to say that [Nelsen, 1999]:

- *Y* ***is almost surely a decreasing function of*** *X* ***if and only if the copula of X and Y is*** *W**,***
- *Y* ***is almost surely an increasing function of*** *X* ***if and only if the copula of X and Y is*** *M**.***

And thus, copulas can capture dependence structures of joint distributions, from the lower Fréchet lower bound *W to the* Fréchet upper bound*M, passing by the independence case,*Π.

Now, let H a joint distribution, F and G its margins, and C the associated copula. The joint density of H, noted h, is given by

$$h(x,y) = \frac{\partial^2 H(x,y)}{\partial x \partial y} = \partial^2_{12} H(x,y) \tag{4.31}$$

and of course the joint density of C, noted c, by

$$c(u,v) = \frac{\partial^2 c(u,v)}{\partial u \partial v} = \partial^2_{12} C(u,v). \tag{4.32}$$

Let us remark that the condition $V_C(B) \geq 0$ for any $B = [u_1,u_2]\times[v_1,v_2] \subset [0,1]^2$, can be written as follow

$$C(u_2,v_2) - C(u_1,v_2) - C(u_2,v_1) + C(u_1,v_1) \geq 0 \tag{4.33}$$

and is equivalent to positivity of the density:

$$c(u,v) = \partial^2_{12} C(u,v) \geq 0 \tag{4.34}$$

when this later exists. Now, using the Sklar's theorem we can write:

$$h(x,y) = c\left(F(x), G(y)\right) f(x) g(y) \tag{4.35}$$

with f and g the densities of respectively F and G. And, if F and G are continuous, we can deduce of this the joint density c:

$$c(u,v) = \frac{h\left(F^{-1}(u), G^{-1}(v)\right)}{f\left(F^{-1}(u)\right) g\left(G^{-1}(v)\right)} \tag{4.36}$$

Example 4.3.3. *Recall the example 4.27 of the normal copula. If we note*

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \tag{4.37}$$

and

$$\phi_\rho(x,y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{x^2 - 2\rho xy + y^2}{2(1-\rho^2)}\right). \tag{4.38}$$

Then knowing the bivariate density normal density and using the expression (4.36)*, we obtain:*

$$\begin{aligned} c\left(u,v;\rho\right) &= \frac{\phi_\rho\left(\Phi^{-1}(u), \Phi^{-1}(v)\right)}{\phi\left(\Phi^{-1}(u)\right)\phi\left(\Phi^{-1}(v)\right)} && (4.39)\\ &= \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{\left(s^2 - 2\rho st + t^2\right)}{2(1-\rho^2)} + \frac{\left(s^2+t^2\right)}{2}\right) && (4.40) \end{aligned}$$

where $s = \Phi^{-1}(u)$ *and* $t = \Phi^{-1}(v)$*.*

The copulas allow us to build non-conventional statistical models. Some joint distributions exist (normal, exponential, gamma, etc...), but in all cases all the margins have the same type [Kotz et al., 2000]. While, with copulas, it is possible to build a joint distribution with the normal copula, an exponential margin, and a beta for the other margin.

Now, for the sake of the illustration we are going to show some copulas densities and the corresponding joint densities using several margins. But before we must introduce what is a concordance measure and a famous of these measures: the *Kendall's tau*. This later will be helpful, when we will want compare different copula with the same measured concordance, but with different dependence structures [Nelsen, 1999].

Definition 4.3.9. *Let* X *and* Y *two random variables,* (x_1, y_1) *and* (x_2, y_2) *two observations of* (X, Y)*. We say that* (x_1, y_1) *and* (x_2, y_2) *are* concordant *if* $(x_1 - x_2)(y_1 - y_2) > 0$*. Similarly* (x_1, y_1) *and* (x_2, y_2) *are* discordant *if* $(x_1 - x_2)(y_1 - y_2) < 0$*.*

Definition 4.3.10. *Let X and Y two continuous random variables, whose copula is C. A numeric measure of association κ (also noted κ_C or $\kappa_{X,Y}$) between X and Y, is a* measure of concordance *if it satisfies the following properties:*

1. *κ is defined for every pair X,Y of continuous variables,*
2. *$-1 \leq \kappa_{X,Y} \leq 1$, $\kappa_{X,X} = 1$, and $\kappa_{X,-X} = -1$,*
3. *$\kappa_{X,Y} = \kappa_{Y,X}$,*
4. *if X and Y are independent, then $\kappa_{X,Y} = \kappa_\Pi = 0$,*
5. *$\kappa_{-X,Y} = \kappa_{X,-Y} = -\kappa_{X,Y}$,*
6. *if C_1 and C_2 are copulas such $C_1(u,v) \leq C_2(u,v)$ for all $u, v \in [0,1]$, then $\kappa_{C_1} \leq \kappa_{C_2}$,*
7. *if$\{X_n, Y_n\}$ is a sequence of continuous random variables with copulas C_n, and if $\{C_n\}$ converges pointwise to C, then $\lim_{n\mapsto\infty} \kappa_{C_n} = \kappa_C$.*

The *Kendall's tau* is a measure of concordance for which the sample version is defined as follow [Kruskal, 1958], [Hollander and Wolfe, 1973], [Lehmann, 1975]: let $\{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$ a random sample of n observations from a vector (X, Y) of continuous random variables. Let c the number of concordant pairs, and d the number of discordant pairs (one pair must be concordant or discordant). The Kendall's tau for the sample is defined by

$$t = \frac{c-d}{c+d} = \frac{c-d}{C_n^2}. \tag{4.41}$$

This measure is, also, the "probability of concordance minus the probability of discordance" for a pair of observation (x_i, y_i) and (x_j, y_j) randomly chosen in the sample. Thus, in a probabilistic way, we have the following definition [Nelsen, 1999].

Definition 4.3.11. *Let (X_1, Y_1) and (X_2, Y_2) be independent and identically distributed random vectors, each with distribution function H. The population version of* Kendall's tau *is the probability of concordance minus the probability of discordance:*

$$\tau = P\left[(X_1 - X_2)(Y_1 - Y_2) > 0)\right] - P\left[(X_1 - X_2)(Y_1 - Y_2) < 0)\right]. \tag{4.42}$$

Then it can be shown that τ depends only of the copulas of the pairs (X_1, Y_1) and (X_2, Y_2) [Nelsen, 1999].

Theorem 4.3.12. *Let X and Y be continuous random variables whose copula is C. Then the population version of* Kendall's tau *for X and Y is given by:*

$$\tau_{X,Y} = \tau_C = 4 \iint_{[0,1]^2} C(u,v)\ dC(u,v) - 1. \tag{4.43}$$

Example 4.3.4. *If C is the normal copula with correlation ρ, then it can be shown that we have the following relation [Cramér, 1971]:*

$$\tau_C = \frac{2}{\pi} \arcsin(\rho). \tag{4.44}$$

The figures 4.56 and 4.57 show the graphs and contour diagrams for the normal copula, with four values of the correlation ρ. Remark that the normal copula "looks like" W, Π and M, when ρ is respectively near -1, 0 and 1. Figures 4.58 and 4.59 show the graphs and contour diagrams for the density of the normal copula.

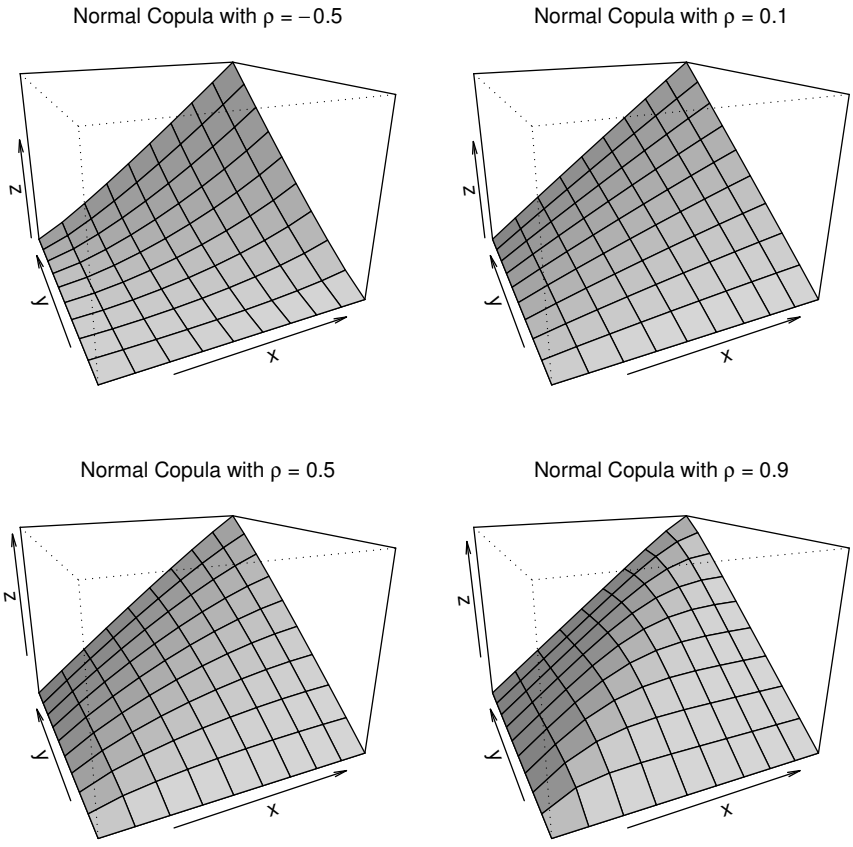


Figure 4.56: Graphs of the normal copula.

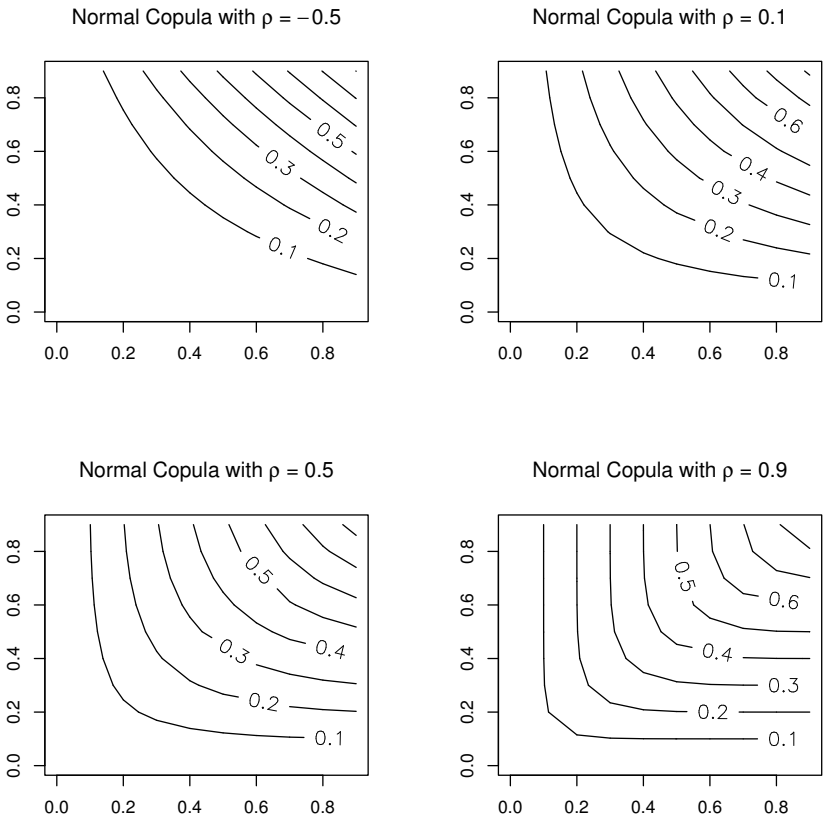


Figure 4.57: Contour diagrams of the normal copula.

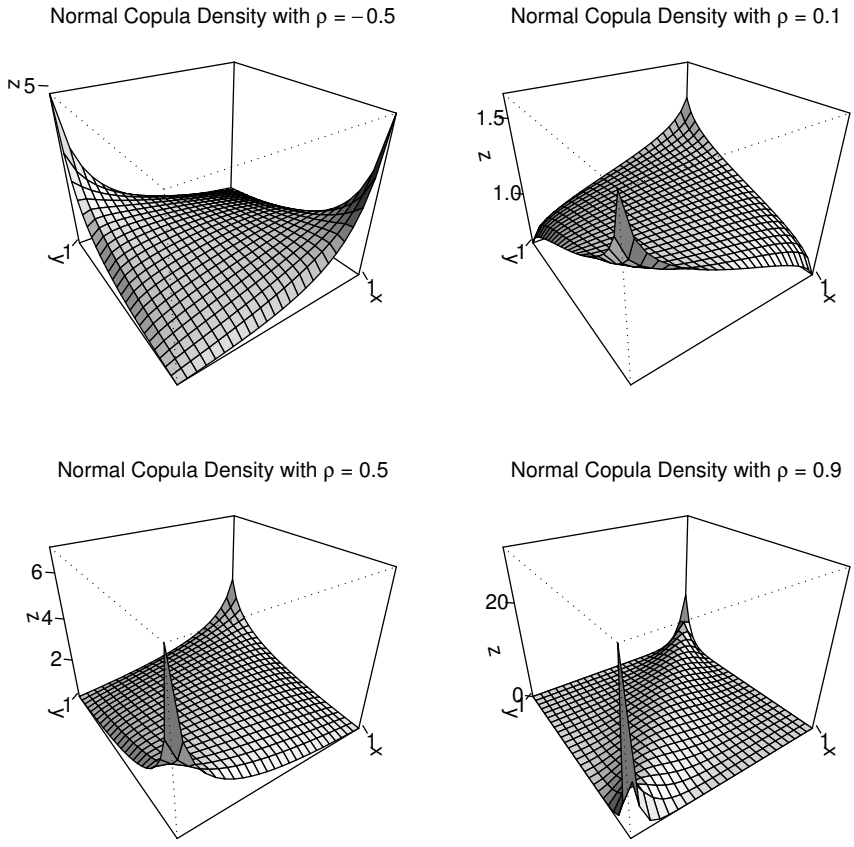


Figure 4.58: Graphs of the density of the normal copula.

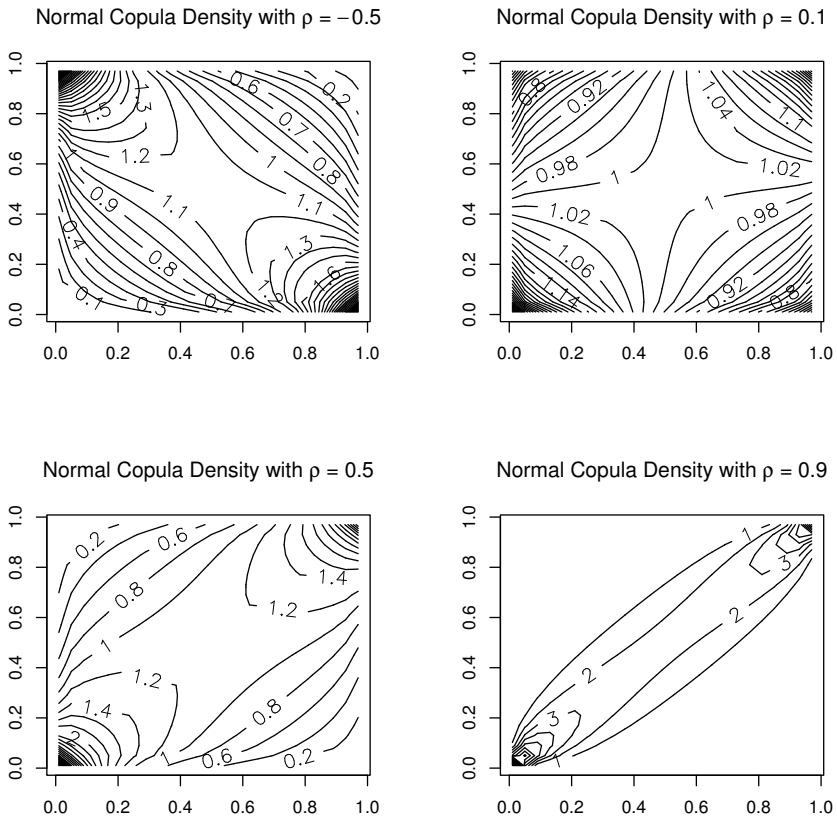


Figure 4.59: Contour diagrams of the density the normal copula.

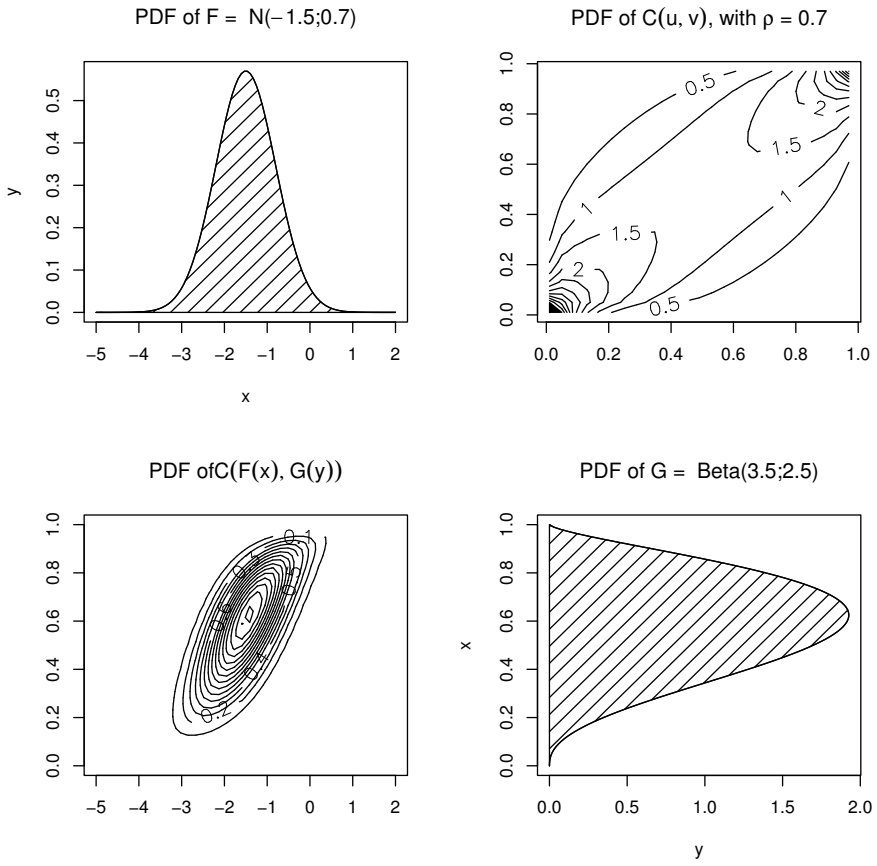


Figure 4.60: Contour diagram of a joint density built with a normal copula.

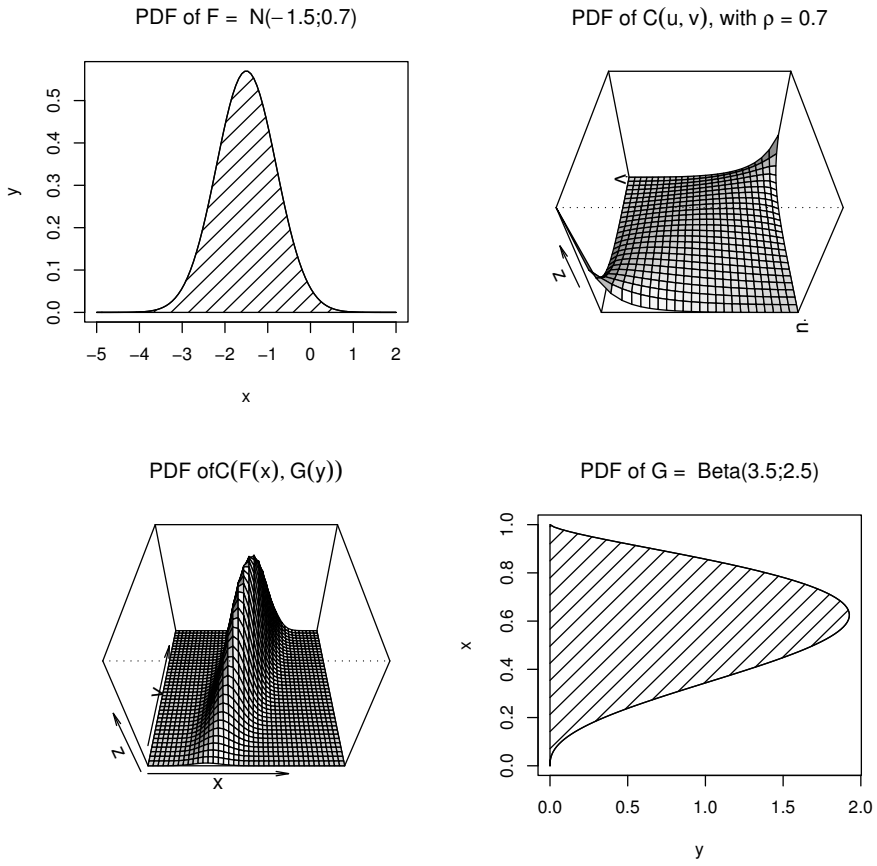


Figure 4.61: Graph of a joint density built with a normal copula.

And to illustrate the use of copulas for the case of distributions with given marginals, figures 4.60 and 4.61 show the densities of the normal copula, with $\rho = 0.7$, and two marginals. As first marginal we have a normal distribution with $\mu = -1.5$ and $\sigma = 0.7$. For second marginal we use a beta distribution with shape parameters 3.5 and 2.5. We have placed in the graph of density of the joint distribution in the left bottom, "surrounded" by the two margins, and in the top right the density of the copula. On the graph of the second margin we have done a rotation of 90^o for the ease of the reading. Figure 4.60 shows the two densities (joint distribution and copula) in a 3D view, while figure 4.61 use use the contour diagram of the densities.

4.4 Bivariate Archimedean Copulas

We have seen how the inversion method (4.24) allows us to build a copula, knowing the joint distribution and margins. Several other methods are described in [Nelsen, 1999](algebraic method, geometric method,...), but a last one based on Laplace transforms and mixtures powers of univariate CDFs gives a great family of copulas (see [Joe, 1997] for the complete details). Let us recall firstly the notion of convexity.

Definition 4.4.1. *A function f defined on a interval $[a, b]$ is convex if for any $x, y \in [a, b]$ and any $t \in [0, 1]$ we have*

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y) \tag{4.45}$$

moreover, if f is continuous on $[a, b]$, then f is convex if and only if

$$f\left(\frac{x+y}{2}\right) \leq \frac{f(x)+f(y)}{2}. \tag{4.46}$$

The function f is said to be concave if $-f$ is convex.

Let us recall also the following properties [Hardy et al., 1934].

Proposition 4.4.1. *If f is a convex function defined on the real interval $[a, b]$, then*

$$f\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) \leq \frac{1}{n}\sum_{i=1}^{n} f(x_i). \tag{4.47}$$

Proposition 4.4.2. *If a function f defined on a interval $[a, b]$ is convex, continuous and decreasing, then f^{-1} is convex on $[f(b), f(a)]$.*

Proof. Let $x', y' \in [f(b), f(a)]$, then there exist $x, y \in [a, b]$ such $x' = f(x)$ and $y' = f(y)$. By hypothesis, we have the relation (4.46), and as f is decreasing, we can write

$$\frac{x+y}{2} \geq f^{-1}\left(\frac{f(x)+f(y)}{2}\right)$$

and then

$$\frac{f^{-1}(x') + f^{-1}(y')}{2} \geq f^{-1}\left(\frac{x'+y'}{2}\right).$$

□

Let us start with the Laplace transfoms in the univariate case. Suppose that M is a univariate *cdf* of a positive random variable (so that $M(0) = 0$) and ψ be the *Laplace transform* of M, i.e.,

$$\psi(s) = \int_0^\infty e^{-sw}\, dM(w),\ s \geq 0. \tag{4.48}$$

For an arbitrary univariate *cdf* F, there exists a unique *cdf* $\mathbb{F}$ such that

$$F(x) = \int_0^\infty \mathbb{F}^\alpha(x) dM(\alpha) = \psi\left(-\log \mathbb{F}(x)\right) \tag{4.49}$$

thus, we have $\mathbb{F}(x) = \exp\{-\psi^{-1}\left(F(x)\right)\}$.
Now, consider two univariate *cdf* F and G, then we have

$$\mathbb{F}(x) = \exp\{-\psi^{-1}\left(F(x)\right)\}$$

and

$$\mathbb{G}(x) = \exp\{-\psi^{-1}\left(G(x)\right)\}.$$

As a product of *cdf* is always a *cdf*, then the following is a CDF in $\mathcal{F}\left(F, G\right)$:

$$\begin{aligned} \int_0^\infty \mathbb{F}(x)^\alpha \mathbb{G}(y)^\alpha dM(\alpha) &= \psi\left(-\log \mathbb{F}(x) - \log \mathbb{G}(y)\right) && (4.50)\\ &= \psi\left(\psi^{-1}\left[F(x)\right] + \psi^{-1}\left[G(y)\right]\right) && (4.51) \end{aligned}$$

and then the following is a *Archimedean copula*:

$$\begin{aligned} C\left(u, v\right) &= \psi\left(\psi^{-1}(u) + \psi^{-1}(v)\right) && (4.52)\\ &= \psi\left(\phi(u) + \phi(v)\right) && (4.53) \end{aligned}$$

with $\phi = \psi^{-1}$.

Of course, the function ϕ must be defined on $[0, 1]$ with image in $[0, \infty]$ or in a finite interval $[0, b]$, then to standardize the notations the concept of *pseudo inverse* on $[0, \infty]$ is introduced [Nelsen, 1999].

Definition 4.4.2. *Let ϕ be a continuous, strictly decreasing function from $[0, 1]$ into $[0, \infty]$ such that $\phi(1) = 0$. The* pseudo-inverse *of ϕ is the function $\phi^{[-1]}$ with domain equal to $[0, \infty]$ and range equal to $[0, 1]$, given by*

$$\phi^{[-1]}(t) = \begin{cases} \phi^{-1}(t), & 0 \leq t \leq \phi(0), \\ 0, & \phi(0) \leq t \leq \infty. \end{cases} \tag{4.54}$$

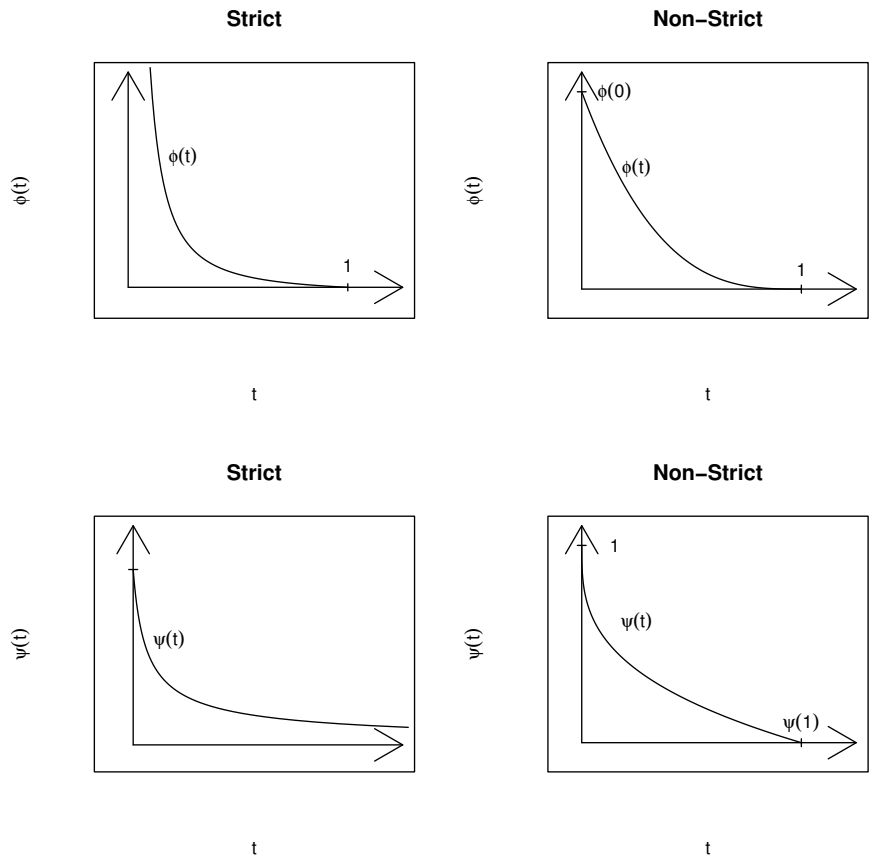


Figure 4.62: Examples of strict and non strict generators (ϕ) and their inverses (ψ).

Now, as Laplace transform, are not always easy to hand, let us give a more easy condition for the function ϕ to define an Archimedean copula [Nelsen, 1999].

Theorem 4.4.3. *Let ϕ be a continuous, strictly decreasing function from $[0,1]$ to $[0,\infty]$ such that $\phi(1)=0$, and $\phi^{[-1]}$ be the pseudo-inverse of ϕ. Then*

$$C(u,v)=\phi^{[-1]}\left(\phi(u)+\phi(v)\right) \tag{4.55}$$

is a copula if and only if ϕ is convex.

The function ϕ is called the *generator of the copula*. If $\phi(0)=\infty$, we say that ϕ is a *strict generator*. The figure 4.62 illustrates generators and their quasi-inverses in the strict and non-strict cases.

Archimedean copulas can be regarded as commutative binary operation in $[0,1]$ as the following theorem illustrate it [Nelsen, 1999].

Theorem 4.4.4. *Let C be an Archimedean copula with generator ϕ. Then:*

1. *C is symmetric (commutative), i.e. $C(u,v) = C(v,u),\ \forall u,v \in [0,1]$;*
2. *C is associative,i.e. $C\left(C(u,v),w\right) = C\left(u,C(v,w)\right),\ \forall u,v,w \in [0,1]$;*
3. *If $C \geq 0$ is any constant, then $c\phi$ is also a generator of C.*

Moreover, if we consider bivariate reducible functions [3], then the expression (4.4.3) is the general solution for associativity functional equation ([Aczel, 1966b]).

Theorem 4.4.5. *Let $u, v \in \mathbb{R}$ and F be a bivariate function. If with u and v, $F(u,v)$ also always lies in a given (possibly infinite) interval and $F(u,v)$ is reducible on both sides,i.e. $F(t,u) = F(t,v)$ or $F(u,w) = F(v,w)$ only if $u = v$, then*

$$F(x,y) = f\left[f^{-1}(x) + f^{-1}(y)\right] \tag{4.56}$$

(with continuous and strictly monotonic f) is the general continuous solution of the associativity functional equation:

$$F\left[F(x,y),z\right] = F\left[x,F(x,y)\right]. \tag{4.57}$$

This link between associativity and Archimedean copulas is also stated by another theorem, due to [Ling, 1965].

Theorem 4.4.6. *Let C be an associative copula such $C(u,u) < u,\ \forall u \in [0,1]$. Then C is Archimedean.*

The following examples shows a sample of Archimedean copulas, their generators and the limiting cases. We have chosen these four copulas because they are also copulas when $n > 2$, which is not always true for all Archimedean generators for bivariate copulas. A list of 22 Archimedean copulas can be found in [Nelsen, 1999].

Example 4.4.1. *The Clayton family, has the following generator*

$$\phi_\theta(t) = t^{-\theta} - 1. \tag{4.58}$$

The function ϕ is convex when $\theta \in [-1,\infty[\backslash\{0\}$, and the copula is given by:

$$C_\theta(u,v) = \max\left(\left[u^{-\theta} + v^{-\theta} - 1\right]^{-\frac{1}{\theta}}, 0\right). \tag{4.59}$$

The limiting cases are the following: $C_{-1} = W$,$C_0 = \Pi$, $C_\infty = M$.

[3] a bivariate function $F(u,v)$ is reducible on both sides if $F(t,u) = F(t,v)$ or $F(u,w) = F(v,w)$ only if $u = v$.

Example 4.4.2. *The Frank family, has the following generator*

$$\phi_\theta(t) = -\ln\left(\frac{e^{-\theta t}-1}{e^{-\theta}-1}\right). \tag{4.60}$$

The function ϕ is convex when $\theta \in]-\infty,\infty[\backslash\{0\}$, and the copula is given by:

$$C_\theta(u,v) = -\frac{1}{\theta}\left(1+\frac{(e^{-\theta u}-1)(e^{-\theta v}-1)}{(e^{-\theta}-1)}\right). \tag{4.61}$$

The limiting cases are the following: $C_{-\infty} = W$, $C_0 = \Pi$, $C_\infty = M$.

Example 4.4.3. *The Gumbel-Hougaard family, has the following generator*

$$\phi_\theta(t) = (-\ln t)^\theta. \tag{4.62}$$

The function ϕ is convex when $\theta \in [1,\infty[$, and the copula is given by:

$$C_\theta(u,v) = \exp\left(-\left[(-\ln u)^\theta + (-\ln v)^\theta\right]^{1/\theta}\right). \tag{4.63}$$

The limiting cases are the following: $C_1 = \Pi$, $C_\infty = M$.

Example 4.4.4. *The Joe family, has the following generator*

$$\phi_\theta(t) = -\ln\left[1-(1-t)^\theta\right]. \tag{4.64}$$

The function ϕ is convex when $\theta \in [1,\infty[$, and the copula is given by:

$$C_\theta(u,v) = 1-\left[1-\left(1-(1-u)^\theta\right)\left(1-(1-v)^\theta\right)\right]^{1/\theta}. \tag{4.65}$$

The limiting cases are the following: $C_1 = \Pi$, $C_\infty = M$.

We can see that Clayton and Frank's copulas can cover dependence structures from W to M, passing by Π, while Gumbel and Joe's copulas cover only from Π to M. Covering dependence structures from W to M passing by Π does not mean that all dependence structures in this range are covered. It is not easy to have a clear idea of the "space of dependence structures", and thus to illustrate the differences between different copulas we are going to shows the graphs and contour diagrams of the different copulas with the same Kendall's tau, i.e. with the same concordance "rate". We have reported that a formula exists which link the Kendall's tau to the correlation ρ for the normal copula (4.44). For the ease of the comparison let us give the formula ([Genest and MacKay, 1986a] and [Genest and MacKay, 1986b]) for the Archimedean copulas.

Theorem 4.4.7. *Let X and Y be random variables with an Archimedean copula C generated by ϕ. Then*

$$\tau_C = 1 + 4\int_0^1 \frac{\phi(t)}{\phi'(t)}dt. \tag{4.66}$$

Example 4.4.5. *For the* Clayton*'s copula, as $\phi(t) = t^{-\theta} - 1$ we have*

$$\frac{\phi(t)}{\phi'(t)} = \theta[t^{\theta+1} - t]$$

and thus it is easy to see that

$$\tau_C = \frac{\theta}{\theta + 2}. \tag{4.67}$$

And in this case the values of τ belong to $[-1, 1] \setminus \{0\}$, which is coherent with the limiting cases shown in example 4.4.3.

Example 4.4.6. *For the* Gumbel-Hougaard*'s copula, as $\phi(t) = (-\ln t)^\theta$, we have*

$$\frac{\phi(t)}{\phi'(t)} = \frac{t \ln t}{\theta}$$

and thus

$$\tau_C = 1 - \frac{1}{\theta} \tag{4.68}$$

and then τ belong to $[0, 1]$.

Example 4.4.7. *For the* Frank*'s copula the Kendall's tau is given by the following expression [Genest, 1987]:*

$$\tau = 1 + \frac{4}{\theta}\left[D_1(\theta) - 1\right] \tag{4.69}$$

where D_k is the Debye's function

$$D_k(x) = \frac{k}{x^k}\int_0^x \frac{t^k}{e^t - 1}dt.$$

Figures from 4.63 to 4.74 show the Normal, Frank, Clayton and Gumbel-Hougaard copulas, each of them plotted using their parameters to have a Kendall's tau equal respectively to 0.2, 0.5 and 0.8. On each figure we give, on the top part, the graph and the contour diagram of the copula with the selected parameter, and on the bottom part we give the graph and the contour diagram of the joint distribution built with this copula coupling two gaussian margins. Of course any other margins can be used, but acting as that, it is more easy to compare the resulting joint distribution to the well known normal case.

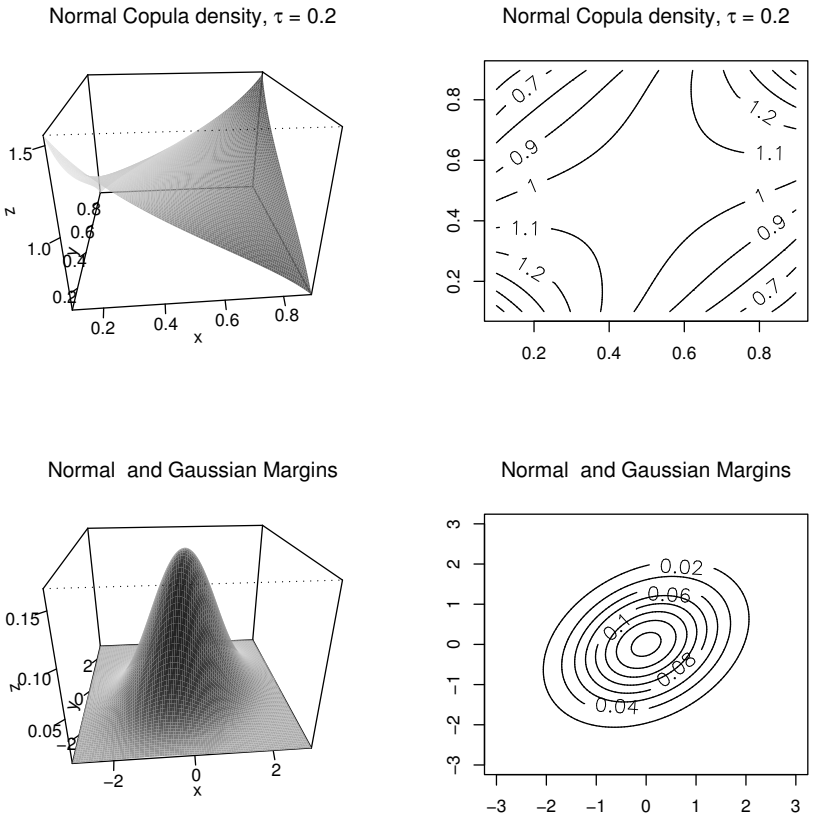


Figure 4.63: Top: The Normal copula with $\tau = 0.2$ and $\rho = 0.31$. Bottom: The same copula using Gaussian margins.

If we take a look at the shapes of these figures, we can see that the Frank copula have a shape more or less close to the normal copula. Although we can see that the curvature of the level curves in the contour diagrams are more important for the normal copula that for the Frank copula, which explain the more quadrangular aspect for the joint distribution build with the later. The Clayton and the Gumbel-Hougaard copulas have their maximum probabilities when the margins are, respectively, in $(0, 0)$ and $(1, 1)$. That gives, when we use gaussian margins, two pear-shaped joint densities, theshapes being in "opposites directions" between the two copulas.

4.5 Multivariate copulas

In the following we will use the notation $\mathbf{u} = (u_1, \ldots, u_n)$. Although some difference exist between multivariate copulas and bivariate copulas, the main

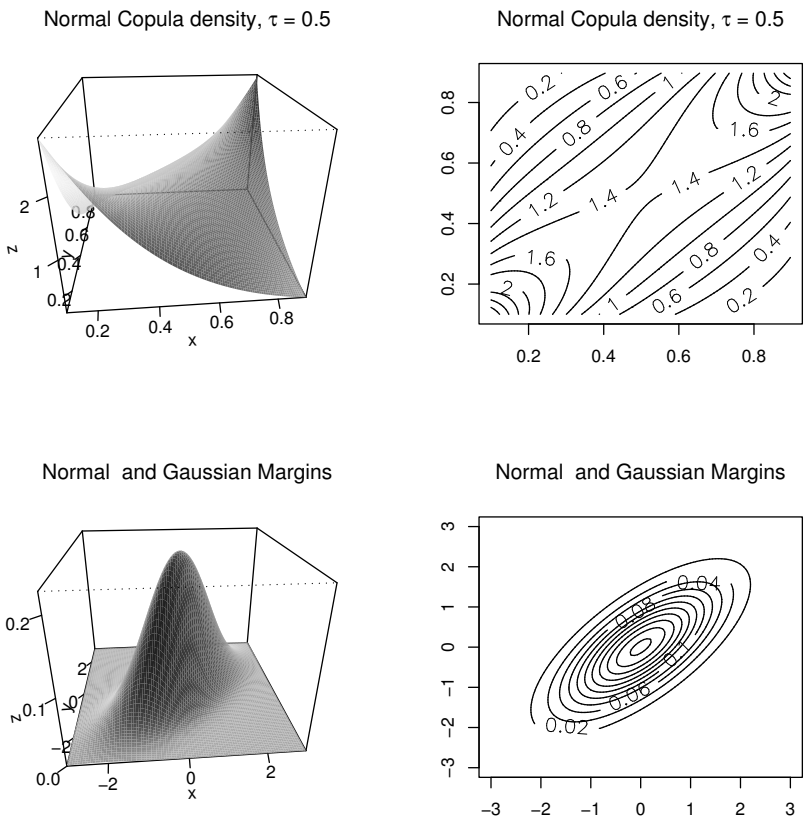


Figure 4.64: Top: The Normal copula with $\tau = 0.5$ and $\rho = 0.71$. Bottom: The same copula using Gaussian margins.

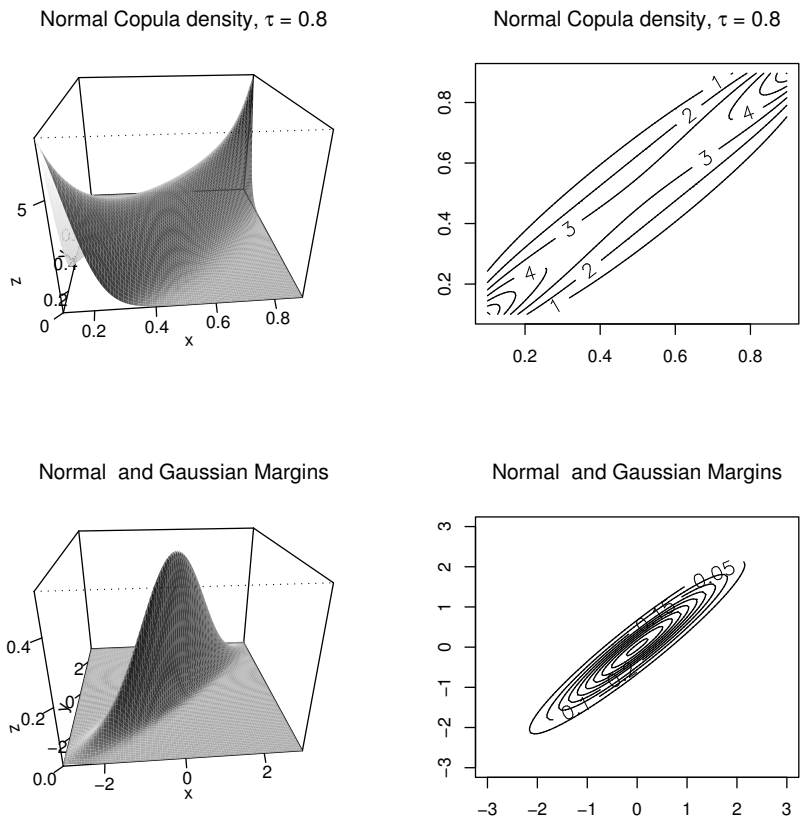


Figure 4.65: Top: The Normal copula with $\tau = 0.8$ and $\rho = 0.95$. Bottom: The same copula using Gaussian margins.

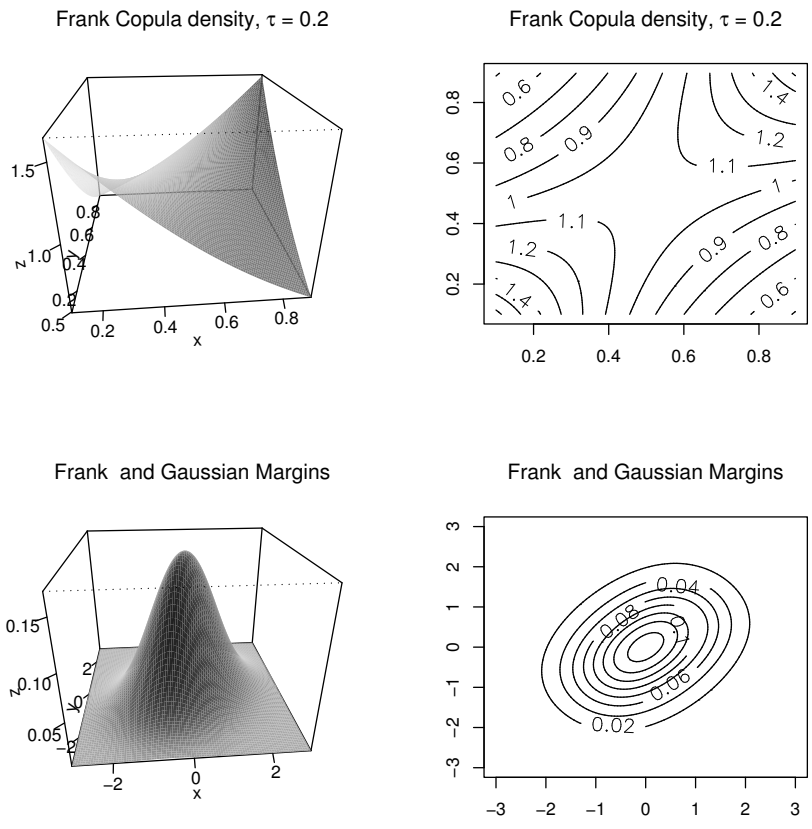


Figure 4.66: Top: The Frank copula with $\tau = 0.2$ and $\theta = 1.86$. Bottom: The same copula using Gaussian margins.

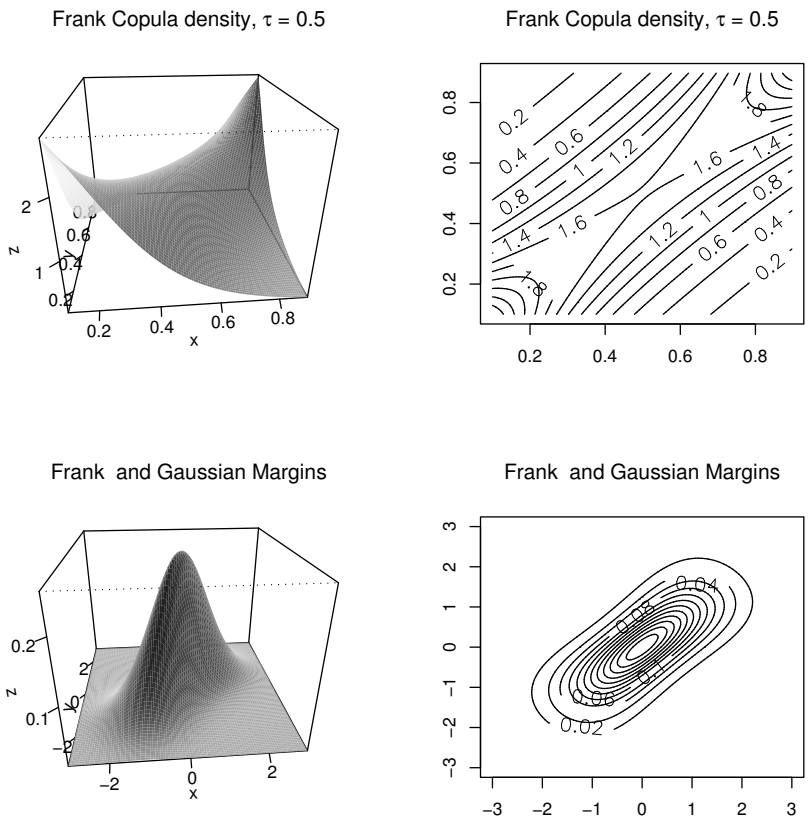


Figure 4.67: Top: The Frank copula with $\tau = 0.5$ and $\theta = 5.73$. Bottom: The same copula using Gaussian margins.

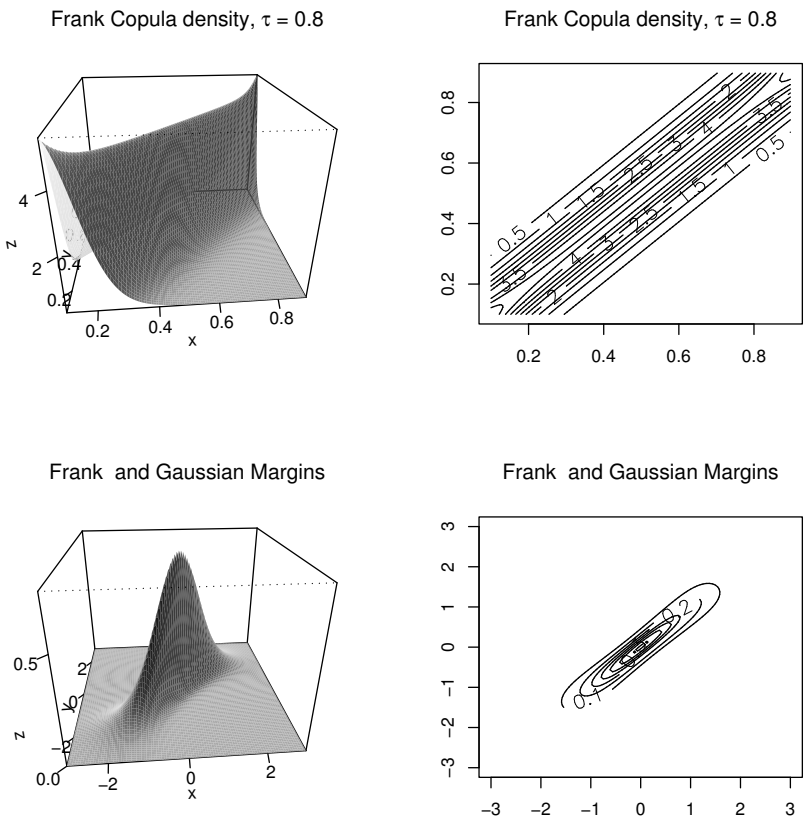


Figure 4.68: Top: The Frank copula with $\tau = 0.8$ and $\theta = 18.19$. Bottom: The same copula using Gaussian margins.

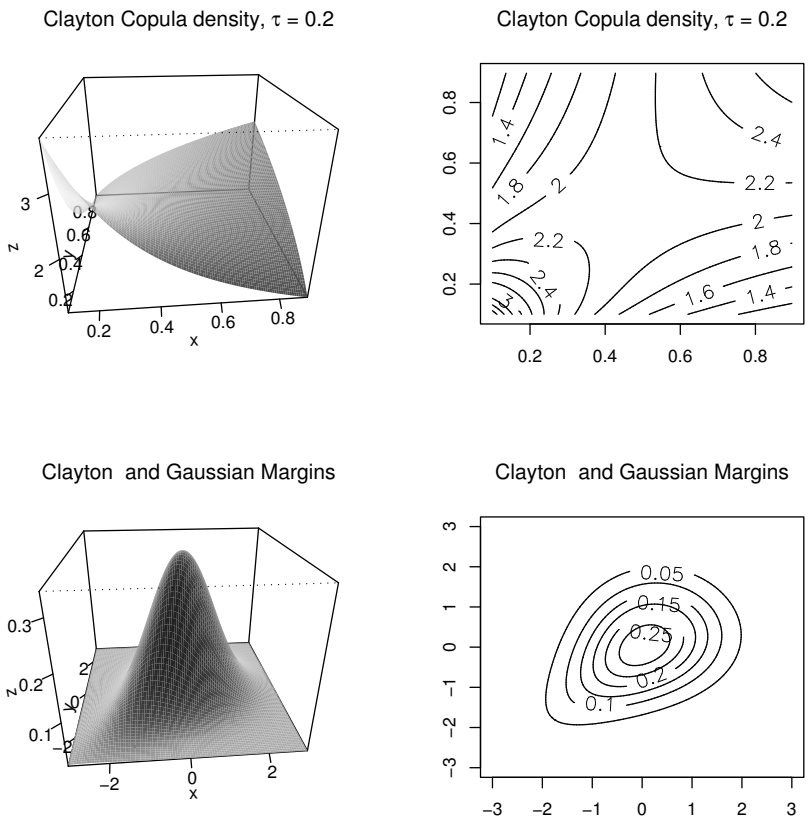


Figure 4.69: Top: The Clayton copula with $\tau = 0.2$ and $\theta = 0.5$. Bottom: The same copula using Gaussian margins.

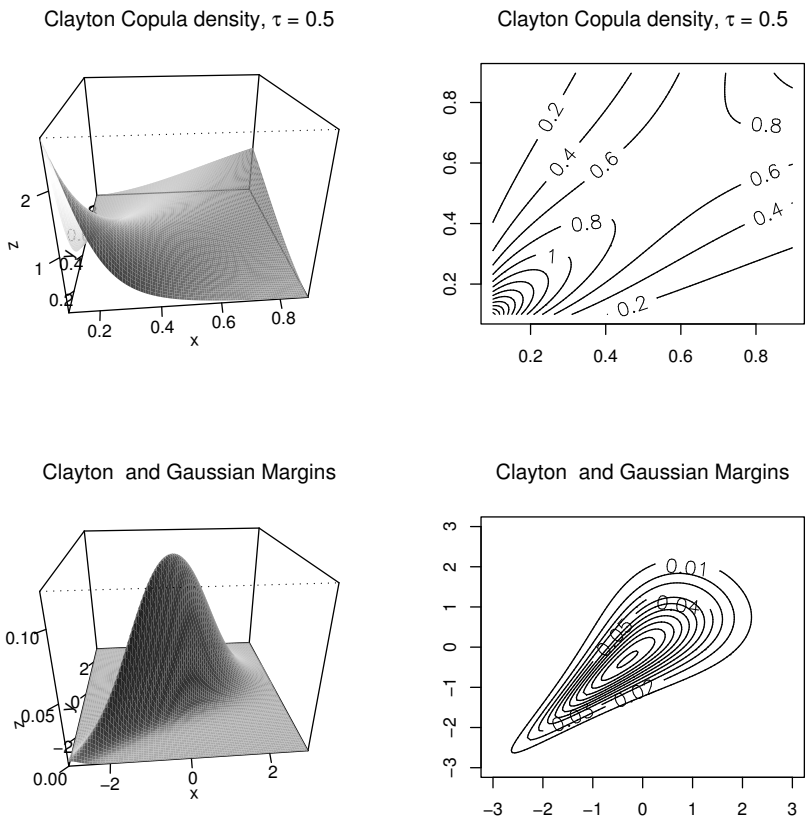


Figure 4.70: Top: The Clayton copula with $\tau = 0.5$ and $\theta = 2$. Bottom: The same copula using Gaussian margins.

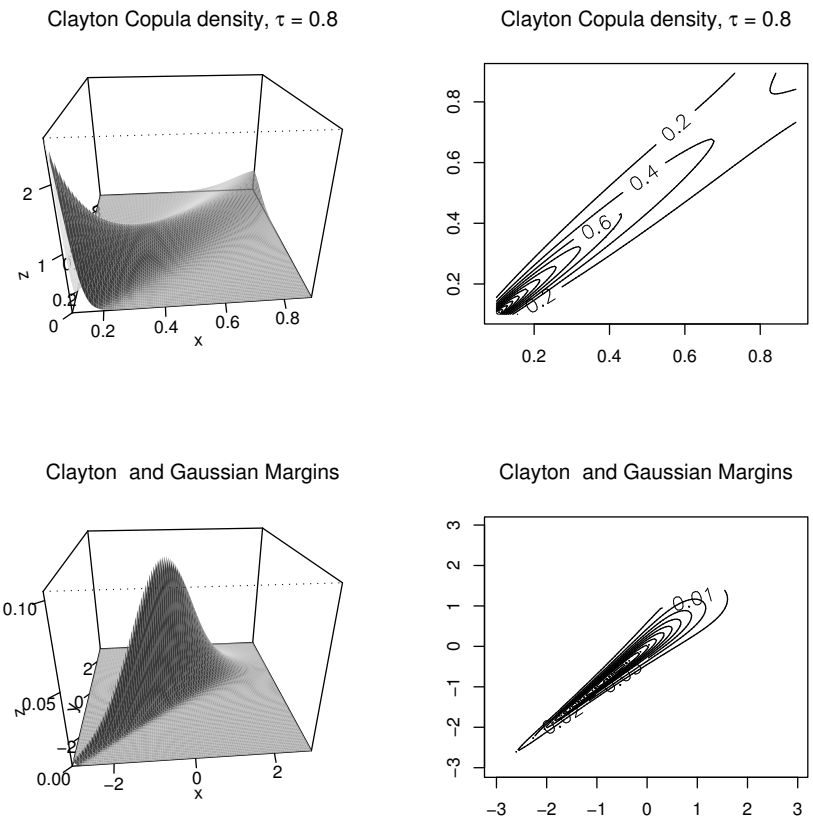


Figure 4.71: Top: The Clayton copula with $\tau = 0.8$ and $\theta = 8$. Bottom: The same copula using Gaussian margins.

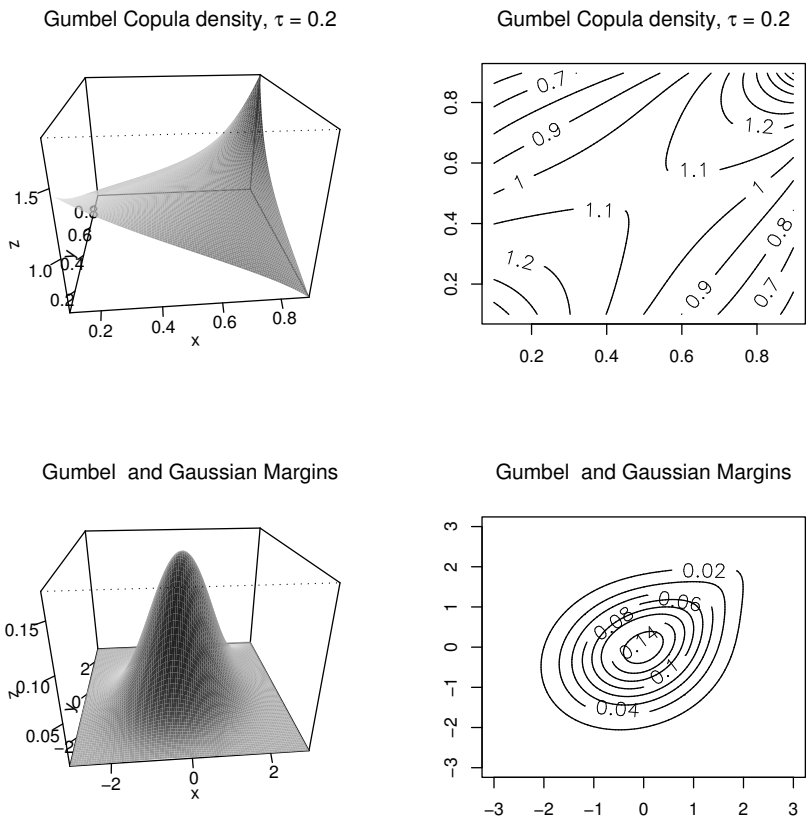


Figure 4.72: Top: The Gumbel-Hougaard copula with $\tau = 0.2$ and $\theta = 1.25$. Bottom: The same copula using Gaussian margins.

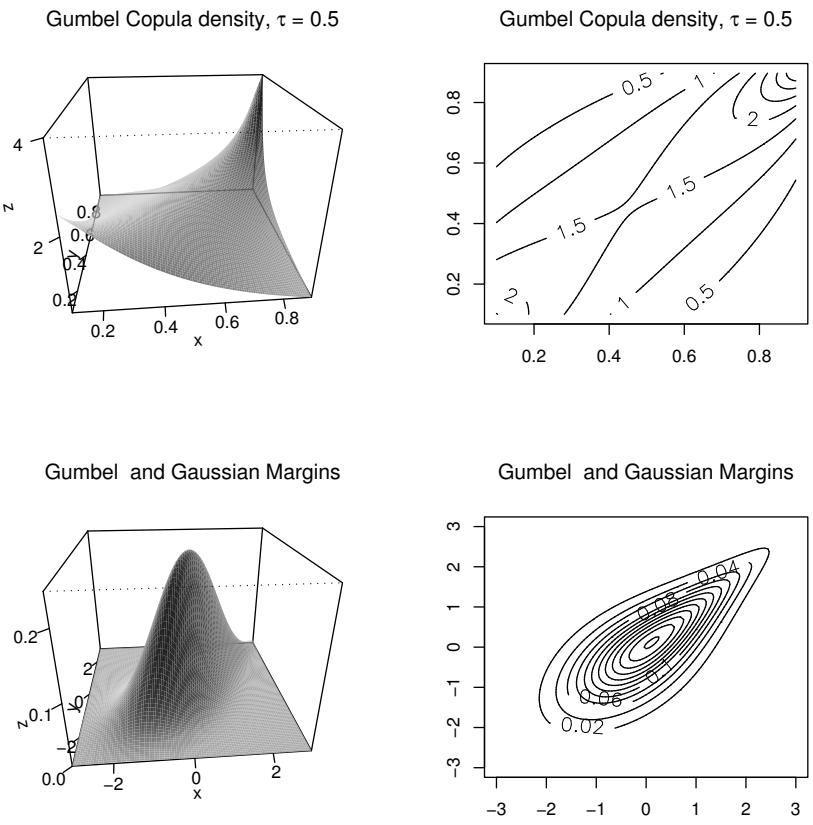


Figure 4.73: Top: The Gumbel-Hougaard copula with $\tau = 0.5$ and $\theta = 2$. Bottom: The same copula using Gaussian margins.

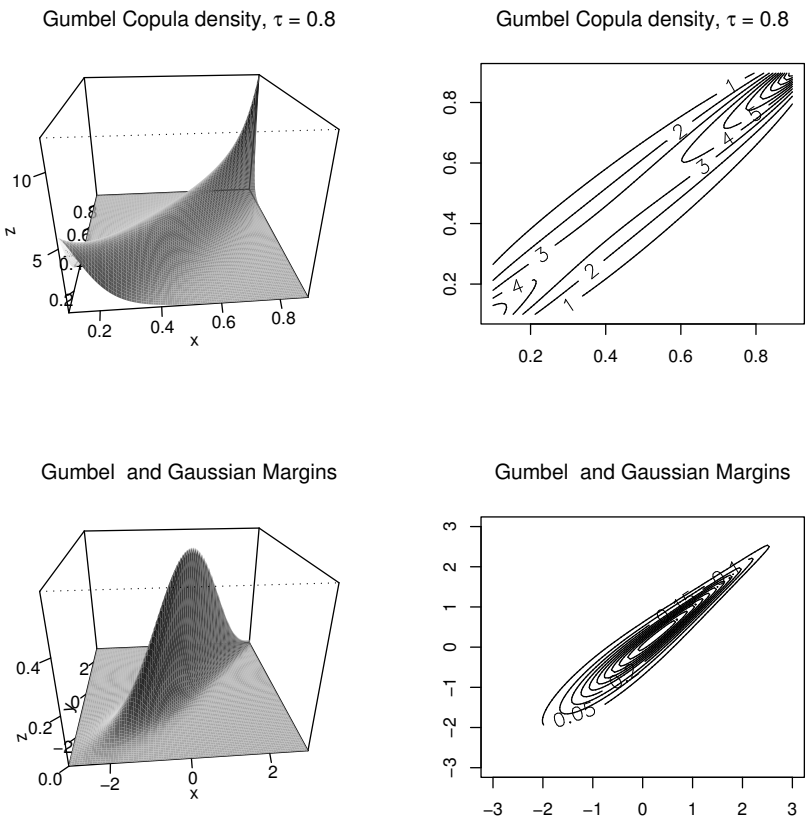


Figure 4.74: Top: The Gumbel-Hougaard copula with $\tau = 0.8$ and $\theta = 5$. Bottom: The same copula using Gaussian margins.

features of the bivariate case are extended for $n > 2$. Let us begin with the extension of the *H-volume* and the *n-increasing* property [Nelsen, 1999] and [Joe, 1997].

Definition 4.5.1. *Let $D_1, \ldots, D_n$ be nonempty subsets of $\mathbb{R}$, and let H be an n-place real function with domain equal to $D_1 \times \ldots \times D_n$. Let $[\mathbf{a}, \mathbf{b}] = \prod_{i=1}^{n}[a_i, b_i]$ be an n-box of whose vertices are all in the domain of H. Then the* H-volume of B *is given by*

$$V_H([\mathbf{a}, \mathbf{b}]) = \sum_{\mathbf{c}} sgn(\mathbf{c}) H(\mathbf{c}) \tag{4.70}$$

where the sum is taken over all vertices $\mathbf{c}$ of B, and $sgn(\mathbf{c})$ is given by

$$sgn(\mathbf{c}) = \begin{cases} 1, & if\ c_k = a_k\ for\ an\ even\ number\ of\ k's, \\ -1, & if\ c_k = a_k\ for\ an\ odd\ number\ of\ k's. \end{cases}$$

Or equivalently

$$V_H([\mathbf{a}, \mathbf{b}]) = \sum_{i_1=1}^{2} \ldots \sum_{i_n=1}^{2} (-1)^{\sum_{j=1}^{n} i_j} H(x_{1i_1}, \ldots, x_{ni_n})$$

where $x_{k1} = a_k$ and $x_{k2} = b_k$.

As we have already mentioned for the bivariate case, the *H-volume* is also the nth order difference of H of $[\mathbf{a}, \mathbf{b}]$:

$$V_H([\mathbf{a}, \mathbf{b}]) = \Delta_{\mathbf{b}}^{\mathbf{a}}(\mathbf{t}) = \Delta_{b_n}^{a_n} \ldots \Delta_{b_1}^{a_1} H(\mathbf{t})$$

where the first order difference is defined by

$$\Delta_{b_k}^{a_k} H(\mathbf{t}) = H(t_1, \ldots, t_{k-1}, b_k, t_{k+1}, \ldots, t_n) - H(t_1, \ldots, t_{k-1}, a_k, t_{k+1}, \ldots, t_n).$$

Example 4.5.1. *Let H be a 3-place function with domain $\mathbb{R}$. The* H-volume *of $B = [a_1, b_1] \times [a_2, b_2] \times [a_3, b_3]$ is given by:*

$$\begin{aligned} V_H(B) &= H(b_1, b_2, b_3) - H(b_1, b_2, a_3) - H(b_1, a_2, b_3) - H(a_1, b_2, b_3) \\ &+ H(b_1, a_2, a_3) + H(a_1, b_2, a_3) + H(a_1, a_2, b_3) - H(a_1, a_2, a_3) \end{aligned}$$

Definition 4.5.2. *An n-place real function H is* n-increasing *if $V_H(B) \geq 0$ for all n-boxes B whose vertices lies in domain of H.*

Definition 4.5.3. *Let H be an n-place function with domain $B_1 \times \ldots \times B_n$, where each B_k has a least element a_k. H is said* grounded *if $H(\mathbf{t}) = 0$ for all $\mathbf{t}$ in the domain of H such $t_k = a_k$ for at least one k. If each B_k is nonempty and has a greatest element b_k, then* one-dimensional margins *of H are the functions H_k given by*

$$H_k(x) = H(b_1, \ldots, b_{k-1}, x, b_{k+1}, \ldots, b_n)\ \forall x \in B_k \tag{4.71}$$

Higher dimensional margins are defined by fixing fewer places in H.

Now, let us give the n-dimensional versions of: the copula definition, the Sklar's theorem and its corollary (inversion method) [Nelsen, 1999].

Definition 4.5.4. *A* n-dimensional copula *is a grounded and n-increasing function* $C : [0,1]^n \mapsto [0,1]$ *with the following additional property:*

$$C(\mathbf{u}) = u_k \tag{4.72}$$

if all coordinates of $\mathbf{u}$ *are* 1 *except* u_k*.*

Theorem 4.5.1 (Sklar's theorem)**.** *Let* H *be an n-dimensional distribution function with margins* $F_1, ..., F_n$*. Then there exists an n-copula* C *such that for all* $x \in \mathbb{R}^n$ *,*

$$H(x_1, ..., x_n) = C(F_1(x_1), ..., F_n(x_n)). \tag{4.73}$$

If $F_1, ..., F_n$ *are all continuous, then* C *is unique; otherwise,* C *is uniquely determined on Range of* $F_1 \times ... \times$ *Range of* F_n*.*

Corollary 4.5.2. *Let* H *is an n-dimensional distribution function with margins* $F_1, ..., F_n$*, and* C *its associated copula. If* $F_1, ..., F_n$ *are continuous and strictly increasing, then for any* $\mathbf{u} = (u_1, \ldots, u_n) \in [0,1]^n$

$$C(u_1, ..., u_n) = H(F_1^{-1}(u_1), \ldots, F_n^{-1}(u_n)) \tag{4.74}$$

Example 4.5.2. *As in the bidimensional case the Normal copula is built by:*

$$C(u_1, ..., u_n; \rho) = \Phi_\rho(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_n)) \tag{4.75}$$

where Φ_ρ *the multidimensional normal law with correlation matrix* ρ*, and* Φ *is the unidimensional Gaussian law.*

Now, as in the bivariate case (definition 4.3.5 and theorem 4.3.6), we can use the classical framework of random variables, and restate the copulas definition and the Sklar's theorem [Nelsen, 1999].

Definition 4.5.5. *A copula is a n-dimensional cumulative distribution function defined on the unit hypercube* $[0,1]^n$ *such the margins distributions are uniform on the interval* $[0,1]$*:*

$$C : [0,1]^n \to [0,1] : (u_1, \ldots, u_n) \mapsto C(u_1, \ldots, u_n).$$

Theorem 4.5.3 (Sklar's theorem)**.** *Let* $X_1, \ldots, X_n$ *n real random variables with distribution functions* $F_1 \ldots, F_n$*, and joint distribution function* H*. Then there exists a n-copula* C *such* (4.73) *holds. If* $F_1 \ldots, F_n$ *are continuous,* C *is unique , otherwise* C *is uniquely determined on range of* $F_1 \times \ldots$ *range of* F_n*.*

The n-dimensional versions of the product copula and the *Fréchet-Hoeffding bounds* are denoted W^n, Π^n and M^n, and are given by:

$$M^n(\mathbf{u}) = min\,(u_1, ..., u_n) \tag{4.76}$$

$$\Pi^n(\mathbf{u}) = u_1 u_2 ... u_n \tag{4.77}$$

$$W^n(\mathbf{u}) = max\,(u_1 + ... + u_n - n + 1, 0) \tag{4.78}$$

The functions M^n and Π are always n-copulas, whereas W^n is a copula only if $n = 2$. But the *Fréchet-Hoeffding bounds* hold as the following theorems stated it [Nelsen, 1999].

Theorem 4.5.4. *If C is a n-copula, then for every $\mathbf{u} \in [0,1]^n$*

$$W^n(\mathbf{u}) \leq C(\mathbf{u}) \leq M^n(\mathbf{u}). \tag{4.79}$$

Theorem 4.5.5. *For any $n \geq 3$ and any $\mathbf{u} \in [0,1]^n$, there exists an n-copula C (which depend on $\mathbf{u}$) such*

$$C(\mathbf{u}) = W^n(\mathbf{u}). \tag{4.80}$$

Although W^n is not a copula, it is the "best-possible" left bound for n-copulas.These n-copulas Π^n and M^n are also well-known dependence structures [Nelsen, 1999].

Theorem 4.5.6. *Let $n \geq 2$ and $X_1, ..., X_n$ be continuous random variables, then:*

- *$X_1, ..., X_n$ are independents if and only if the n-copula of $X_1, ..., X_n$ is Π^n.*
- *All variables of $X_1, ..., X_n$ are almost surely a strictly increasing function of any of the others if and only if the n-copula of $X_1, ..., X_n$ is M^n.*

Now, let $H(x_1, \ldots, x_n)$ a joint distribution, $F_1, \ldots, F_n$ its margins, and C the associated copula. The joint density of H, noted h, is given by

$$h(x_1, \ldots, x_n) = \frac{\partial^n H(x_1, \ldots, x_n)}{\partial x_1 \ldots \partial x_n} = \partial^n_{1\ldots n} H(x_1, \ldots, x_n) \tag{4.81}$$

and of course the joint density of C, noted c, by

$$c(u_1, \ldots, u_n) = \frac{\partial^n C(u_1, \ldots, u_n)}{\partial u^1 \ldots \partial u^n} = \partial^n_{1\ldots n} C(u_1, \ldots, u_n). \tag{4.82}$$

Again, let us remark that the condition $V_C(B) \geq 0$ for any $B = [a_1, b_1] \times \ldots \times [a_n, b_n] \subset [0,1]^n$, is equivalent to positivity of the density when this later exists. Using the Sklar's theorem we have the following relation:

$$h(x_1, \ldots, x_n) = c\left(F_1(x_1), \ldots, F_n(x_n)\right) \prod_{i=1}^{n} f_i(x_i) \tag{4.83}$$

where $f_1, \ldots, f_n$ are the densities of respectively $F_1, \ldots, F_n$. And, if $F_1, \ldots, F_n$ are continuous, we can deduce of this the joint density c:

$$c(u_1, \ldots, u_n) = \frac{h\left(F_1^{-1}(u_1), \ldots, F_n^{-1}(u_n)\right)}{\prod_{i=1}^{n} f_i\left(F_i^{-1}(u_i)\right)}. \tag{4.84}$$

Example 4.5.3. *Recall the multivariate normal density of dimension n:*

$$\phi_\Sigma(x_1, \ldots, x_n) = \frac{1}{(2\pi)^{n/2}\left|\Sigma\right|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu)^\top \rho^{-1}(x-\mu)\right) \tag{4.85}$$

where $\mu = (\mu_1, \ldots, \mu_n)$ is a vector of means, and Σ a covariance matrix. Suppose that $\mu_i = 0$ and $\sigma_i = 1$ for all i, where σ_i is the standard deviation of the i^{th} component. Then the joint density is given by

$$\phi_\rho(x_1, \ldots, x_n) = \frac{1}{(2\pi)^{n/2}\left|\rho\right|^{1/2}} \exp\left(-\frac{1}{2}x^\top \rho^{-1} x\right) \tag{4.86}$$

where ρ is correlation matrix. Then the density margins are the gaussian densities:

$$\phi(x_i) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x_i^2}{2}\right).$$

Using (4.84) *we can write:*

$$\begin{aligned} c(u_1, \ldots, u_n) &= \frac{\frac{1}{(2\pi)^{n/2}|\rho|^{1/2}} \exp\left(-\frac{1}{2}\Phi^{-1}(u)^\top \rho^{-1}\Phi^{-1}(u)\right)}{\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\Phi^{-1}(x_i)^2}{2}\right)} \\ &= \frac{1}{\left|\rho\right|^{1/2}} \exp\left(-\frac{1}{2}\Phi^{-1}(u)^\top \left(\rho^{-1} - \mathbf{I}\right)\Phi^{-1}(u)\right) \end{aligned} \tag{4.87}$$

where $\Phi^{-1}(u) = (\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_n))$ and $\mathbf{I}$ is the $n \times n$ identity matrix.

4.6 Multivariate Archimedean Copulas

The extension of the formula (4.55) for a n-dimensional vector $\mathbf{u} = (u_1, \ldots, u_n)$ is very "natural":

$$C^n(\mathbf{u}) = \phi^{[-1]}\left(\sum_{i=1}^{n} \phi(u_i)\right) \tag{4.88}$$

$$= \psi\left(\sum_{i=1}^{n} \phi(u_i)\right) \tag{4.89}$$

where $\psi = \phi^{[-1]}$, but what is the condition on the generator ϕ that the above expression (4.89) is distribution? The answer to the reverse question is more easy to find. Suppose that for a differentiable real function function f we use the following notation:

$$f^{(k)}(x) = \frac{d^k}{dx^k} f(x) \tag{4.90}$$

then if we suppose that the expression (4.89) is a copula, then its density is given by:

$$c(\mathbf{u}) = c(u_1, \ldots, u_n) = \frac{\partial^n C(\mathbf{u})}{\partial u_1 \ldots \partial u_n} = \psi^{(n)} \left[\sum_{i=1}^{n} \phi(u_i) \right] \cdot \prod_{i=1}^{n} \phi'(u_i) \tag{4.91}$$

with $\phi'(t) = \phi^{(1)}(t)$.

The bivariate margin of (4.89) for two chosen dimensions $k \neq l$, denoted C_{kl} is given by:

$$\begin{aligned} C_{kl}(u_k, u_l) &= C(1, \ldots, u_k, \ldots, u_l, \ldots, 1) \\ &= \psi \left(\phi(u_k) + \phi(u_l) + \sum_{i \neq k,l} \phi(1) \right) \\ &= \psi \left(\phi(u_k) + \phi(u_l) \right) \end{aligned} \tag{4.92}$$

and is a bivariate copula using the same generator ϕ. Thus this latter must be fulfill the conditions of the theorem 4.4.3: ϕ must be a decreasing function, i.e. $\phi^{(1)}(u_i) < 0$, $\forall i \in 1, \ldots, n$. Then the sign of the product $\prod_{i=1}^{n} \phi^{(1)}(u_i)$ in expression (4.91) is given by $(-1)^n$, and, as a density is always positive, we must have:

$$(-1)^n \psi^{(n)} \left[\sum_{i=1}^{n} \phi(u_i) \right] > 0$$

which is the property of *completely monotonic* function.

Definition 4.6.1. *[Widder, 1941] A continuous function $f(t)$ defined on an interval $[a, b]$ is* completely monotonic *if and only if*

$$(-1)^k \frac{d^k}{dt^k} f(t) \geq 0 \tag{4.93}$$

$\forall a < x < b$.

We can also find in [Widder, 1941] that if $f(t)$ is completely monotonic in $]a, b[$ and is zero at $x = c > a$, then it is identically zero. So if the pseudo inverse $\phi^{[-1]}$ of an archimedean generator ϕ is completely monotonic, then it must be positive on $[0, \infty[$, i.e. ϕ is strict and $\phi^{[-1]} = \phi^{-1}$.

The proof of necessary and sufficient conditions for function ϕ to be an archimedean generator for any $n \geq 2$ can be found in [Kimberling, 1974] and in [Schweizer and Sklar, 1983].

Theorem 4.6.1. *Let ϕ be a continuous strictly decreasing function from $[0,1]$ to $[0,\infty[$ such $\lim_{t\to 0}\phi(t)=\infty$ and $\phi(1)=0$. If C^n is the function from $[0,1]^n \to [0,1]$ given by (4.89), then C^n is an n-copula for all $n \geq 2$ if and only if $\psi = \phi^{-1}$ is completely monotonic on $[0,\infty[$.*

The following theorem, due to Bernstein [Bernstein, 1928], close the loop, since we have introduced archimedean copulas using expression(4.48). Several proofs of this theorem can be found in [Widder, 1941].

Theorem 4.6.2 (Bernstein Theorem)**.** *A necessary and sufficient condition that $f(x)$ should be completely monotonic in $0 \leq x < \infty$ is that*

$$f(x) = \int_0^\infty e^{-xt} d\alpha(t) \tag{4.94}$$

where $\alpha(t)$ is bounded and non decreasing and the integral converges for $0 \leq x < \infty$.

We give below the n-dimensional version of the four copulas shown in examples 4.4.1, 4.4.2, 4.4.3 and 4.4.4. We can see that the domain of the parameter for the generators are subset of the domains for bivariate copulas.

Example 4.6.1. *The Clayton family, with generator given by expression (4.58), for which ϕ^{-1} is completely monotonic when $\theta > 0$, is given by:*

$$C_\theta^n(\mathbf{u}) = \left(1 - n + \sum_{i=1}^n u_i^{-\theta}\right)^{-1/\theta}. \tag{4.95}$$

Example 4.6.2. *The Frank family, with generator given by expression (4.60) for which ϕ^{-1} is completely monotonic when $\theta > 0$, is given by*

$$C_\theta^n(\mathbf{u}) = -\frac{1}{\theta}\left(1 + \frac{\prod_{i=1}^n (e^{-\theta u_i} - 1)}{(e^{-\theta} - 1)^{n-1}}\right). \tag{4.96}$$

Example 4.6.3. *The Gumbel-Hougaard family, with generator given by expression (4.62) for which ϕ^{-1} is completely monotonic when $\theta \geq 1$, is given by:*

$$C_\theta^n(\mathbf{u}) = \exp\left(-\left[\sum_{i=1}^n (-\ln u_i)^\theta\right]^{1/\theta}\right). \tag{4.97}$$

Example 4.6.4. *The Joe family, with generator given by expression (4.64) for which ϕ^{-1} is completely monotonic when $\theta > 1$, is given by:*

$$C_\theta^n(\mathbf{u}) = 1 - \left[1 - \prod_{i=1}^n \left(1 - (1-u_i)^\theta\right)\right]^{1/\theta}. \tag{4.98}$$

Remark 4.6.1. *By the theorem 4.6.1 we know that expression* (4.89) *is a copula when* $\psi = \phi^{-1}$ *completely monotonic. The "price" of this property is given in the theorem below [Nelsen, 1999]:*

> ***Archimedean n-copulas can only capture dependence structures from independence*** Π ***to completely positive dependence*** M***.***

Theorem 4.6.3. *Let* ϕ *be a strict archimedean generator. If* ϕ^{-1} *is completely monotonic, then copula* C *built with expression* (4.89) *is such*

$$C(\mathbf{u}) \geq \Pi(\mathbf{u})$$

for all $\mathbf{u}$ *in* $[0,1]^n$*.*

Another restriction to the use of Archimedean n-copulas is the fact that it is not really easy to find a general formula for for the density given in expression (4.91) for any n. The difficult part is to find a general formula for $\psi^{(n)}$. For the Clayton family it is very easy, but for the Frank, Gumbel-Hougaard and Joe families it is harder. It is our main contribution in this field to have found recurrence formulas for these three later families.

Proposition 4.6.4. *If* ϕ *is the generator of the Clayton copula, then we have the following formulas:*

$$\begin{aligned}
\phi(t) &= (t^{-\theta} - 1) && (4.99)\\
\phi'(t) &= -\theta u^{-\theta-1} && (4.100)\\
\psi(t) &= (t+1)^{-1/\theta} && (4.101)\\
\psi^{(n)}(t) &= (-1)^n (t+1)^{-\frac{1}{\theta}-n} \prod_{i=1}^{n} \left(\frac{1}{\theta} + i - 1\right) && (4.102)
\end{aligned}$$

Proof. : By recurrence:

If $n = 1$:

$$\psi^{(1)}(t) = -\frac{1}{\theta}(t+1)^{-\frac{1}{\theta}-1}$$

If the formula (4.102) hold for $n-1$, then:

$$\begin{aligned}
\psi^{(n)}(t) &= \left\{(-1)^{n-1}(t+1)^{-\frac{1}{\theta}-n+1} \prod_{i=1}^{n-1} \left(\frac{1}{\theta} + i - 1\right)\right\}'\\
&= \left\{(-1)^{n-1}(t+1)^{-\frac{1}{\theta}-n} \left(-\frac{1}{\theta} - n + 1\right) \prod_{i=1}^{n-1} \left(\frac{1}{\theta} + i - 1\right)\right\}'\\
&= (-1)^n (t+1)^{-\frac{1}{\theta}-n} \prod_{i=1}^{n} \left(\frac{1}{\theta} + i - 1\right)
\end{aligned}$$

□

Proposition 4.6.5. *If ϕ is the generator of the Frank copula, then we have the following formulas:*

$$\begin{aligned}
\phi(t) &= -\ln\left(\frac{e^{-\theta t}-1}{e^{-\theta}-1}\right) && (4.103)\\
\phi'(t) &= \frac{\theta e^{-\theta t}}{e^{-\theta t}-1} && (4.104)\\
\psi(t) &= -\frac{1}{\theta}\ln\left[1-\mathcal{E}e^{-t}\right] && (4.105)\\
\psi^{(n)}(t) &= \frac{(-1)^n}{\theta}\left[\sum_{i=1}^{n-1}\mathcal{K}_n^i\left(\mathcal{E}e^{-t}\right)^i\right][1-\mathcal{E}e^{-u}]^{-n} && (4.106)
\end{aligned}$$

where:

- $\mathcal{E}=1-e^{-\theta}$,
- $\mathcal{K}_1^1=1$,
- $\mathcal{K}_n^0=\mathcal{K}_n^n=0$,
- $\mathcal{K}_n^i=(n-i)\mathcal{K}_{n-1}^{i-1}+i\mathcal{K}_{n-1}^i$.

Proof. :

By recurrence:

If $n=1$:

$$\psi^{(1)}(t)=-\frac{1}{\theta}[\mathcal{E}e^{-t}][1-\mathcal{E}e^{-t}]^{-1}$$

If the formula (4.106) hold for $n-1$, then:

$$\begin{aligned}
\psi^{(n)}(t) &= \left\{ \frac{(-1)^{n-1}}{\theta} \left[\sum_{i=1}^{n-2} \mathcal{K}_{n-1}^{i} \mathcal{E}^{i} e^{-it} \right] [1-\mathcal{E}e^{-t}]^{-(n-1)} \right\}' \\
&= \frac{(-1)^{n-1}}{\theta} \left\{ \left[\sum_{i=1}^{n-2} (-i) \mathcal{K}_{n-1}^{i} \mathcal{E}^{i} e^{-it} \right] [1-\mathcal{E}e^{-t}]^{-n+1} \right. \\
&\quad \left. + \left[\sum_{i=1}^{n-2} \mathcal{K}_{n-1}^{i} \mathcal{E}^{i+1} e^{-(i+1)t} \right] (-n+1)[1-\mathcal{E}e^{-t}]^{-n} \right\} \\
&= \frac{(-1)^{n}}{\theta} [1-\mathcal{E}e^{-t}]^{-n} \left\{ \left[\sum_{i=1}^{n-2} i \mathcal{K}_{n-1}^{i} \mathcal{E}^{i} e^{-it} \right] [1-\mathcal{E}e^{-t}] \right. \\
&\quad \left. + \left[\sum_{i=1}^{n-2} \mathcal{K}_{n-1}^{i} \mathcal{E}^{i+1} e^{-(i+1)t} \right] (n-1) \right\} \\
&= \frac{(-1)^{n}}{\theta} [1-\mathcal{E}e^{-t}]^{-n} \left\{ \sum_{i=1}^{n-2} i \mathcal{K}_{n-1}^{i} \mathcal{E}^{i} e^{-it} \right. \\
&\quad \left. + \sum_{i=1}^{n-2} i \mathcal{K}_{n-1}^{i} \mathcal{E}^{i+1} e^{-(i+1)t} + \sum_{i=1}^{n-2} (n-1) \mathcal{K}_{n-1}^{i} \mathcal{E}^{i+1} e^{-(i+1)t} \right\} \\
&= \frac{(-1)^{n}}{\theta} [1-\mathcal{E}e^{-t}]^{-n} \left\{ \sum_{i=1}^{n-2} i \mathcal{K}_{n-1}^{i} \mathcal{E}^{i} e^{-it} \right. \\
&\quad \left. + \sum_{j=2}^{n-1} (j-1) \mathcal{K}_{n-1}^{j-1} \mathcal{E}^{j} e^{-jt} + \sum_{j=2}^{n-1} (n-1) \mathcal{K}_{n-1}^{j-1} \mathcal{E}^{j} e^{-j)t} \right\} \\
&= \frac{(-1)^{n}}{\theta} [1-\mathcal{E}e^{-t}]^{-n} \\
&\quad \cdot \left\{ \mathcal{K}_{n-1}^{1} \mathcal{E}e^{t} + \mathcal{K}_{n-1}^{n-2} \mathcal{E}^{n-1} e^{-(n-1)t} + \sum_{i=1}^{n-2} \mathcal{E}^{i} e^{-it} \left[(n-i) \mathcal{K}_{n-1}^{i-1} + i \mathcal{K}_{n-1}^{i} \right] \right\}
\end{aligned}$$

If we set

$$\mathcal{K}_{n}^{i} = (n-i) \mathcal{K}_{n-1}^{i-1} + i \mathcal{K}_{n-1}^{i}$$

and $\mathcal{K}_{1}^{1} = 1$, $\mathcal{K}_{n}^{0} = \mathcal{K}_{n}^{n} = 0$, then

$$\mathcal{K}_{n}^{1} = (n-1) \mathcal{K}_{n-1}^{0} + \mathcal{K}_{n-1}^{1} = \mathcal{K}_{n-1}^{1}$$

and

$$\mathcal{K}_{n}^{n-1} = (n-n+1) \mathcal{K}_{n-1}^{n-2} + \mathcal{K}_{n-1}^{n-1} = \mathcal{K}_{n-1}^{n-2}$$

and the formula (4.106) hold for n. □

	i									
n	0	1	2	3	4	5	6	7	8	9
1	0	1								
2	0	1	0							
3	0	1	1	0						
4	0	1	4	1	0					
5	0	1	11	11	1	0				
6	0	1	26	66	26	1	0			
7	0	1	57	302	302	57	1	0		
8	0	1	120	1191	2416	1191	120	1	0	
9	0	1	247	4293	15619	15619	4293	247	1	0

Table 4.2: Coefficients $\mathcal{K}_n^i$ for n^{th} derivatives of the inverse of Frank generator for $n \in \{1, \ldots, 9\}$.

The coefficients $\mathcal{K}_n^i$ in the preceding theorem can be found by recurrence as the binomial coefficients. The table 4.2 shows the values of these coefficients for $n = 1$ to $n = 9$, and we can remark the symmetry of the table, as in the binomial case.

Proposition 4.6.6. *If ϕ is the generator of the Gumbel-Hougaard copula, then we have the following formulas:*

$$\phi(t) = (-\ln t)^\theta \tag{4.107}$$

$$\phi'(t) = -\frac{\theta}{t}(-\ln t)^{\theta-1} \tag{4.108}$$

$$\psi(t) = e^{-t^{1/\theta}} \tag{4.109}$$

$$\psi^{(n)}(t) = e^{-t^{1/\theta}} \left[\sum_{i=1}^{n} \mathcal{K}_n^i t^{\frac{1}{\theta}-n}\right] \tag{4.110}$$

where:

- $\mathcal{K}_0^0 = 1$,
- $\mathcal{K}_n^0 = \mathcal{K}_n^{n+1} = 0$,
- $\mathcal{K}_n^i = -\frac{1}{\theta}\mathcal{K}_{n-1}^{i-1} + \left(\frac{i}{\theta} - n + 1\right)\mathcal{K}_{n-1}^i$.

Proof. :

By recurrence:

If $n = 1$:

$$\psi^{(1)}(t) = -\frac{1}{\theta} t^{\frac{1}{\theta}-1} e^{-t^{1/\theta}}$$

If the formula (4.110) hold for $n-1$, then:

$$\begin{aligned}
\psi^{(n)}(t) &= \left\{ e^{-t^{1/\theta}} \sum_{i=1}^{n-1} \mathcal{K}_{n-1}^{i} t^{\frac{i}{\theta}-n+1} \right\}' \\
&= e^{-t^{1/\theta}} \left\{ -\frac{t^{\frac{1}{\theta}-1}}{\theta} \sum_{i=1}^{n-1} \mathcal{K}_{n-1}^{i} t^{\frac{i}{\theta}-n+1} + \sum_{i=1}^{n-1} \mathcal{K}_{n-1}^{i} t^{\frac{i}{\theta}-n} \left(\frac{i}{\theta} - n + 1 \right) \right\} \\
&= e^{-t^{1/\theta}} \left\{ \sum_{i=1}^{n-1} \mathcal{K}_{n-1}^{i} t^{\frac{i+1}{\theta}-n} \left(-\frac{1}{\theta} \right) + \sum_{i=1}^{n-1} \mathcal{K}_{n-1}^{i} t^{\frac{i}{\theta}-n} \left(\frac{i}{\theta} - n + 1 \right) \right\} \\
&= e^{-t^{1/\theta}} \left\{ \sum_{j=2}^{n} \mathcal{K}_{n-1}^{j-1} t^{\frac{j}{\theta}-n} \left(-\frac{1}{\theta} \right) + \sum_{i=1}^{n-1} \mathcal{K}_{n-1}^{i} t^{\frac{i}{\theta}-n} \left(\frac{i}{\theta} - n + 1 \right) \right\} \\
&= e^{-t^{1/\theta}} \left\{ \sum_{j=2}^{n-1} t^{\frac{j}{\theta}-n} \left[\mathcal{K}_{n-1}^{i-1} \left(-\frac{1}{\theta} \right) + \mathcal{K}_{n-1}^{i} \left(\frac{i}{\theta} - n + 1 \right) \right] \right. \\
&\quad \left. + \mathcal{K}_{n-1}^{1} \left(\frac{1}{\theta} - n + 1 \right) t^{\frac{1}{\theta}-n} + \mathcal{K}_{n-1}^{n-1} \left(-\frac{1}{\theta} \right) t^{\frac{n}{\theta}-n} \right\}
\end{aligned}$$

If we set

$$\mathcal{K}_{n}^{i} = -\frac{1}{\theta} \mathcal{K}_{n-1}^{i-1} + \left(\frac{i}{\theta} - n + 1 \right) \mathcal{K}_{n-1}^{i}$$

and $\mathcal{K}_{0}^{0} = 1$, $\mathcal{K}_{n}^{0} = \mathcal{K}_{n}^{n+1} = 0$, then

$$\mathcal{K}_{n}^{1} = -\frac{1}{\theta} \mathcal{K}_{n-1}^{0} + \left(\frac{1}{\theta} - n + 1 \right) \mathcal{K}_{n-1}^{1} = \left(\frac{1}{\theta} - n + 1 \right) \mathcal{K}_{n-1}^{1}$$

and

$$\mathcal{K}_{n}^{n} = -\frac{1}{\theta} \mathcal{K}_{n-1}^{n-1} + \left(\frac{n}{\theta} - n + 1 \right) \mathcal{K}_{n-1}^{n} = -\frac{1}{\theta} \mathcal{K}_{n-1}^{n-1}$$

and finally

$$\mathcal{K}_{1}^{1} = -\frac{1}{\theta} \mathcal{K}_{0}^{0} + \left(\frac{1}{\theta} - 1 + 1 \right) \mathcal{K}_{0}^{1} = -\frac{1}{\theta}$$

and the formula (4.110) hold for n. □

Tables 4.3, 4.4 and 4.5 show the $\mathcal{K}_{n}^{i}$ coefficients for the Gumbel-Hougaard copula, with θ equal respectively to 0.5, 1 and 2.

Proposition 4.6.7. *If ϕ is the generator of the Joe copula, then we have*

	i									
n	0	1	2	3	4	5	6	7	8	9
0	1	0								
1	0	-2	0							
2	0	-2	4	0						
3	0	0	12	-8	0					
4	0	0	12	-48	16	0				
5	0	0	0	-120	160	-32	0			
6	0	0	0	-120	720	-480	64	0		
7	0	0	0	0	1680	-3360	1344	-128	0	
8	0	0	0	0	1680	-13440	13440	-3584	256	0
9	0	0	0	0	0	-30240	80640	-48384	9216	-512

Table 4.3: Coefficients $\mathcal{K}_n^i$ for n^{th} derivatives of the inverse of Gumbel-Hougaard generator for $n \in \{1, \ldots, 9\}$ and $\theta = 0.5$.

	i									
n	0	1	2	3	4	5	6	7	8	9
0	1	0								
1	0	-1	0							
2	0	0	1	0						
3	0	0	0	-1	0					
4	0	0	0	0	1	0				
5	0	0	0	0	0	-1	0			
6	0	0	0	0	0	0	1	0		
7	0	0	0	0	0	0	0	-1	0	
8	0	0	0	0	0	0	0	0	1	0
9	0	0	0	0	0	0	0	0	0	-1

Table 4.4: Coefficients $\mathcal{K}_n^i$ for n^{th} derivatives of the inverse of Gumbel-Hougaard generator for $n \in \{1, \ldots, 9\}$ and $\theta = 1$.

	i								
n	0	1	2	3	4	5	6	7	8
0	1	0							
1	0	-0,5	0						
2	0	0,25	0,25	0					
3	0	-0,37	-0,37	-0,125	0				
4	0	0,93	0,93	0,37	0,06	0			
5	0	-3,28	-3,28	-1,40	-0,31	-0,03	0		
6	0	14,76	14,76	6,56	1,64	0,23	0,01	0	
7	0	-81,21	-81,21	-36,91	-9,84	-1,64	-0,16	-0,01	0
8	0	527,87	527,87	243,63	67,67	12,30	1,47	0,11	0,003

Table 4.5: Coefficients $\mathcal{K}_n^i$ for n^{th} derivatives of the inverse of Gumbel-Hougaard generator for $n \in \{1, \ldots, 8\}$ and $\theta = 2$.

the following formulas:

$$\phi(t) = -\ln\left[1-(1-t)^{\theta}\right] \tag{4.111}$$

$$\phi'(t) = -\frac{\theta(1-u)^{\theta-1}}{1-(1-u)^{\theta}} \tag{4.112}$$

$$\psi(t) = 1-\left(1-e^{-u}\right)^{\frac{1}{\theta}} \tag{4.113}$$

$$\psi^{(n)}(t) = \left(1-e^{-t}\right)^{\frac{1}{\theta}-n}\left[\sum_{i=1}^{n}\mathcal{K}_n^i e^{-iu}\right] \tag{4.114}$$

where:

- $\mathcal{K}_1^1 = -\frac{1}{\theta}$,
- $\mathcal{K}_n^0 = \mathcal{K}_n^{n+1} = 0$,
- $\mathcal{K}_n^i = \mathcal{K}_{n-1}^{i-1}\left(\frac{1}{\theta}-n+i\right) - i\mathcal{K}_{n-1}^i$.

Proof. : By recurrence:

If $n = 1$:

$$\psi^{(1)}(t) = e^{-t}\left(1-e^{t}\right)^{\frac{1}{\theta}-1}\left(-\frac{1}{\theta}\right)$$

If the formula (4.114) hold for $n-1$, then:

$$\begin{aligned}
\psi^{(n)}(t) &= \left\{ \left(1-e^{-t}\right)^{\frac{1}{\theta}-n+1} \left[\sum_{i=1}^{n-1} \mathcal{K}_{n-1}^{i} e^{-it}\right] \right\}' \\
&= \left(1-e^{-t}\right)^{\frac{1}{\theta}-n} \left(\frac{1}{\theta}-n+1\right) \left[\sum_{i=1}^{n-1} \mathcal{K}_{n-1}^{i} e^{-(i+1)t}\right] \\
&\quad + \left(1-e^{-t}\right)^{\frac{1}{\theta}-n+1} \left[\sum_{i=1}^{n-1} \mathcal{K}_{n-1}^{i}(-i) e^{-it}\right] \\
&= \left(1-e^{-t}\right)^{\frac{1}{\theta}-n} \left\{ \left(\frac{1}{\theta}-n+1\right) \left[\sum_{i=1}^{n-1} \mathcal{K}_{n-1}^{i} e^{-(i+1)t}\right] \right. \\
&\quad \left. + \left(1-e^{-t}\right) \left[\sum_{i=1}^{n-1} \mathcal{K}_{n-1}^{i}(-i) e^{-it}\right] \right\} \\
&= \left(1-e^{-t}\right)^{\frac{1}{\theta}-n} \left\{ \left[\sum_{i=1}^{n-1} \left(\frac{1}{\theta}-n+1\right) \mathcal{K}_{n-1}^{i} e^{-(i+1)t}\right] \right. \\
&\quad \left. + \left[\sum_{i=1}^{n-1} \mathcal{K}_{n-1}^{i}(-i) e^{-it}\right] + \left[\sum_{i=1}^{n-1} \mathcal{K}_{n-1}^{i} i e^{-(i+1)t}\right] \right\} \\
&= \left(1-e^{-t}\right)^{\frac{1}{\theta}-n} \left\{ \left[\sum_{j=2}^{n} \left(\frac{1}{\theta}-n+1\right) \mathcal{K}_{n-1}^{j-1} e^{-jt}\right] \right. \\
&\quad \left. + \left[\sum_{i=1}^{n-1} \mathcal{K}_{n-1}^{i}(-i) e^{-it}\right] + \left[\sum_{j=2}^{n} \mathcal{K}_{n-1}^{j-1}(j-1) e^{-jt}\right] \right\} \\
&= \left(1-e^{-t}\right)^{\frac{1}{\theta}-n} \left\{ \left[\sum_{i=2}^{n-1} e^{-it} \left[\left(\frac{1}{\theta}-n+i\right) \mathcal{K}_{n-1}^{i-1} - i\mathcal{K}_{n-1}^{i}\right]\right] \right. \\
&\quad \left. + (-1)\mathcal{K}_{n-1}^{1} e^{-u} + \frac{1}{\theta}\mathcal{K}_{n-1}^{n-1} e^{-nt} \right\}
\end{aligned}$$

If we set

$$\mathcal{K}_{n}^{i} = \left(\frac{1}{\theta}-n+i\right) \mathcal{K}_{n-1}^{i-1} - i\mathcal{K}_{n-1}^{i}$$

and $\mathcal{K}_{n-1}^{0} = \mathcal{K}_{n-1}^{n} = 0$, then

$$\mathcal{K}_{n}^{1} = \left(\frac{1}{\theta}-n+1\right) \mathcal{K}_{n-1}^{0} - \mathcal{K}_{n-1}^{1} = -\mathcal{K}_{n-1}^{1}$$

and

$$\mathcal{K}_{n}^{n} = \left(\frac{1}{\theta}-n+n\right) \mathcal{K}_{n-1}^{n-1} - n\mathcal{K}_{n-1}^{n} = \frac{1}{\theta}\mathcal{K}_{n-1}^{n-1}$$

	i									
n	0	1	2	3	4	5	6	7	8	9
1	0	-1	0							
2	0	1	-1	0						
3	0	-1	2	-1	0					
4	0	1	-3	3	-1	0				
5	0	-1	4	-6	4	-1	0			
6	0	1	-5	10	-10	5	-1	0		
7	0	-1	6	-15	20	-15	6	-1	0	
8	0	1	-7	21	-35	35	-21	7	-1	0
9	0	-1	8	-28	56	-70	56	-28	8	-1

Table 4.6: Coefficients $\mathcal{K}_n^i$ for n^{th} derivatives of the inverse of Joe generator for $n \in \{1, \ldots, 9\}$ and $\theta = 1$.

and the formula (4.114) hold for n. □

Tables 4.6 and 4.7 show the $\mathcal{K}_n^i$ coefficients for the Joe copula, with θ equal respectively to 1 and 2.

Using the results of theorems 4.6.4, 4.6.5, 4.6.6 and 4.6.7 with expressions (4.83) and (4.91), it is now possible to compute easily the density of joint distribution build with the Clayton, Frank, Gumbel-Hougaard and Joe copulas.

Original contribution(s) 3. *In this section we give formulas which permit to calculate the joint density, in any dimension, for the Clayton, Frank, Gumbel-Hougaard and Joe families of Archimedean copulas. Then, with these formulas, it is possible to compute the joint density of any joint distribution built on these copulas.*

4.7 Parameters Estimation

Let us suppose that we have a set of realizations of a vector of random variables $\mathbf{X}$: $\{\mathbf{x_1}, \ldots, \mathbf{x_N}\}$. A classical task is to determine the underlying joint distribution H. If we suppose that the parametric family of H is already known, then it "remains" to estimates the parameters of H. Then the most common method for estimating the parameters is maximum likelihood method [Wasserman, 2004].

If $h_\theta(\mathbf{x})$ is the joint density defined by expression (3.20) with parameter θ, then the *likelihood function* is defined by

$$\mathcal{L}(\theta) = \prod_{i=1}^{N} h_\theta(\mathbf{x}_i) \tag{4.115}$$

	i								
n	0	1	2	3	4	5	6	7	8
1	0	-0,5	0						
2	0	0,5	-0,25	0					
3	0	-0,5	0,25	-0,12	0				
4	0	0,5	0,25	0,25	-0,06	0			
5	0	-0,5	-1,75	-1,12	0,12	-0,03	0		
6	0	0,5	5,25	7,75	1,18	0,09	-0,01	0	
7	0	-0,5	-12,75	-41,62	-24,12	-2,25	0,04	-0,01	0
8	0	0,5	28,25	182,25	242,18	71,56	3,09	0,03	-0,003

Table 4.7: Coefficients $\mathcal{K}_n^i$ for n^{th} derivatives of the inverse of Joe generator for $n \in \{1, \ldots, 8\}$ and $\theta = 2$.

and since maxima are unaffected by monotone transformations, one can take the logarithm of this expression to turn it into a sum and then, define the *log-likelihood function* by

$$\mathcal{L}^*(\theta) = \sum_{i=1}^{N} \log h_\theta(\mathbf{x}_i). \tag{4.116}$$

The method of maximum likelihood estimates θ by finding the value of θ that maximizes $\mathcal{L}^*(\theta)$. This is the maximum likelihood estimator (MLE) of θ:

$$\widehat{\theta} = \arg\max_{\theta} \mathcal{L}^*(\theta) \tag{4.117}$$

But in the case of the decomposition of H_θ using copulas the joint density is given by expression (4.83)

$$h_\theta(x_1, \ldots, x_n) = c_{\theta_C}\left(F_{\theta_1}(x_1), \ldots, F_{\theta_n}(x_n)\right) \prod_{i=1}^{n} f_{\theta_j}(x_j)$$

where $\theta = (\theta_1, \ldots, \theta_n; \theta_C)$, and then the log-likelihood become

$$\mathcal{L}^*(\theta) = \sum_{i=1}^{N} \log c_{\theta_C}\left(F_{\theta_1}(x_{1,i}), \ldots, F_{\theta_n}(x_{n,i})\right) + \sum_{i=1}^{N} \sum_{j=1}^{n} \log f_{\theta_j}(x_{j,i}). \tag{4.118}$$

But the MLE method could be time-consuming especially for great values of N. However the use of copulas suggests a two step estimation [Shih and Louis, 1995]:

1. estimate parameters $\theta_1, \ldots, \theta_n$ for the margins,

2. using the above results estimate θ_C.

In other words we have

$$\widehat{\theta}_j = \arg\max_{\theta_j} \sum_{i=1}^{N} \log f_{\theta_j}(x_{j,i}) \tag{4.119}$$

and

$$\widehat{\theta}_C = \arg\max_{\theta_C} \sum_{i=1}^{N} \log c_{\theta_C}\left(F_{\widehat{\theta}_1}(x_{1,i}), \ldots, F_{\widehat{\theta}_n}(x_{n,i})\right). \tag{4.120}$$

This method is called two-stage parametric ML method in [Shih and Louis, 1995] and inference functions for margins (IFM) in [Joe and Xu, 1996] (see also [Joe, 1997]).

Other methods exist: the omnibus estimator (see [Genest et al., 1995], [Shih and Louis, 1995] and [Genest and Werker, 2002]) and a method using the Kendall's tau for Archimedean copulas [Genest and Rivest, 1993], but in the following we will use the IFM method, because it is less time consuming, and more easy to use practically.

4.8 FCDF and copulas

So, from here, we are able to evaluate expression (3.51) using multivariate copulas, and then to try to compute the *fcdf* of an *frv* using the expression (4.1). The expression (3.51), re-writed using (4.73) in conjunction with (4.9), become:

$$\begin{aligned} P[\mathcal{A}_{\mathcal{X},\mathcal{D},n}(u)] &= H_{\mathcal{X}_{t_1},\ldots,\mathcal{X}_{t_n}}(u(t_1),\ldots,u(t_n)) && (4.121)\\ &= C_{\mathcal{X},n}(G_{\mathcal{X},\mathcal{D}}[t_1;u],\ldots,G_{\mathcal{X},\mathcal{D}}[t_n;u]) && (4.122) \end{aligned}$$

where $C_{\mathcal{X},n}$ denotes the copula which links together the margins $G_{\mathcal{X}_{t_k},\mathcal{D}}$ of an *frv* $\mathcal{X}$.

Let us remember that an important class of stochastic process is the class of stationary processes. A stochastic process is said to be *strictly stationary* (see [Burril, 1972] and [Cox and Miller, 1965]) if its distribution do not change with time.

Definition 4.8.1. *A stochastic process X_t is said* stationary *if $\forall t_1, ..., t_k$ and for any h, the joint distribution of $(X_{t_1+h}, \ldots, X_{t_k+h})$ does not depend on hi.e. if its distribution do not change with time.*

As corollary of this definition, the margins of the joint distribution are always the same, i.e. the univariate distributions of X_{t_i} is also the distribution of X_{t_i+h}. Copulas allow us to a broader definition of the stationarity.

Definition 4.8.2. *A stochastic process X_t is said* copula stationary *if $\forall t_1, ..., t_k$ and for any h, the copula of $(X_{t_1+h}, \ldots, X_{t_k+h})$ does not depend on h, i.e. its copula does not change with time.*

That means that if we note F_t the distribution of X_t, then $\forall t_1, ..., t_n \in \mathcal{D}$, in the following expressions

$$C_0 \left(F_{t_1}(t_1), ..., F_{t_k}(t_k)\right)$$

$$C_h \left(F_{t_1+h}(t_1), ..., F_{t_k+h}(t_k)\right)$$

C_h is always equal to C_0, but F_t could be different from F_s, if $s \neq t$. In other words, the distribution of X_t can be change over time, but the relation between the distribution of X_s and in X_t $(s \neq t)$ is always the same.

If an *frv* is also a *copula stationary* stochastic process, then we call it a *copula stationary frv*.

Although any copula can be used for a *copula stationary frv*, Archimedean copulas are really well suited for this kind of stationarity. In this case the expression (4.122) can be rewritten:

$$\begin{aligned} P[\mathcal{A}_{\mathcal{X},\mathcal{D},n}(u)] &= H_{\mathcal{X}_{t_1},...,\mathcal{X}_{t_n}}\left(u(t_1), \ldots, u(t_n)\right) \\ &= C_{\mathcal{X},n}\left(G_{\mathcal{X}_{t_1},\mathcal{D}}[t_1; u], \ldots, G_{\mathcal{X}_{t_n},\mathcal{D}}[t_n; u]\right) \\ &= \psi\left(\sum_{i=1}^{n} \phi\left(G_{\mathcal{X}_{t_i},\mathcal{D}}\left[t_i; u\right]\right)\right). \end{aligned} \tag{4.123}$$

And the associated multivariate density is then given, using expression 4.83 and 4.91, by

$$\begin{aligned} & h_{\mathcal{X}_{t_1},...,\mathcal{X}_{t_n}}\left(u(t_1), \ldots, u(t_n)\right) \\ = \; & \psi^{(n)}\left[\sum_{i=1}^{n} \phi(G_{\mathcal{X}_{t_i},\mathcal{D}}\left[t_i; u\right])\right] \prod_{i=1}^{n} \phi'(G_{\mathcal{X}_{t_i},\mathcal{D}}\left[t_i; u\right]) g_{\mathcal{X}_{t_i},\mathcal{D}}\left[t_i; u\right]. \end{aligned} \tag{4.124}$$

Let $0 < k < n$, then for any $0 \leq i \leq n-k$, the correspondent margins is given by

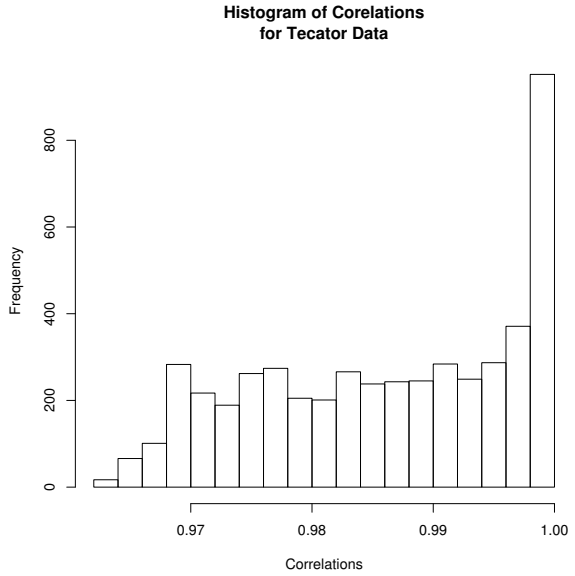
$$\begin{aligned} & H_{\mathcal{X}_{t_1},...,\mathcal{X}_{t_n}}\left(1, \ldots, 1, u(t_{i+1}), \ldots, u(t_{i+k}), 1, \ldots, 1\right) \\ = \; & H_{\mathcal{X}_{t_{1+i}},...,\mathcal{X}_{t_{n+i}}}\left(u(t_{i+1}), \ldots, u(t_{i+k})\right) \\ = \; & \psi\left(\sum_{j=1}^{k} \phi\left(G_{\mathcal{X}_{t_{i+j}},\mathcal{D}}\left[t_{i+j}; u\right]\right)\right) \end{aligned} \tag{4.125}$$

then all k-margins have the same copulas.

When the copula of a *copula stationary frv* is an Archimedean copula generated by a function ϕ, it means that, for any chosen $k \neq l$, the structure

dependence between $\mathcal{X}_{t_k}$ and $\mathcal{X}_{t_l}$ is always the same, and is given by the Archimedean bivariate copula with generator ϕ (see expression 4.92). In this case, we can say that, the *frv* has a "constant dependence structure".

The "dependence structure constance" of *copula stationary frv* given by an Archimedean copula is good approximation model of the dependence structures found in functional data. For the sake of illustration, in figures 4.8, 4.8 and 4.8, we the histograms of the correlations for the discrete versions of the Tecator data (see 2.11 p.21), the growth curves (see fig. 2.9 p.19)and the 2.10 p.20. We can see in these graphs that we have a strong dependence between couples of real random variables $\mathcal{X}_{t_k}$ and $\mathcal{X}_{t_l}$, even if it is with some variations between cases.

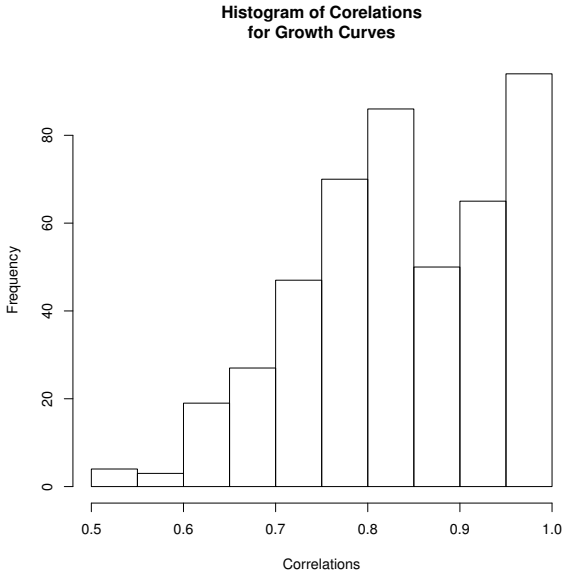


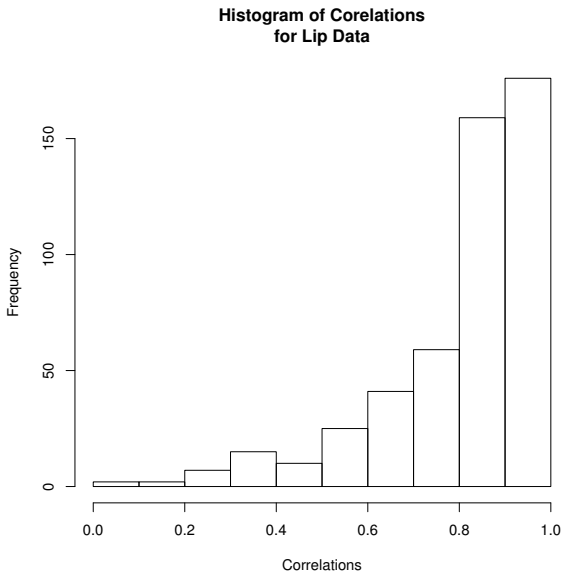
Functional copula stationarity and Archimedean copulas were used in in the framework of *symbolic data analysis* for build finite dimensional distributions of an *frv* for classification using mixture decomposition (see [Vrac et al., 2001] and [Cuvelier and Noirhomme-Fraiture, 2005]).

But even if the results was good using finite dimensional distributions, it is impossible to use the Archimedean copulas to build an *fcdf* using the limit (4.1), indeed we can show this limit is almost always zero for Archimedean copulas when $n \to \infty$ (see [Cuvelier and Noirhomme-Fraiture, 2006])!

Proposition 4.8.1. *Let $u \in \mathcal{I}^{\mathcal{D}}$, and $E = \{t \in \mathcal{D} : \mathbf{G}_{\mathcal{X},\mathcal{D}}[t;u] < 1\}$, and for $n \in \mathbb{N}$, we divide $\mathcal{D}$ in n interval of equal length, i.e.*

- $\left\{t_1^n, \ldots, t_{n+1}^n\right\}$ *are equidistant points of* $\mathcal{D}$
- $t_1^n = \inf(\mathcal{D})$ *and* $t_{n+1}^n = \sup(\mathcal{D})$,


Histogram of Corelations
for Growth Curves
Frequency
0
20
40
60
80
0.5
0.6
0.7
0.8
0.9
1.0
Correlations


Histogram of Corelations
for Lip Data
Frequency
0
50
100
150
0.0
0.2
0.4
0.6
0.8
1.0
Correlations

- $\forall i \in \{1, \ldots, n\} \ \left|t_{i+1}^n - t_i^n\right| = \frac{|\mathcal{D}|}{n} = \Delta_t$.

If $E \neq \emptyset$ *then*

$$\lim_{n\to\infty} \psi \left[\sum_{i=1}^{n} \phi\left(\mathbf{G}_{\mathcal{X},\mathcal{D}}\left[t_i^n; u\right]\right)\right] = 0 \tag{4.126}$$

Proof. First we have the following simplification

$$\begin{aligned}
\psi \left[\sum_{i=1}^{n} \phi\left(\mathbf{G}_{\mathcal{X},\mathcal{D}}\left[t_i^n; u\right]\right)\right] &= \psi \left[\sum_{t_i^n \in E} \phi\left(\mathbf{G}_{\mathcal{X},\mathcal{D}}\left[t_i^n; u\right]\right) + \sum_{t_i^n \notin E} \phi\left(\mathbf{G}_{\mathcal{X},\mathcal{D}}\left[t_i^n; u\right]\right)\right] \\
&= \psi \left[\sum_{t_i^n \in E} \phi\left(\mathbf{G}_{\mathcal{X},\mathcal{D}}\left[t_i^n; u\right]\right)\right].
\end{aligned}$$

Let $p = max\left\{\mathbf{G}_{\mathcal{X},\mathcal{D}}[t; u] | t \in E\right\}$, and so $\forall t \in E$

$$\begin{aligned}
1 > p \geq \mathbf{G}_{\mathcal{X},\mathcal{D}}[t; u] \\
0 < \phi(p) \leq \phi\left(\mathbf{G}_{\mathcal{X},\mathcal{D}}[t; u]\right)
\end{aligned}$$

and thus $\forall n \in \mathbb{N}$ if we write $E_n = \{t_i^n \in E\}$ and suppose that $\# E_n = \nu_n$ we have

$$0 < \nu \cdot \phi(p) \leq \sum_{t_i^n \in E} \phi\left(\mathbf{G}_{\mathcal{X},\mathcal{D}}\left[t_i^n; u\right]\right).$$

We can suppose, without lack of generality that, $|E|/|\mathcal{D}| = \nu_n/n$ for all n, thus it suffice to remark that for any $\varepsilon > 0$, and for $n > \frac{\phi(\varepsilon)}{\phi(p)} \cdot \frac{|\mathcal{D}|}{|E|}$ we have

$$\sum_{t_i^n \in E} \phi\left(\mathbf{G}_{\mathcal{X},\mathcal{D}}\left[t_i^n; u\right]\right) \geq \nu \cdot \phi(p) > \phi(\varepsilon)$$

and then

$$\psi \left[\sum_{t_i^n \in E} \phi\left(\mathbf{G}_{\mathcal{X},\mathcal{D}}\left[t_i^n; u\right]\right)\right] \leq \psi\left[\nu \cdot \phi(p)\right] < \varepsilon$$

□

Another objection to the use of this type of joint distribution is something which we could call *volumetric behavior*: let $p \in]0, 1[$, if we consider the following sequence of interval, square, cube, n-cubes: $[0, p]$,$[0, p]^2$, $[0, p]^3$, $[0, p]^4$,... subsets of respectively $[0, 1]$,$[0, 1]^2$,$[0, 1]^3$,$[0, 1]^4$,... Then the corresponding euclidean measures of volume give the following decreasing sequence p, p^2, p^3, p^4,... whose limit is obviously equal to 0! So for a chosen p there is a sort of "dilution" of $[0, p]^n$ in $[0, 1]^n$ when $n \to \infty$.

In the same way, let us suppose that $U_{\mathcal{X},\mathcal{D},p}$ $(0 < p < 1)$ is a functional quantile (see definition 4.2.3, p. 76) of an *frv* $\mathcal{X}$ defined on $\mathcal{D}$, and suppose that the finite dimensional distribution of $\mathcal{X}$ are computed using expression (4.123), then

$$\begin{aligned} P[\mathcal{A}_{\mathcal{X},\mathcal{D},n}(U_{\mathcal{X},\mathcal{D},p})] &= \psi\left[\sum_{i=1}^{n}\phi\left(\mathbf{G}_{\mathcal{X},\mathcal{D}}\left[t_i^n; U_{\mathcal{X},\mathcal{D},p}\right]\right)\right] \\ &= \psi\left[\sum_{i=1}^{n}\phi(p)\right] \\ &= \psi\left(n\cdot\phi\left(p\right)\right) < p. \end{aligned} \tag{4.127}$$

Worse,

$$\begin{aligned} m > n &\Rightarrow m\phi(p) > m\phi \\ &\Rightarrow \psi(m\phi(p)) < \psi(n\phi) \end{aligned}$$

and then

$$m > n \Rightarrow P[\mathcal{A}_{\mathcal{X},\mathcal{D},m}(U_{\mathcal{X},\mathcal{D},p})] < P[\mathcal{A}_{\mathcal{X},\mathcal{D},n}(U_{\mathcal{X},\mathcal{D},p})]. \tag{4.128}$$

Then, the more we try to have a better approximation for a *functional quantile* of order p, the more we move away from reference value p toward zero. This instability of the estimation of $P[\mathcal{A}_{\mathcal{X},\mathcal{D}}(U_{\mathcal{X},\mathcal{D},p})]$ is very disappointing.

In fact this property is a direct consequence of the Archimedean property, which gives its name to this class of copula. This term was first given by [Ling, 1965]. Let us recall the classical arithmetic Archimedean property [Nelsen, 1999]: for any positive real numbers x, y, there exists an integer n such

$$\underbrace{x + \cdots + x}_{n \text{ terms}} = nx > y.$$

Remember that a bivariate Archimedean copulas can be seen as binary operators (cf. theorems 4.4.4, 4.4.5 and 4.4.6, p.108). With a given generator ϕ we can define a binary operator $\diamond$ on $[0, 1]$:

$$x \diamond y = \psi\left(\phi(x) + \phi(y)\right).$$

This binary operation is commutative and associative (see theorem 4.4.4), and thus can be extended in the following manner:

$$x_1 \diamond \ldots \diamond x_n = \psi\left(\sum \phi(x_i)\right)$$

The operator $\diamond$ defines an implicit operator $\bullet$ on $\mathbb{N} \times [0, 1]$:

$$n \bullet x = \underbrace{x \diamond \cdots \diamond x}_{n \text{ terms}} = \psi\left(n\phi(x)\right)$$

with result in $]0,1]$: for $n \in \mathbb{N}$ and $x \in [0,1]$, we have

$$n\phi(x) \geq \phi(x)$$

and so

$$n \bullet x \leq x.$$

The above results tell us a bad new: Archimedean copulas fail to compute the wanted limit of the finite dimensional distributions (4.1), but the following result tells us a worse new: the limit (4.122) is almost always equal to zero for a large class of copulas, and then, for a large class of classical joint distribution.

Proposition 4.8.2. *Let $\{C_i\}$ be a sequence of copulas such, for all i:*

1. $C_i(\vec{\mathbf{u}}_i) = C_i(u_1, \ldots, u_i) = C_{i+1}(u_1, \ldots, u_{i+1}, 1)$,
2. $C_i < M$,
3. C_i *is strictly monotonic for each variables.*

Then for any sequence $\{u_i\}$ such $\forall i, 0 < u_i < 1$ we have

$$\lim_{i \to \infty} C_i(u_i) = 0. \tag{4.129}$$

Proof. Let $\{u_i\}$ such $\forall i, 0 < u_i < 1$, then using 1 and 2 we have

$$0 < C_i(\vec{\mathbf{u}}_i) < C_{i-1}(\vec{\mathbf{u}}_{i-1}) < 1.$$

Then if we set

$$\varepsilon_i = 1 - \frac{C_i(\vec{\mathbf{u}}_i)}{C_{i-1}(\vec{\mathbf{u}}_{i-1})}$$

then

$$C_i(\vec{\mathbf{u}}_i) = C_{i-1}(\vec{\mathbf{u}}_{i-1})(1 - \varepsilon_i)$$

with $0 < \varepsilon_i < 1$.

Let $\varepsilon_m = \min_i \varepsilon_i$ and $\varepsilon_\delta = \varepsilon_m - \delta$, where $0 < \delta < \varepsilon_m$. Then

$$1 - \varepsilon_\delta < 1 - \varepsilon_i$$

for all i, thus,

$$\begin{aligned} C_i(\vec{\mathbf{u}}_i) &< C_{i-1}(\vec{\mathbf{u}}_{i-1})(1 - \varepsilon_\delta) \\ &< C_{i-2}(\vec{\mathbf{u}}_{i-2})(1 - \varepsilon_\delta)^2 \\ &< \ldots \\ &< (1 - \varepsilon_\delta)^i \end{aligned}$$

and $\lim_{i \to \infty}(1 - \varepsilon_\delta)^i = 0$.

□

Let us notice that, conditions of theorem 4.8.2 are not restrictives at all: the first one[4] permits to have an embedded sequence of copulas, and is easy to fulfill, because by definition of copulas, once we have the copula C_n, all copulas C_k for $k \leq n$ are automatically defined. The second condition requires only, that each off the corresponding random variables are not completely dependent from the others (see theorem 4.5.6), which is the least common case[5]. And the last condition is fulfilled by classical copulas (normal copula, Archimedean copulas,...).

Thus the *fcdf* of a functional quantile of order p , estimated with a finite dimensional distribution is diluted when n growth, because the joint distribution of dimension n is a probability measure of the corresponding n-cube as shown in expression (3.16). And proportionally this n-cube tends to take a lesser part of the total volume when n growth. Thus using a "volumetric probability measure" of dimension n as an approximation of the n^{th} degree of the *fcdf* of an *frv* has an unusual behavior: *the precision vanish to* 0 *when* n *tends to infinity.*

And, like in the case of Archimedean copulas, we have the same instability of the estimation of $P[\mathcal{A}_{\mathcal{X},\mathcal{D}}(U_{\mathcal{X},\mathcal{D},p})]$ with functional quantiles. Let $U_{\mathcal{X},\mathcal{D},p}$ a functional quantile of order p, and let $n < m$:

$$\begin{aligned} P[\mathcal{A}_{\mathcal{X},\mathcal{D},m}(U_{\mathcal{X},\mathcal{D},p})] &= C_m(\underbrace{p,\ldots,p}_{n \text{ times}}, \underbrace{p,\ldots,p}_{m-n \text{ times}}) \\ &< C_m(\underbrace{p,\ldots,p}_{n \text{ times}}, \underbrace{1,\ldots,1}_{m-n \text{ times}}) \\ &= C_n(\underbrace{p,\ldots,p}_{n \text{ times}}) \\ &< C_n(p, \underbrace{1,\ldots,1}_{n-1 \text{ times}}) = p. \end{aligned}$$

Then, for $1 < n < m$, again we have

$$P[\mathcal{A}_{\mathcal{X},\mathcal{D},m}(U_{\mathcal{X},\mathcal{D},p})] < P[\mathcal{A}_{\mathcal{X},\mathcal{D},n}(U_{\mathcal{X},\mathcal{D},p})] < p. \tag{4.130}$$

To illustrate this instability we give in tables 4.8, 4.9 and 4.10, the normal joint distributions of three functional quantiles $U_{\mathcal{X},\mathcal{D},p}$ with the respective values for p: 0.2, 0.5 and 0.8. These finite dimensional distributions are computed in dimensions from 10 to 100, and for different values for the Kendall's tau τ[6]. Then, for a given value p, what is the best approximation of $P[\mathcal{A}_{\mathcal{X},\mathcal{D}}(u_{\mathcal{X},\mathcal{D},p})]$ in these 101 different values? And thus what is the real

[4]This condition is equivalent to the condition (3.48) of Kolmogorov's extension theorem, but for copulas

[5]This condition could be easily relaxed in $C_i < M$, $\forall i \in \mathbb{N} \setminus M$, where M is a finite set for which we could have $C_i \leq M$.

[6]see expression (4.44) p.98 for the relation connecting τ to the classical correlation ρ.

$\tau \setminus n$	10	20	30	40	50	60	70	80	90	100
0.0	e-07	e-14	e-21	e-28	e-35	e-42	e-49	e-56	e-63	e-70
0.1	e-04	e-06	e-07	e-08	e-09	e-09	e-10	e-10	e-10	e-11
0.2	e-03	e-04	e-05	e-05	e-05	e-06	e-06	e-06	e-06	e-06
0.3	e-02	e-03	e-04	e-04	e-04	e-04	e-04	e-04	e-04	e-04
0.4	0.02	e-03	e-03	e-03	e-03	e-03	e-03	e-03	e-03	e-03
0.5	0.03	0.02	0.01	0.01	e-03	e-03	e-03	e-03	e-03	e-03
0.6	0.05	0.04	0.03	0.03	0.02	0.02	0.02	0.02	0.02	0.02
0.7	0.08	0.06	0.06	0.05	0.05	0.05	0.04	0.04	0.04	0.04
0.8	0.11	0.10	0.09	0.09	0.09	0.08	0.08	0.08	0.08	0.08
0.9	0.15	0.15	0.14	0.14	0.14	0.13	0.13	0.13	0.13	0.13
1.0	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20

Table 4.8: Normal joint distribution of a functional quantile $Q_{\mathcal{X},\mathcal{D},p}$ with $p = 0.2$, using several τ and several dimensions n.

$\tau \setminus n$	10	20	30	40	50	60	70	80	90	100
0.0	e-03	e-06	e-09	e-12	e-15	e-18	e-21	e-24	e-27	e-30
0.1	0.01	e-03	4e-4	e-04	4e-5	e-05	e-05	e-05	e-04	e-06
0.2	0.03	0.01	5e-3	2e-3	2e-3	e-03	e-03	e-03	e-03	e-04
0.3	0.07	0.03	0.01	0.01	0.01	0.01	e-02	e-02	0.01	e-02
0.4	0.12	0.07	0.05	0.05	0.03	0.03	0.03	0.02	0.02	0.02
0.5	0.17	0.12	0.11	0.12	0.08	0.07	0.07	0.06	0.07	0.05
0.6	0.23	0.18	0.17	0.15	0.14	0.13	0.12	0.12	0.12	0.11
0.7	0.29	0.26	0.24	0.23	0.21	0.21	0.20	0.20	0.18	0.19
0.8	0.36	0.33	0.33	0.29	0.30	0.30	0.29	0.29	0.29	0.28
0.9	0.43	0.41	0.39	0.39	0.39	0.39	0.39	0.39	0.37	0.38
1.0	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50

Table 4.9: Normal joint distribution of a functional quantile $Q_{\mathcal{X},\mathcal{D},p}$ with $p = 0.5$, using several τ and several dimensions n.

$\tau \setminus n$	10	20	30	40	50	60	70	80	90	100
0.0	0.11	0.01	e-03	e-04	e-05	e-06	e-07	e-08	e-09	e-10
0.1	0.20	0.07	0.04	0.02	0.01	e-03	e-03	e-03	e-03	e-03
0.2	0.28	0.16	0.11	0.08	0.06	0.05	0.04	0.04	0.03	0.03
0.3	0.36	0.25	0.19	0.16	0.14	0.12	0.11	0.10	0.09	0.09
0.4	0.43	0.34	0.29	0.25	0.23	0.22	0.20	0.19	0.18	0.17
0.5	0.51	0.43	0.38	0.36	0.34	0.32	0.31	0.30	0.29	0.28
0.6	0.57	0.51	0.48	0.46	0.44	0.43	0.42	0.41	0.40	0.39
0.7	0.64	0.59	0.57	0.55	0.54	0.53	0.52	0.52	0.51	0.51
0.8	0.69	0.67	0.65	0.64	0.64	0.63	0.63	0.62	0.62	0.61
0.9	0.75	0.74	0.73	0.73	0.72	0.72	0.72	0.72	0.72	0.71
1.0	0.80	0.80	0.80	0.80	0.80	0.80	0.80	0.80	0.80	0.80

Table 4.10: Normal joint distribution of a functional quantile $Q_{\mathcal{X},\mathcal{D},p}$ with $p = 0.8$, using several τ and several dimensions n.

value of $P[\mathcal{A}_{\mathcal{X},\mathcal{D}}(U_{\mathcal{X},\mathcal{D},p})]$? How manage the fact that the probability of the same function has so many different approximations? Having different values with different τ seem acceptable, because it depends of the *frv* $\mathcal{X}$, but the *instability* for the values for the different finite dimensional distributions, depending of the chosen dimension, is more disturbing.

Another objection to the extended use of finite dimensional distributions is the *quasi-impossibility of share the same cylinder*. Let us suppose that $\mathcal{X}(t)$ is an *frv* where $t \in \mathcal{D}$ is the time. And suppose that we can sample in laboratories the realizations of this *frv* with a Δ_t depending of the laboratory's equipment. And suppose also, that three laboratories Lab_1, Lab_2 and Lab_3, did the same sampling experiment, but with different values for Δ_t:

$$\frac{\mathcal{D}}{\Delta_{t_1}} = 90,\ \frac{\mathcal{D}}{\Delta_{t_2}} = 150,\ \frac{\mathcal{D}}{\Delta_{t_3}} = 210,$$

where Δ_{t_1}, Δ_{t_2} and Δ_{t_3} are respectively chosen by Lab_1, Lab_2 and Lab_3. Then we have

$$\Delta_{t_1} = \frac{\mathcal{D}}{90},\ \Delta_{t_2} = \frac{\mathcal{D}}{150},\ \Delta_{t_3} = \frac{\mathcal{D}}{210},$$

and thus, there exists a $\Delta_{t_0} = \mathcal{D}/30$ such

$$\Delta_{t_0} = 3\Delta_{t_1} = 5\Delta_{t_2} = 7\Delta_{t_3}.$$

Then in the best case, i.e. when the initial time measure $t_{i,0}$ is the same for the three laboratories, the common cylinder (see p.49) is generated by the only 30 measure defined when using by Δ_{t_0}, and then only 33% of the Lab_1's measures are used, 20% of the Lab_2's measures and 14% of the Lab_3's measures.

Of course, as $P(t_{1,0} = t_{2,0} = t_{3,0}) = 0$, in most of the cases there will be no common cylinder for the results of the three laboratories, i.e. the worse case is the more likelihood.

A natural objection to the weakness illustrated above, is that all the measures in $\mathcal{D}$ are not always necessary, and we can use the measure in t_i ($\mathcal{X}_{t_i}$) as an approximation for what happens between t_i and t_j. I agree with this idea and we will use it in the following chapter, but it is not the usual interpretation of finite dimensional distributions.

A last objection to the use of finite dimensional distributions is the *accuracy issue*. Table 4.11 show the normal joint distribution of several functional quantiles $Q_{\mathcal{X},\mathcal{D},p}$, with $p = 0.1, \ldots, 0.9$, and several values of τ, with finite distributions of dimension 100[7]. Table 4.11 illustrate the fact that, when the measure of association τ is small then the finite dimensional distributions take very small values. For the sake of comparison, this table was computed with the R system [R Development Core Team, 2005], and

[7]Let us remark that this latter value for the dimensions, is not excessive, because it is the dimension of the Tecator data (see p.21).

$\tau \setminus p$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.0	e-99	e-70	e-53	e-40	e-31	e-23	e-16	e-10	e-05
0.1	e-13	e-11	e-09	e-07	e-06	e-05	e-04	e-03	0.03
0.2	e-07	e-06	e-05	e-04	e-04	e-03	e-03	0.03	0.11
0.3	e-05	e-04	e-04	e-03	e-03	0.01	0.04	0.09	0.22
0.4	e-04	e-03	e-03	0.01	0.02	0.05	0.09	0.17	0.34
0.5	e-03	e-03	0.02	0.03	0.06	0.10	0.17	0.28	0.47
0.6	e-03	0.03	0.04	0.07	0.12	0.18	0.27	0.39	0.58
0.7	0.01	0.04	0.08	0.13	0.19	0.27	0.38	0.51	0.68
0.8	0.03	0.08	0.14	0.20	0.29	0.38	0.49	0.61	0.77
0.9	0.06	0.13	0.21	0.30	0.39	0.49	0.60	0.71	0.84
1.0	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90

Table 4.11: Normal joint distribution of functional quantiles $Q_{\mathcal{X},\mathcal{D},p}$, computed in dimension 100, with several τ and several p

1	.Machine$double.xmin	=	2.225074	e−308
2	.Machine$double.xmax	=	1.797693	e+308
3	.Machine$double.eps	=	2.220446	e−016
4	.Machine$double.neg.eps	=	1.110223	e−016

Table 4.12: Precision of the R system.

in the precision of this system is shown in the table 4.12. The first two values of table 4.12 are respectively the smallest non-vanishing normalized floating-point power of the radix, and the largest finite floating-point number. These values are far away from these in table 4.11, but these latter one are close to the last two values of table 4.12. However .Machine$double.eps and .Machine$double.neg.eps are respectively the smallest positive floating-point number x such that $1 + x \neq 1$ and the small positive floating-point number x such that $1 - x \neq 1$. Thus, when we compute finite dimensional distributions in high dimensions with small value of measure of concordance, we are not sure that the comparison between these small value make sense.

Moreover the time and memory consumption should not be neglected: in dimension 100, a 100×100 matrix of correlation is a necessary parameter for the normal joint distribution , and these 110 were computed in 125 seconds, which far away from real time computing.

Original contribution(s) 4. *In this section we show that, in the classical framework, the limit of a finite dimensional distribution can not leads us to a usable expression for* fcdf. *For this, we show firstly that the limit for any approximation of an* fcdf *built on an Archimedean copula is always equal to zero when n, the number of points of discretization, tends toward infinity. Moreover, we showed the same result for a very broad class of copulas. If*

we do not want to take the limit, and remain in finite dimensions, we show how the estimation of the fcdf *calculated in n points, varies with n. That is what we call the instability in the estimation of the* fcdf *using classical finite dimensional distributions.*

4.9 Conclusion

In this chapter we have firstly introduced the concept of surface of margins which give the probability $P[\mathcal{X}_t \leq y]$ for any t in the domain $\mathcal{D}$ and for any y in the image $\mathcal{I}$ of an *frv* $\mathcal{X}$.

After this, we have introduced the copulas, which permits to divide any joint distribution in two parts: the margins in one hand and the dependence structure in other hand. We have also introduced a particular and important family of copulas: the Archimedean copulas. All these copulas permit to build (or re-build) a joint distribution starting from the margins, thus it is possible to build any finite dimensional distribution of an *frv* $\mathcal{X}$ using surfaces of margins and copulas.

In the last part of the chapter we have shown that copulas are unable to lead directly to a *fcdf* defined in the infinite dimensional space $\mathcal{I}^{\mathcal{D}}$ using the limit of expression (4.1). And this impossibility is shown, firstly in the particular case of Archimedean copulas, and then in a wider case. We have also enumerated several weakness for the use of finite dimensional distributions:

1. the "volumetric behavior" which implies instability of estimations depending of the dimension and the vanishing to 0 when n tends to infinity,

2. the quasi-impossibility, in real situations, to share the same cylinder,

3. the precision issue in case of high dimensionality.

Then, even if they use were of inestimable utility in lack of other general solutions, finite dimensional distributions are not completely satisfactory solutions to build a *fcdf* for a*frv* $\mathcal{X}$.

Chapter 5

Building FCDF Using Quasi Arithmetic-Means

Trouver quelque chose en mathématiques, c'est vaincre une inhibition et une tradition.

Laurent Schwartz

Although this may seem a paradox, all exact science is dominated by the idea of approximation.

Bertrand Russell

5.1 Introduction

In section 3.5 we have seen that the Daniell-Kolmogorov's extension theorem (theorem 3.5.1, p.50,) requires the fulfillment of two compatibility conditions (expressions (3.47) and (3.48)) to prove the existence of a probability on $(\mathbb{R}^\infty, \mathcal{R}^\infty)$. We have also seen how solutions based on this second compatibility condition lead to the instability of the probability of functional quantiles (see proposition 4.128, p.144). In other words it is not possible to define an *fcdf* being the limit of finite dimensional distributions which fulfill the second compatibility condition of the Daniell-Kolmogorov's extension theorem, i.e. the classical framework where the discretized version of an *frv* is considered as a vector of *rrv* is not adapted for probability distributions in the functional case.

Then, we propose an adapted framework with the replacement of the second compatibility condition, by a condition which imposes the stability for the probability of functional quantiles. This change leads to a distribution directly defined in the infinite dimensional space of stochastic process

$\mathcal{X}_t$ ($t \in \mathcal{D}$). Moreover we will prove that this solution will also solve "zero limit" problem met in the proposition 4.8.1 (p.141).

This solution is built firstly on the functional quantiles (see expression (4.11) p.76), and on the fact that, when the surface of margins is strictly increasing for each t, then any function is the limit of sequences of functions built as a sum of pieces of functional quantiles: the quantilized functions.

This solution is built secondly, on the *quasi-arithmetic mean* of given margins. *Quasi-arithmetic means* are extensions of most common means (arithmetic, quadratic, geometric and harmonic means). The first axiomatic characterization of the means was given for a vector of numbers in 1930 by [Nagumo, 1930] and, independently, by [Kolmogorov, 1930], and extended in 1931 to probability distributions by [De Finetti, 1931]. The earliest axiomatic construction for the arithmetic mean was due to [Schiaparelli, 1868] and [Schiaparelli, 1875]. [Aczel, 1966b] and [Aczel, 1966a] used functional equations for the characterization of *quasi-arithmetic means*. A complete review of *quasi-arithmetic means* can be found in [Hardy et al., 1934], [Bullen et al., 1988] or [Beliakov et al., 2007].

The use of *quasi-arithmetic means* in conjunction with the surfaces of margins (see section 4.2) leads firstly to a new type of joint distribution: the *Quasi-Arithmetic Means of Margins* (QAMM) distributions, built with *quasi-arithmetic means* in conjunction with generators of Archimedean copulas. These QAMM distributions are discrete approximations of *Quasi-Arithmetic Means of Margins Limit* (QAMML) distributions, which are probability distributions directly defined in the infinite dimensional space of functions. Built upon generators of Archimedean copulas of dimension higher than 2, QAMM and QAMML distributions share the properties and weaknesses of these latter: the inability to capture structure dependence "less " than the independence. That is why we have to define a generalized QAMML, which can be seen as mean of means.

New solutions bring new problems, and as QAMML are defined directly in the infinite dimensional space of functions, it is not possible anymore to use classical joint density. Our solution is based on the directional derivative defined by René Gâteaux in two papers: [Gâteaux, 1919a] and [Gâteaux, 1922]. We define then a new kind of density: the Gâteaux density. Using this density we propose to estimate parameters of a QAMML distribution using a maximum likelihood method, based on the IFM method for copulas (see section 4.7).

5.2 Quantilized functions

In this chapter, functional quantiles and their "conservation" will lead to a new kind of joint probability distribution. But for this we are going to define a new type of approximation for a given function u belonging to $\mathcal{I}^{\mathcal{D}}$.

Let us suppose that we know a function u in n points of $\mathcal{D}$: $t_1, \ldots, t_n$: $u(t_1), \ldots, u(t_n)$. How to deal with the unknown values of u, i.e. the values $u(t)$ for any $t \notin \{t_1, \ldots, t_n\}$?

In real analysis, the concept of simple function answer to this question in the framework of the definition of the Lebesgues integral (see [Rudin, 1987] or [Bourbaki, 1967]). A simple function is finite linear combination of indicator functions on measurable sets. If $A_1, \ldots, A_n$ is a partition of $\mathcal{D}$, and $a_1, \ldots, a_n$ a sequence of real numbers, then a simple function is a function of the form

$$f(t) = \sum_{k=1}^{n} a_k \mathbf{1}_{A_k}(t), \quad \forall t \in \mathcal{D} \tag{5.1}$$

where $\mathbf{1}_A(t)$ is the indicator function of A.

With this concept, a function u known in $t_1, \ldots, t_n$ is easily approximated with the following simple function:

$$u_n(t) = \sum_{k=1}^{n} u(t_k) \mathbf{1}_{[t_k, t_{k+1}[}(t). \tag{5.2}$$

And then, for any $t \notin \{t_1, \ldots, t_n\}$: $u(t) = u(t_k)$, where k is such $t \in [t_k, t_{k+1}[$.

In other words, in real analysis, it seems natural that the approximation (simple) function u_n is constant for an "unknown" t.

In the same way, in probability, when we try to compute $P[u \leq_{\mathcal{D}} \mathcal{X}]$ using a finite dimensional distribution $H(t_1, \ldots, t_n)$, an acceptable approximation for u could be a function u_n, such for any $t \notin \{t_1, \ldots, t_n\}$: $P[u(t) \leq \mathcal{X}_t] = P[u(t_k) \leq \mathcal{X}_{t_k}]$, where k is such $t \in [t_k, t_{k+1}[$.

Remark 5.2.1. *In the following, for the ease of notations, we will suppose that if $p \neq q$, then $V_{\mathcal{X},\mathcal{D},p}$ and $V_{\mathcal{X},\mathcal{D},q}$ will be two different functional quantiles.*

We define this type of approximation function, and call it quantilized function, a finite "functional" combination of indicators using the functional quantiles.

Definition 5.2.1. *Let $\mathcal{A} = \{A_1, \ldots, A_n\}$ be a partition of $\mathcal{D}$, and $\mathbf{p} = (p_1, \ldots, p_n) \in]0,1[^n$. We define a* Quantilized Function *based on $\mathcal{A}$ and $\mathbf{p}$ as follow:*

$$f(t) = \sum_{k=1}^{n} U_{\mathcal{X},\mathcal{D},p_k}(t) \mathbf{1}_{A_k}(t), \quad \forall t \in \mathcal{D}, \tag{5.3}$$

where for each k, $U_{\mathcal{X},\mathcal{D},p_k}$ is a functional quantile of order p_k for the frv $\mathcal{X}$ *defined on $\mathcal{D}$.*

Figures 5.1 and 5.2 show two examples of quantilized functions, based on the surface of margins of the Tecator dataset (see

fig. 2.11, p. 21) using the Normal distribution for computing $G_{\mathcal{X},\mathcal{D}}$ (see expression 4.4, p. 55). In figure 5.1, we have chosen (arbitrarily) $\mathcal{A} = \{[850, 900[, [900, 950[,[950, 1000[, [1000, 1048[\}$ and $\mathbf{p} = (0.35, 0.85, 0.2, 0.7)$. The figure shows the quantilized function, made of bold lines, and for comparison, nine level curves (functional quantiles of values: $0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8$ and 0.9) with thin lines. We also show in bold, the bounds of elements of $\mathcal{A}$. And in figure 5.2, $\mathcal{A} = \{[850, 875[, [875, 900[,[900, 925[, [925, 950[, [950, 975[, [975, 1000[,[1000, 1025[, [1025, 1048[\}$ and $\mathbf{p} = (0.35,0.45, 0.55,0.65, 0.75,0.6, 0.5, 0.4)$. The use of the normal distribution to compute $G_{\mathcal{X},\mathcal{D}}(t, .) = F_{\mathcal{X}_t}(.)$ allows us to only care about the lowest functional quantiles $Q_{\mathcal{X},\mathcal{D}}$ (see expression 4.12 p.76, and remark 4.2.1, p. 77).

Then for any $u \in \mathcal{I}^{\mathcal{D}}$, only known in n points of $\mathcal{D}$: $\{t_1, \ldots, t_n\}$, a natural "quantilized" approximation of the function u is given by:

$$u_n(t) = \sum_{i=1}^{n} V_{\mathcal{X},\mathcal{D},p_i}(t)\mathbf{1}_{[t_i,t_{i+1}[}(t) \tag{5.4}$$

where for all $i \in \{1, \ldots, n\}$, $p_i = G_{\mathcal{X},\mathcal{D}}[t_i; u]$ and $V_{\mathcal{X},\mathcal{D},p_i}$ is a functional quantile of order p_i.

Now, let us prove that, when $G_{\mathcal{X},\mathcal{D}}(t, .) = F_{\mathcal{X}_t}(.)$ is continuous and strictly increasing, then any function of $\mathcal{I}^{\mathcal{D}}$ is the limit of sequences of functions built on the quantile functions.

Proposition 5.2.1. *If $G_{\mathcal{X},\mathcal{D}}(t, y)$ is continuous and strictly increasing for any $t \in \mathcal{D}$, then any function $v \in \mathcal{I}^{\mathcal{D}}$ is the limit of two sequences of quantilized functions, one increasing and the other decreasing. These two sequences are uniformly convergent toward v.*

Proof. Let $v \in \mathcal{I}^{\mathcal{D}}$ and $i \in \mathbb{N}$, then we define the following partition of $[0, 1]$:

$$\begin{aligned} I_{i,k} &= \left[\tfrac{k-1}{2^i}, \tfrac{k}{2^i}\right[\quad if\ k = 1, \ldots, 2^i - 1 \\ &= \left[\tfrac{2^i-1}{2^i}, 1\right] \quad if\ k = 2^i. \end{aligned}$$

With this partition of $[0, 1]$ we can define a partition on $\mathcal{D}$:

$$B_{i,k} = \left\{ t \in \mathcal{D} : \frac{k-1}{2^i} \leq \mathbf{G}_{\mathcal{X},\mathcal{D}}[t; v] < \frac{k}{2^i} \right\} \tag{5.5}$$

which can be written symbolically: $B_{i,k} = {\mathbf{G}_{\mathcal{X},\mathcal{D}}}^{-1}[I_{i,k}; v]$. We can define now the two following quantilized functions:

$$u_i(t) = \sum_{k=1}^{2^i} Q_{\frac{k-1}{2^i}}(t)\ \mathbf{1}_{B_{i,k}}(t), \tag{5.6}$$

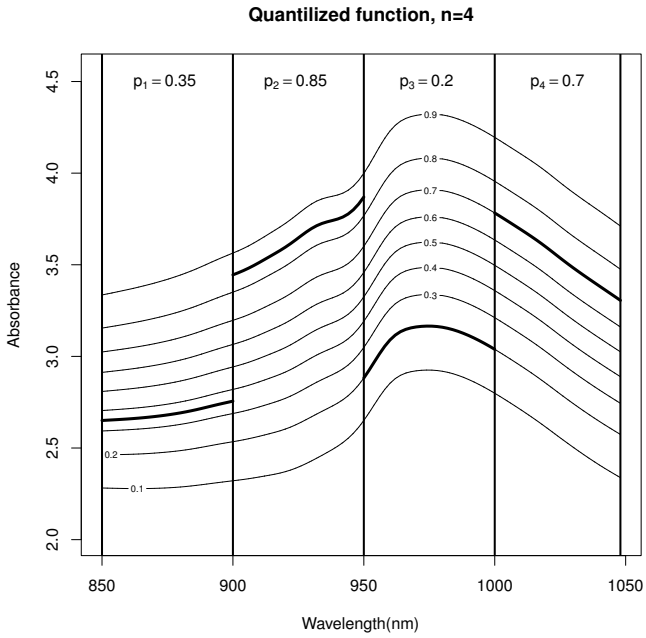


Figure 5.1: Quantilized function with $\mathcal{A}$ ={[850, 900[,[900, 950[,[950, 1000[, [1000, 1048[} and $\mathbf{p} = (0.35,0.85,0.2,0.7)$.

$$w_i(t) = \sum_{k=1}^{2^i} Q_{\frac{k}{2^i}}(t)\ \mathbf{1}_{B_{i,k}}(t). \tag{5.7}$$

First, let us show that u_i is an increasing sequence. We prove this property for a fixed t_0, chosen randomly in $\mathcal{D}$. Thus for $t_0 \in \mathcal{D}$, there exists one, and only one $1 \leq k_0 \leq 2^i$ such $t_0 \in B_{i,k_0}$, and then

$$u_i(t_0) = \sum_{k=1}^{2^i} Q_{\frac{k-1}{2^i}}(t_0)\ \mathbf{1}_{B_{i,k}}(t_0) = Q_{\frac{k_0-1}{2^i}}(t_0).$$

By construction, $B_{i,k}$ sets, are such

$$B_{i,k_0} = B_{i+1,2k_0-1} \cup B_{i+1,2k_0}.$$

Then we have

$$\begin{aligned} u_{i+1}(t_0) &= Q_{\frac{2k_0-2}{2^{i+1}}}(t_0) \quad if\ t_0 \in B_{i+1,2k_0-1}, \\ &= Q_{\frac{2k_0-1}{2^{i+1}}}(t_0) \quad if\ t_0 \in B_{i+1,2k_0}, \end{aligned}$$

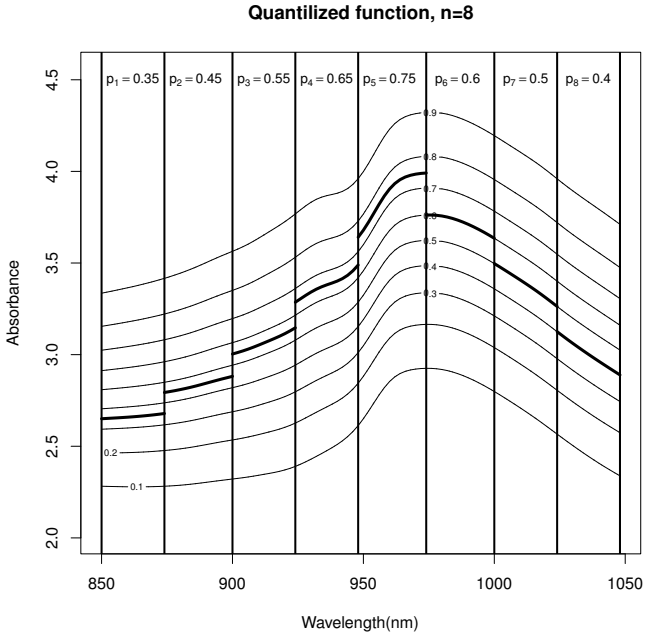


Figure 5.2: Quantilized function with $\mathcal{A} = \{[850, 875[, [875, 900[, [900, 925[, [925, 950[, [950, 975[, [975, 1000[, [1000, 1025[, [1025, 1048[\}$ and $\mathbf{p} = (0.35, 0.45, 0.55, 0.65, 0.75, 0.6, 0.5, 0.4)$.

and thus, using proposition 4.2.1:

$$\begin{aligned} u_{i+1}(t_0) &= Q_{\frac{k_0-1}{2^i}}(t_0) \quad if\ t_0 \in B_{i+1,2k_0-1}, \\ &> Q_{\frac{k_0-1}{2^i}}(t_0) \quad if\ t_0 \in B_{i+1,2k_0}, \end{aligned}$$

and then

$$u_{i+1}(t_0) \geq u_i(t_0).$$

And the reasoning can be used for any $t_0 \in \mathcal{D}$. The decreasing property of $w_i(t)$ is proven similarly.

Now, let us prove the uniformly convergence, i.e.

$$\lim_{i\to\infty} ||v - u_i||_\infty = 0$$

where for a given $u \in \mathcal{I}^{\mathcal{D}}$

$$||u||_\infty = \sup\{\, |u(t)| : t \in \mathcal{D}\}\,.$$

First, if $\mathbf{G}_{\mathcal{X},\mathcal{D}}{}^u$ denotes the function $\mathcal{D} \to [0,1] : t \mapsto \mathbf{G}_{\mathcal{X},\mathcal{D}}[t;u]$, let us remark that, as $G_{\mathcal{X},\mathcal{D}}(t,y)$ is continuous and strictly increasing for each $t \in \mathcal{D}$, we have

$$||\mathbf{G}_{\mathcal{X},\mathcal{D}}{}^u - \mathbf{G}_{\mathcal{X},\mathcal{D}}{}^v||_\infty = 0 \quad \Leftrightarrow \quad u(t) = v(t) \quad \forall t \in \mathcal{D} \tag{5.8}$$

indeed

$$\begin{array}{lll} ||\mathbf{G}_{\mathcal{X},\mathcal{D}}{}^u - \mathbf{G}_{\mathcal{X},\mathcal{D}}{}^v||_\infty = 0 & \Leftrightarrow & |\mathbf{G}_{\mathcal{X},\mathcal{D}}[t;u] - \mathbf{G}_{\mathcal{X},\mathcal{D}}[t;v]| = 0 \quad \forall t \in \mathcal{D} \\ & \Leftrightarrow & \mathbf{G}_{\mathcal{X},\mathcal{D}}[t;u] = \mathbf{G}_{\mathcal{X},\mathcal{D}}[t;v] \quad \forall t \in \mathcal{D} \\ & \Leftrightarrow & u(t) = v(t) \quad \forall t \in \mathcal{D}. \end{array}$$

So, we will prove that

$$\lim_{i\to\infty} ||\mathbf{G}_{\mathcal{X},\mathcal{D}}{}^u - \mathbf{G}_{\mathcal{X},\mathcal{D}}{}^{u_i}||_\infty = 0.$$

Let $0 < \varepsilon < 1$, if we set $i = \left\lfloor \log_2(\frac{1}{\varepsilon}) \right\rfloor + 1$, then $2^{-i} < \varepsilon$. For a randomly chosen $t_0 \in \mathcal{D}$, $\exists!\ 1 \leq k_0 \leq 2^i$ such $t_0 \in B_{i,k_0}$. Thus

$$u_i(t_0) = Q_{\frac{k_0-1}{2^i}}(t_0)$$

and

$$\mathbf{G}_{\mathcal{X},\mathcal{D}}[t_0;u_i] = \frac{k_0-1}{2^i}$$

and by construction of Bi,k sets

$$\mathbf{G}_{\mathcal{X},\mathcal{D}}[t_0;v] < \frac{k_0}{2^i}$$

so

$$\mathbf{G}_{\mathcal{X},\mathcal{D}}[t_0; v] - \mathbf{G}_{\mathcal{X},\mathcal{D}}[t_0; u_i] \leq \frac{1}{2^i} < \varepsilon.$$

And, again, as the reasoning can be used for any $t_0 \in \mathcal{D}$ for the same value i, then we have proved the proposition for the increasing sequence of quantilized functions. The same demonstration can be held for the decreasing sequence. □

In the preceding proposition, the quantilized functions are built upon intervals $B_{i,k}$, these latter built on intervals of equal lengths belonging in $[0, 1]$. The relation $B_{i,k} = {\mathbf{G}_{\mathcal{X},\mathcal{D}}}^{-1}[I_{i,k}; v]$ (see expression 5.5) do not assures us that $B_{i,k}$ is an interval (a compact). In the following proposition we are going to prove that, if we divide $\mathcal{D}$ in intervals of equal lengths we can also build two convergent sequences of quantilized functions.

Proposition 5.2.2. *Let $v \in \mathcal{I}^{\mathcal{D}}$, $a = \min(\mathcal{D})$ and $b = \max(\mathcal{D})$. For $i \in \mathbb{N}$, we define the following partition for $\mathcal{D}$:*

$$\begin{array}{rcll} C_{i,l} & = & \left[a + \frac{l-1}{2^i}(b-a), a + \frac{l}{2^i}(b-a)\right[& if\ l = 1, \ldots, 2^i - 1 \\ & = & \left[a + \frac{2^i-1}{2^i}(b-a), b\right] & if\ l = 2^i. \end{array} \tag{5.9}$$

Let, also,

$$p_{i,l} = \min_{t \in C_{i,l}} \mathbf{G}_{\mathcal{X},\mathcal{D}}[t; v] \quad \& \quad q_{i,l} = \max_{t \in C_{i,l}} \mathbf{G}_{\mathcal{X},\mathcal{D}}[t; v].$$

If $G_{\mathcal{X},\mathcal{D}}(t, .) = F_{\mathcal{X}_t}(.)$ is continuous and strictly increasing for each $t \in \mathcal{D}$, and moreover, if the function ${\mathbf{G}_{\mathcal{X},\mathcal{D}}}^{v} : \mathcal{D} \to [0, 1] : t \mapsto \mathbf{G}_{\mathcal{X},\mathcal{D}}[t; v]$ is continuous then, v is the limit of two following sequences of a quantilized functions:

$$u_i(t) = \sum_{l=1}^{2^i} Q_{p_{i,l}}(t)\ \mathbf{1}_{C_{i,l}}(t), \tag{5.10}$$

$$w_i(t) = \sum_{l=1}^{2^i} Q_{q_{i,l}}(t)\ \mathbf{1}_{C_{i,l}}(t). \tag{5.11}$$

Moreover u_n is an increasing sequence, w_n is a decreasing sequence, and the two sequences are uniformly convergent toward v.

Proof. As in the preceding proof, let us, first, show that u_i is an increasing sequence. We prove this property for a fixed t_0, chosen randomly in $\mathcal{D}$. Thus for $t_0 \in \mathcal{D}$, then $\exists!\ 1 \leq l_0 \leq 2^i$ such $t_0 \in C_{i,l_0}$, and then

$$u_i(t_0) = \sum_{l=1}^{2^i} Q_{p_{i,l}}(t_0)\ \mathbf{1}_{C_{i,l}}(t_0) = Q_{p_{i,l_0}}(t_0).$$

By the construction of the $C_{i,l}$ sets, we have

$$C_{i,l_0} = C_{i+1,2l_0-1} \cup C_{i+1,2l_0}.$$

Then we have

$$\begin{aligned} u_{i+1}(t_0) &= Q_{p_{i+1,2l_0-1}}(t_0) \quad if\ t_0 \in C_{i+1,2l_0-1}, \\ &= Q_{p_{i+1,2l_0}}(t_0) \quad if\ t_0 \in C_{i+1,2l_0}. \end{aligned}$$

But we have the following relation

$$p_{i,l_0} = \min(p_{i+1,2l_0-1}, p_{i+1,2l_0})$$

and thus, using proposition 4.2.1:

$$u_{i+1}(t_0) \geq u_i(t_0).$$

And the reasoning can be used for any $t_0 \in \mathcal{D}$. The decreasing property of $w_i(t)$ is proven similarly.

As in the preceding proof, it is sufficient to prove that $\lim_{i\to\infty} ||\mathbf{G}_{\mathcal{X},\mathcal{D}}{}^v - \mathbf{G}_{\mathcal{X},\mathcal{D}}{}^{u_i}||_\infty = 0$.

Let $0 < \varepsilon < 1$. For $j \in \mathbb{N}$ we define:

$$\begin{aligned} I_{j,k} &= \left[\tfrac{k}{2^j}, \tfrac{k+1}{2^j}\right[\quad if\ k = 0, \ldots, 2^j - 2 \\ &= \left[\tfrac{2^j-1}{2^j}, 1\right] \quad if\ k = 2^j - 1. \end{aligned}$$

If we set $j = \left\lfloor \log_2(\frac{1}{\varepsilon}) \right\rfloor + 1$, then $2^{-j} < \varepsilon$. Thus if we define also $B_{j,k} = (\mathbf{G}_{\mathcal{X},\mathcal{D}}{}^v)^{-1}[I_{j,k}]$, as in the preceding proof, then for any $t_1, t_2 \in B_{j,k}$, we have $|\mathbf{G}_{\mathcal{X},\mathcal{D}}[t_1; v] - \mathbf{G}_{\mathcal{X},\mathcal{D}}[t_2; v] < 2^{-j}| < \varepsilon$.

As $\mathbf{G}_{\mathcal{X},\mathcal{D}}{}^v$ is continuous, the $B_{j,k}$ is also an interval, and we will denote it $[b_{j,k}, b_{j,k+1}[$. Now, if we set

$$i = \left\lfloor \log_2 \left(\frac{b-a}{\min_k |B_{j,k}|} \right) \right\rfloor + 1$$

then

$$\frac{b-a}{2^i} < \min_k |B_{j,k}|$$

i.e. the length of intervals $C_{i,l}$ are less than the minimal length of sets $B_{j,k}$. However, it is not sure that each intervals $C_{i,l}$ is completely included in an interval Bj,k, as shown in figure 5.3. To ensure that, we can denote

$$c_{i,l} = a + \frac{l-1}{2^i}(b-a)$$

where $l = 1, \ldots, 2^i$, then if we set

$$n = \left\lfloor \frac{1}{\min\{|b_{j,k} - c_{i,l}| > 0,\ 1 \leq k \leq 2^j, 1 \leq l \leq 2^i\}} \right\rfloor + 1$$

then $||\mathbf{G}_{\mathcal{X},\mathcal{D}}{}^v - \mathbf{G}_{\mathcal{X},\mathcal{D}}{}^{u_n}||_\infty < \varepsilon$. The same proof can be held for w_n. □

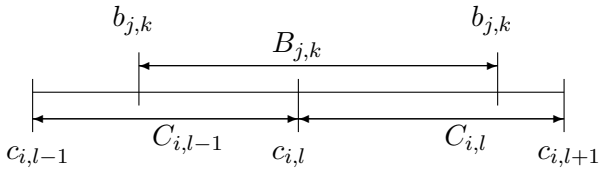


Figure 5.3: Possible relative positions of $B_{j,k}$, and the $C_{i,l}$.

This result is more general that it sounds, because by the Lusin's theorem [Lusin, 1912], if ${\mathbf{G}_{\mathcal{X},\mathcal{D}}}^v$ is measurable, then given ε,there exists a compact $E \subset [a,b]$ such that v restricted to E is continuous and

$$\mu(E^C) < \varepsilon.$$

where E^C denotes the complement of E.

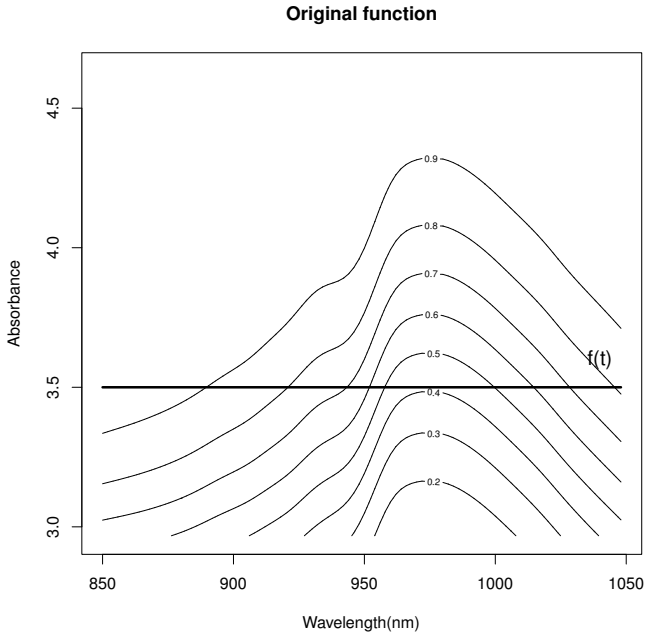


Figure 5.4: A constant function to be quantilized.

If for some $t \in \mathcal{D}$, $G_{\mathcal{X},\mathcal{D}}(t,.) = F_{\mathcal{X}_t}(.)$ is not strictly increasing, then the relation 5.8, is not true anymore, and we have only the following implication

$$||{\mathbf{G}_{\mathcal{X},\mathcal{D}}}^u - {\mathbf{G}_{\mathcal{X},\mathcal{D}}}^v||_\infty = 0 \quad \Leftarrow \quad u(t) = v(t) \quad \forall t \in \mathcal{D}. \tag{5.12}$$

Then, for a given function v and a sequence of functions v_i built using expressions 5.6 or 5.7, we only have the following implication

$$\lim_{i\to\infty} ||{\mathbf{G}_{\mathcal{X},\mathcal{D}}}^v - {\mathbf{G}_{\mathcal{X},\mathcal{D}}}^{v_i}||_\infty = 0 \Rightarrow \mathbf{G}_{\mathcal{X},\mathcal{D}}[t;v] = \lim_{i\to\infty} \mathbf{G}_{\mathcal{X},\mathcal{D}}[t;v_i] \ \forall t \in \mathcal{D}.$$

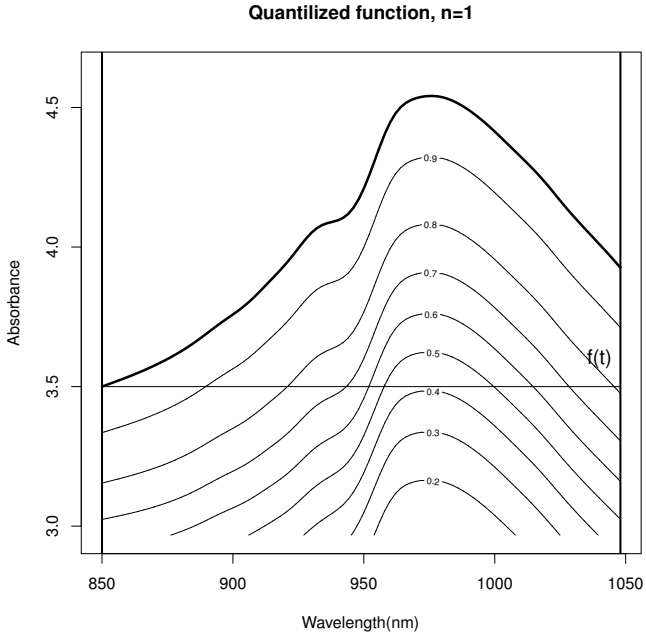


Figure 5.5: Quantilized version of a constant function with 1 part.

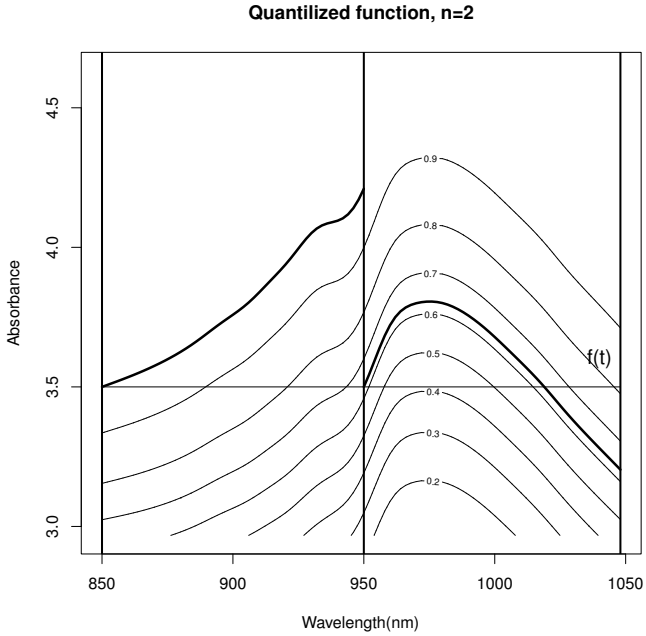


Figure 5.6: Quantilized version of a constant function with 2 parts.

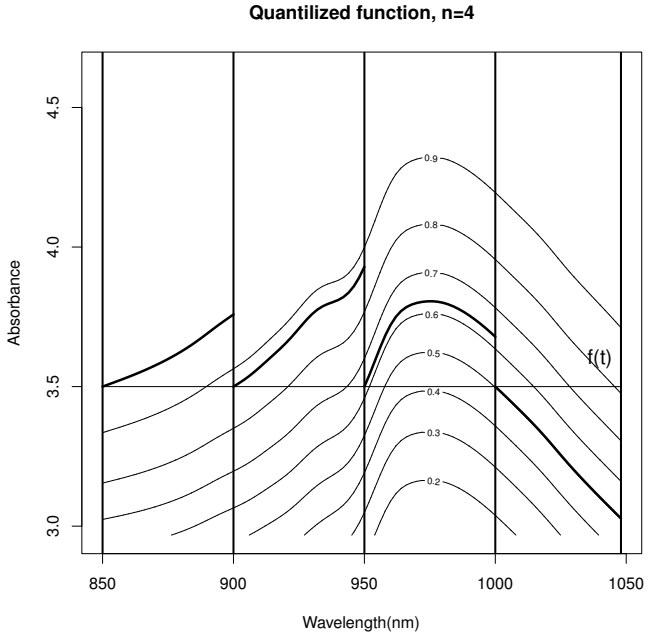


Figure 5.7: Quantilized version of a constant function with 4 parts.

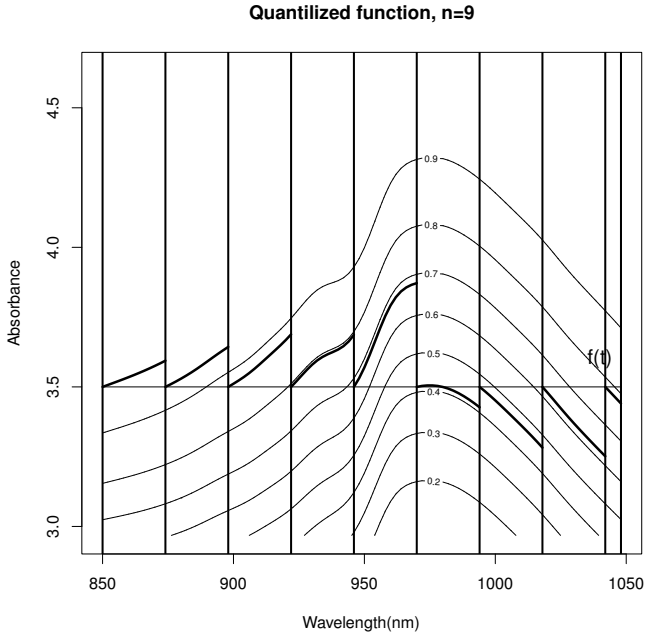


Figure 5.8: Quantilized version of a constant function with 9 parts.

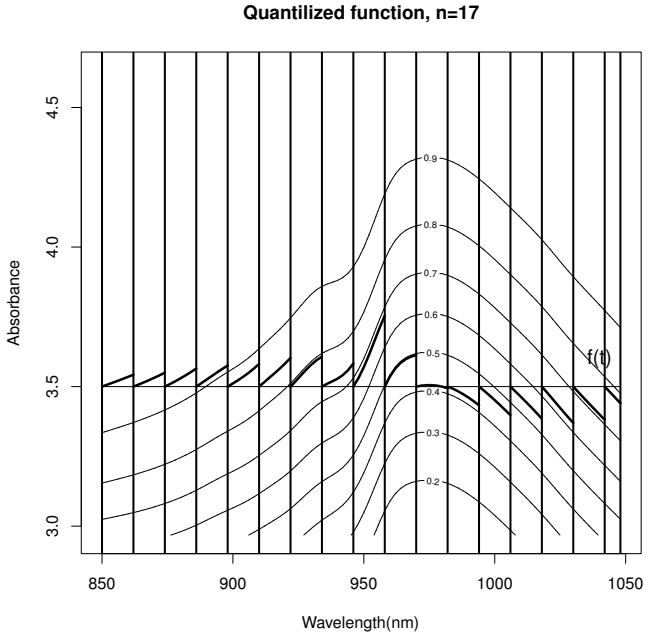


Figure 5.9: Quantilized version of a constant function with 17 parts.

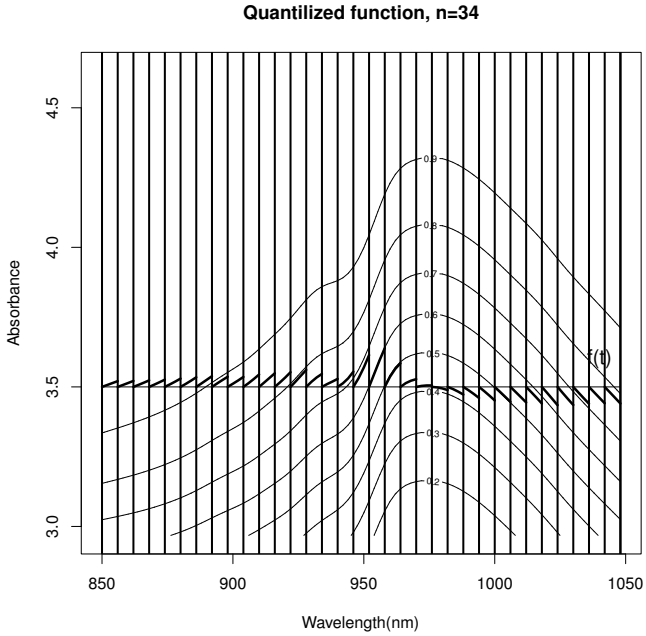


Figure 5.10: Quantilized version of a constant function with 35 parts.

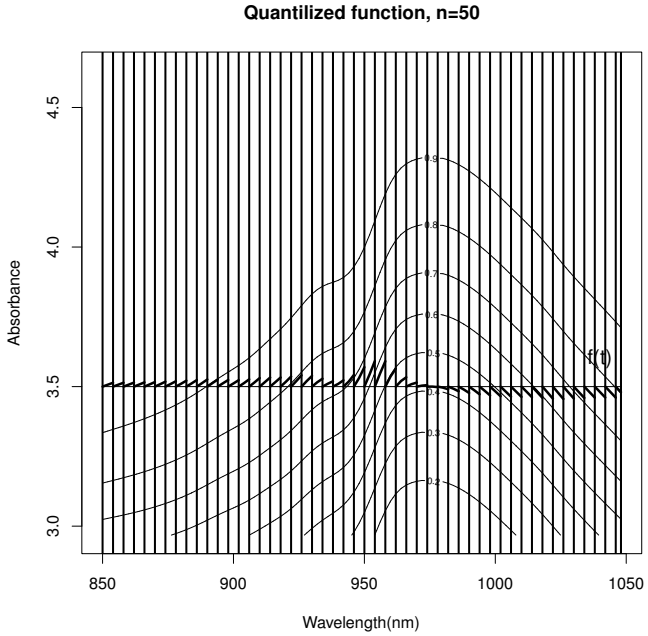


Figure 5.11: Quantilized version of a constant function with 50 parts.

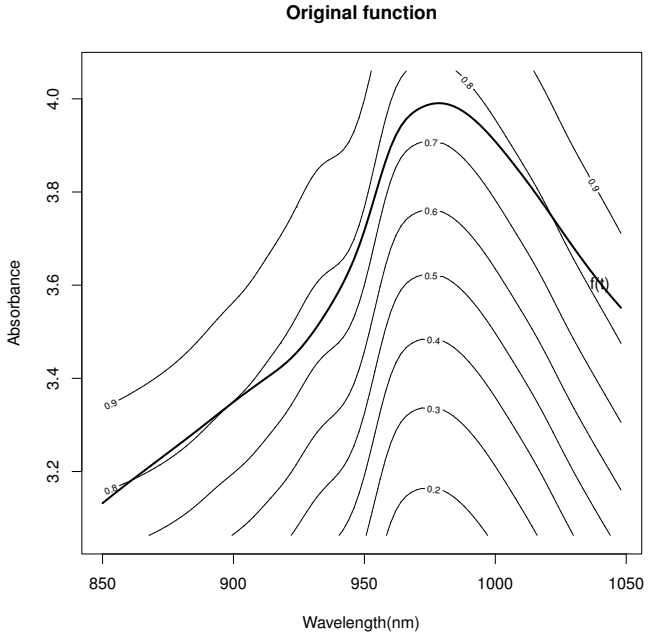


Figure 5.12: A function of the Tecator dataset to be quantilized.

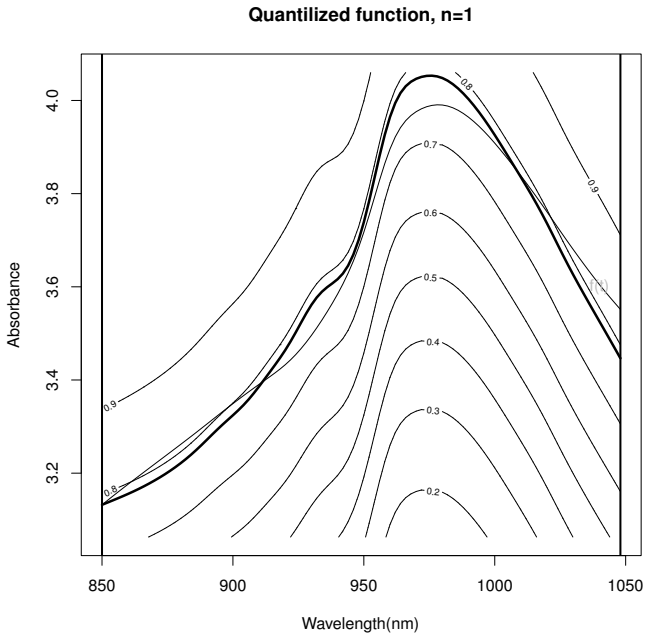


Figure 5.13: Quantilized version of a function of the Tecator dataset with 1 part.

We do not prove it here, but a similar proof could be used in this case for the proposition 5.2.1 if, instead of expressions 5.6 and 5.7, we use the following quantilized functions:

$$u_i(t) = \sum_{k=1}^{2^i} U_{\frac{k-1}{2^i}}(t)\ \mathbf{1}_{B_{i,k}}(t), \tag{5.13}$$

$$w_i(t) = \sum_{k=1}^{2^i} W_{\frac{k}{2^i}}(t)\ \mathbf{1}_{B_{i,k}}(t). \tag{5.14}$$

Of course, the above expressions 5.13 and 5.14 define two infinities of sequences u_i and w_i such, respectively, ${\mathbf{G}_{\mathcal{X},\mathcal{D}}}^{u_i} \to {\mathbf{G}_{\mathcal{X},\mathcal{D}}}^{v}$ and ${\mathbf{G}_{\mathcal{X},\mathcal{D}}}^{w_i} \to {\mathbf{G}_{\mathcal{X},\mathcal{D}}}^{v}$, when $i \to \infty$. In both case it should be possible to find at most a sequence such, respectively, $u_i \to v$ and $w_i \to v$, when $i \to \infty$.

To illustrate the fact that, when $G_{\mathcal{X},\mathcal{D}}(t,.) = F_{\mathcal{X}_t}(.)$ is continuous and strictly increasing, then any function of $\mathcal{I}^{\mathcal{D}}$ is the limit of quantilized functions, we give hereafter, seven quantilized versions of two different functions. These quantilized versions are based on the functional quantiles shown given using expression 4.4 (p. 55) for the surface of margins of the Tecator dataset (see figure 2.11, p.21).

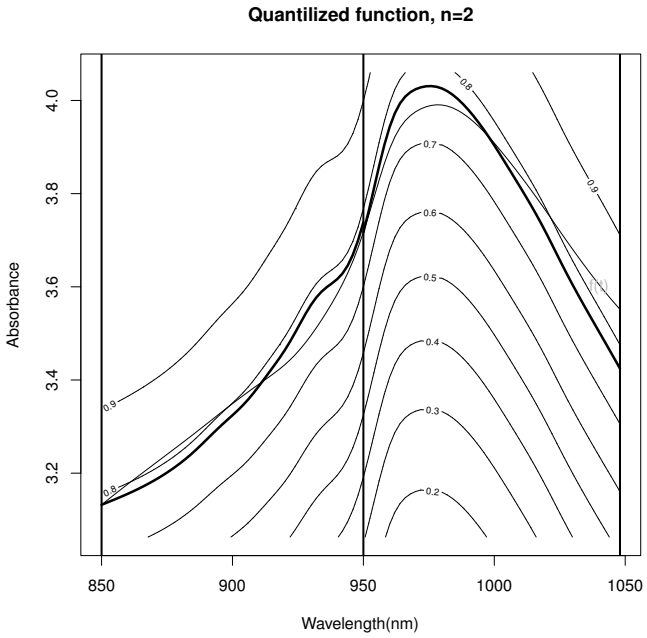


Figure 5.14: Quantilized version of a function of the Tecator dataset with 2 parts.

Figure 5.4 shows the first function: it is a simple constant function $f(t) = 3.5$. Figures from 5.5 to 5.11 show quantilized versions (or approximations) of this constant function in bold line(s), with respectively 1, 2, 4, 9, 17, 35 and 50 parts. Figures show also, for the sake of illustration, nine level curves $(Q_{\mathcal{X},\mathcal{D},0.1}, Q_{\mathcal{X},\mathcal{D},0.2}, \ldots, Q_{\mathcal{X},\mathcal{D},0.9})$ with thin lines. We see clearly how the sequence of quantilized functions tends toward the chosen function.

Figure 5.12 shows the second function which is one function of the Tecator dataset. Figures from 5.13 to 5.19 show quantilized versions (or approximations) of this function, with respectively 1, 2, 4, 9, 17, 35 and 50 parts. Again we see the sequence of quantilized functions tending toward the chosen function.

Original contribution(s) 5. *In this section we define the concept of quantilized functions. In real analysis a simple function is a piecewise constant function, while a quantilized function is a piecewise functional quantile function. Any discretized version of a function defines an implicit quantilized function: the function composed of the functional quantiles whose values are given by margins calculated at the points of discretization. We have shown that, when the surface of margins is strictly increasing for each t, then any function is the limit of at most a sequence of quantilized functions. Quantilized functions permit to consider a functional object (a function or an* frv*)*

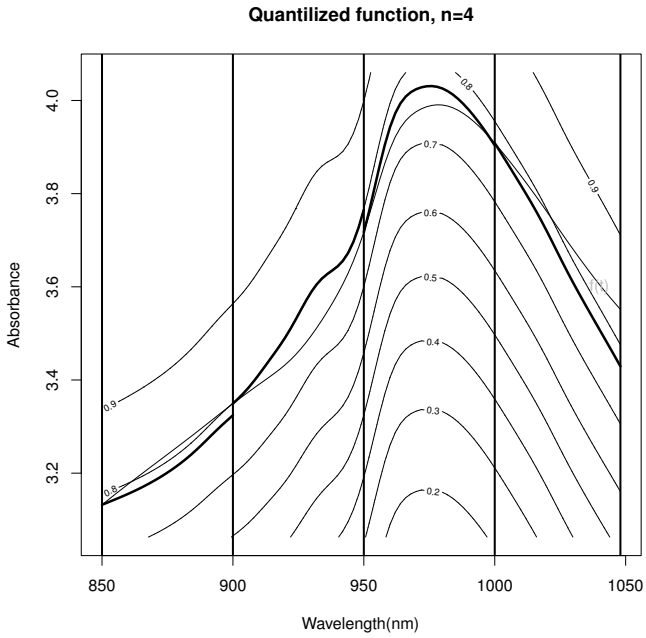


Figure 5.15: Quantilized version of a function of the Tecator dataset with 4 parts.

always as a whole, even if we only know it in a discretized version. Then, the values of a discretized version of a functional object are not anymore considered as vector of rrvs*, but as the determinative parts of a quantilized function.*

5.3 Fractal and quantile based distributions

Let us remember how the *fcdf* of an *frv* is defined by the expression (3.30) (p.44):

$$\begin{aligned} \mathcal{F}_{\mathcal{X},\mathcal{D}}(u) &= P[\mathcal{X} \leq_{\mathcal{D}} u] \\ &= P[\mathcal{X}(t) \leq u(t), \forall t \in \mathcal{D}]. \end{aligned}$$

And this distribution is the probability of the set defined by the expression (3.34) (p.44)

$$\mathcal{A}_{\mathcal{X},\mathcal{D}}(u) = \{\omega \in \Omega : \mathcal{X} \leq_{\mathcal{D}} u\}.$$

Remember also that the classical approach, uses instead the following set (expression (3.50), p.51):

$$\mathcal{A}_{\mathcal{X},\mathcal{D},n}(u) = \bigcap_{i=1}^{n} \{\omega \in \Omega : \mathcal{X}(t_i, \omega) \leq u(t_i)\}$$

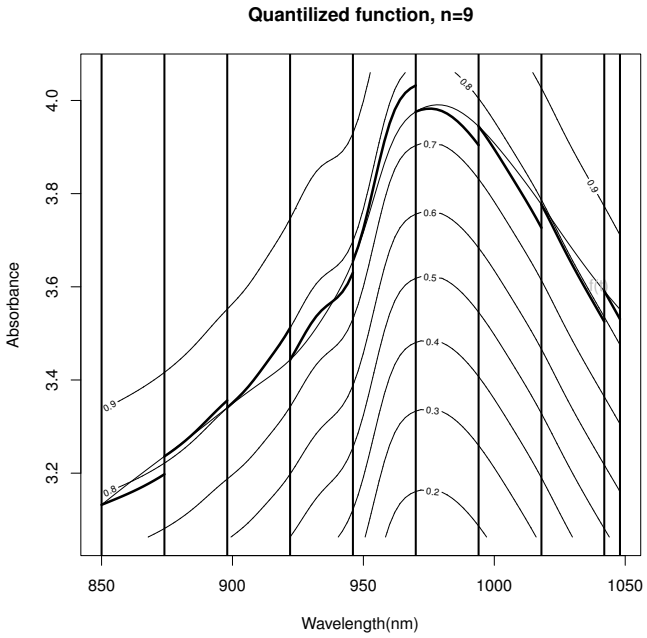


Figure 5.16: Quantilized version of a function of the Tecator dataset with 9 parts.

whose probability is computed with a classical joint distribution $H_{\mathcal{X}_{t_1},\ldots,\mathcal{X}_{t_n}}$ (expression (3.51), p.51), respecting the compatibility conditions for the Daniell-Kolmogorov's extension theorem (p.50):

$$P[\mathcal{A}_{\mathcal{X},\mathcal{D},n}(u)] = H_{\mathcal{X}_{t_1},\ldots,\mathcal{X}_{t_n}}\left(u(t_1),\ldots,u(t_n)\right).$$

But, unfortunately, as exposed at the end of the section 4.8 (p.139), the volumetric behavior of classical joint distributions implies: an unwanted instability in the estimation of the *fcdf*, the impossibility to avoid any approximation by taking the limit $n \to \infty$, the difficulty to share the same cylinder and the precision issue in high dimensions.

To avoid these problems, we propose a solution built on quantilized functions which preserves the probability of any functional quantile.

Let us suppose, that $\mathcal{D}$ is divided in n ($n \in \mathbb{N}_0$) intervals of equal lengths, using the $n+1$ equidistant points $\{t_1,\ldots,t_{n+1}\}$, such $t_1 = \min(\mathcal{D})$ and $t_{n+1} = \max(\mathcal{D})$. Then for any $u \in \mathcal{I}^{\mathcal{D}}$, this partition of $\mathcal{D}$ implicitly defines the following "quantilized" approximation of the function u as defined in the preceding section:

$$u_n(t) = \sum_{i=1}^{n} V_{\mathcal{X},\mathcal{D},p_i}(t)\mathbf{1}_{[t_i,t_{i+1}[}(t) \tag{5.15}$$

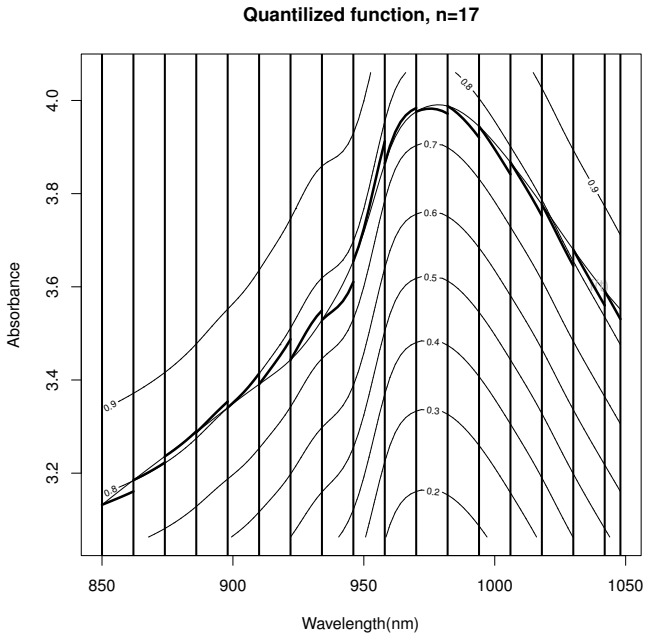


Figure 5.17: Quantilized version of a function of the Tecator dataset with 17 parts.

where for all $i \in \{1, \ldots, n\}$, $V_{\mathcal{X},\mathcal{D},p_i}$ is a functional of order p_i, with $p_i = G_{\mathcal{X},\mathcal{D}}[t_i; u]$.

Thus, for every n we have

$$\begin{array}{ccccccc} \mathcal{F}_{\mathcal{X},[t_1,t_2[}\ (u_n) & = & \mathcal{F}_{\mathcal{X},[t_1,t_2[}\ (V_{\mathcal{X},\mathcal{D},p_1}) & = & \mathbf{G}_{\mathcal{X},\mathcal{D}}[t_1; u] & = & p_1 \\ & \vdots & & \vdots & & \vdots & \\ \mathcal{F}_{\mathcal{X},[t_n,t_{n+1}[}(u_n) & = & \mathcal{F}_{\mathcal{X},[t_n,t_{n+1}[}(V_{\mathcal{X},\mathcal{D},p_n}) & = & \mathbf{G}_{\mathcal{X},\mathcal{D}}[t_n; u] & = & p_n. \end{array}$$

Remember that, in the classical case, using copulas, the finite dimensional distribution is given by expression (4.122)

$$\begin{array}{rcl} P[\mathcal{A}_{\mathcal{X},\mathcal{D},n}(u)] & = & H_{\mathcal{X}_{t_1},\ldots,\mathcal{X}_{t_n}}\left(u(t_1), \ldots, u(t_n)\right) \\ & = & C_{\mathcal{X},n}\left(G_{\mathcal{X},\mathcal{D}}[t_1; u], \ldots, G_{\mathcal{X},\mathcal{D}}[t_n; u]\right) \\ & = & C_{\mathcal{X},n}\left(p_1, \ldots, p_n\right) \end{array}$$

which can be rewritten as follows:

$$P[\mathcal{A}_{\mathcal{X},\mathcal{D},n}(u)] = C_{\mathcal{X},n}\left(\mathcal{F}_{\mathcal{X},[t_1,t_2[}\left(u_n\right), \ldots, \mathcal{F}_{\mathcal{X},[t_n,t_{n+1}[}\left(u_n\right)\right).$$

Thus, the copula $C_{\mathcal{X},n}$ can be seen as an aggregation operation on the values $\mathcal{F}_{\mathcal{X},[t_i,t_{i+1}[}\left(u_n\right)$, but with the weaknesses explained in section 4.8. One of

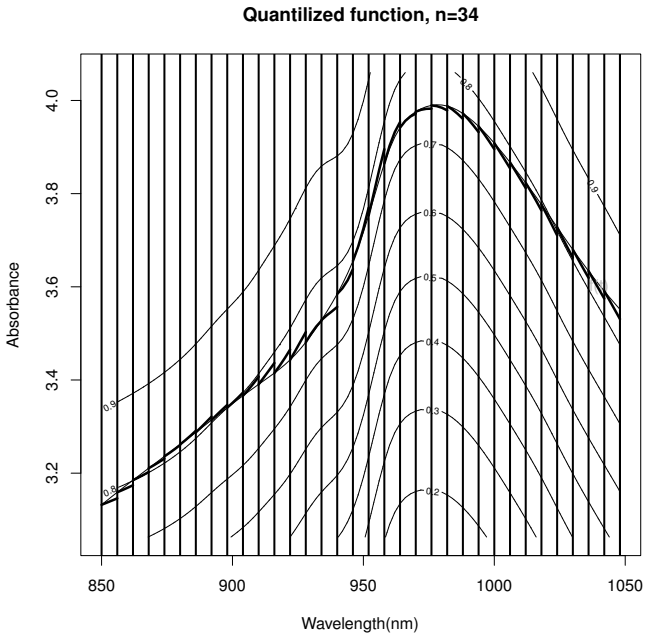


Figure 5.18: Quantilized version of a function of the Tecator dataset with 35 parts.

these weaknesses is the instability of the estimation for functional quantiles: this estimation depends of n, decreases when n increases, and finaly tends toward 0 when $n \to \infty$. Then, the framework of joints probability distributions designed for finite vectors of real random variables is not adapted to define a probability distribution for *frvs*.

Such situations are classical in mathematics: for a given framework there is no solution for a specific and important problem. And, a classical response to this type of situations, is to define a new framework more adapted to solve the problem. In the present case we propose to set as an necessary property the fact that the probability of a functional quantile must be the same at "any granularity of observation", i.e. for any n.

Let $V_{\mathcal{X},\mathcal{D},p}$ a functional quantile of order p ($p \in]0,1[$) of the *frv* $\mathcal{X}$ belonging to $\mathcal{I}^{\mathcal{D}}$, then by definition 4.2.3 we have

$$\mathbf{G}_{\mathcal{X},\mathcal{D}}[t; V_{\mathcal{X},\mathcal{D},p}] = p, \ \forall t \in \mathcal{D}.$$

We want that for any $S \subseteq \mathcal{D}$ we have

$$\mathcal{F}_{X,\mathcal{S}}\left(V_{\mathcal{X},S,p}\right) = p.$$

We will say that the *fcdf* $\mathcal{F}_{\mathcal{X},\mathcal{D}}$ is *conservative for the probability of the functional quantiles*. And like in this case we have obviously the following

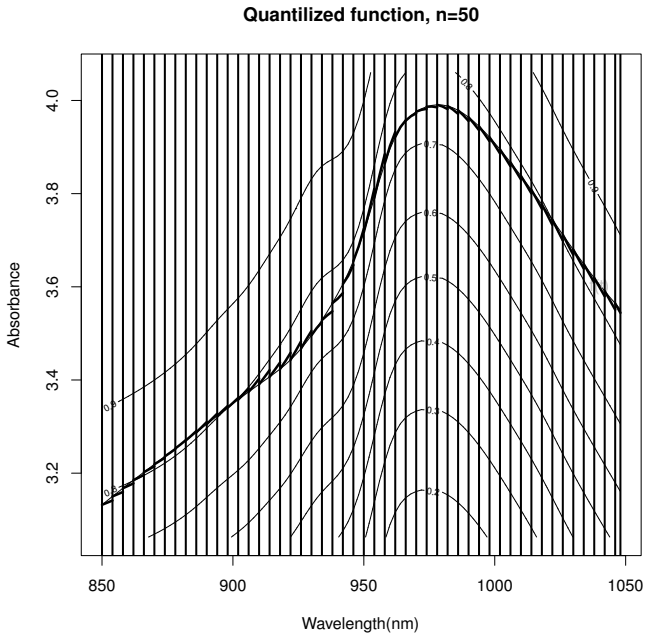


Figure 5.19: Quantilized version of a function of the Tecator dataset with 50 parts.

limit

$$\lim_{t_{i+1}\to t_i} \mathcal{F}_{\mathcal{X},[t_i,t_{i+1}]}\left(V_{\mathcal{X},\mathcal{D},p_i}\right) = G_{\mathcal{X},\mathcal{D}}[t_i; V_{\mathcal{X},\mathcal{D},p_i}] = p_i \tag{5.16}$$

we will say that such an *fcdf* $\mathcal{F}_{\mathcal{X},\mathcal{D}}$ is a *fractal distribution for functional quantiles*, in the sense that there is a *self-similarity of the probability distribution for functional quantiles*.

Thus we search for an aggregation operator

$$\begin{aligned} \mathcal{F}_{\mathcal{X},\mathcal{D}}\left(u_n\right) &= \mathcal{M}^{(n)}\left(\mathcal{F}_{\mathcal{X},[t_1,t_2[}\left(u_n\right),\ldots,\mathcal{F}_{\mathcal{X},[t_n,t_{n+1}[}\left(u_n\right)\right) && (5.17)\\ &= \mathcal{M}^{(n)}\left(G_{\mathcal{X},\mathcal{D}}[t_1;u],\ldots,G_{\mathcal{X},\mathcal{D}}[t_n;u]\right) && (5.18)\\ &= \mathcal{M}^{(n)}\left(p_1,\ldots,p_n\right) && (5.19) \end{aligned}$$

such that $\mathcal{M}^{(n)}\left(p_1,\ldots,p_n\right)$ is still a probability.

With such an operator we will compute the following limit

$$\mathcal{F}_{X,\mathcal{D}}\left(u\right) = \lim_{i\to\infty} \mathcal{F}_{X,\mathcal{D}}\left(u_i\right), \tag{5.20}$$

where $\{u_i\}$ is a sequence of quantilized functions tending toward u.

Definition 5.3.1. *Let* $\mathcal{X}$ *an* frv *belonging in* $\mathcal{I}^{\mathcal{D}}$*, and* $G_{\mathcal{X},\mathcal{D}}$ *its surface of margins. The* fcdf $\mathcal{F}_{\mathcal{X},\mathcal{D}}$ *of* $\mathcal{X}$ *will be called a* fractal and quantile based distribution *iff:*

Property 1. *For any, $p \in [0,1]$, any $S \subseteq \mathcal{D}$, and any functional quantile $V_{\mathcal{X},S,p}$*

$$\mathcal{F}_{X,\mathcal{S}}\left(V_{\mathcal{X},S,p}\right) = p. \tag{5.21}$$

Property 2. *For any u_n quantilized function belonging in $\mathcal{I}^{\mathcal{D}}$ defined by a partition $\mathcal{A} = A_1, \ldots, A_n$ of $\mathcal{D}$ and a set of probabilities $\mathbf{p} = (p_1, \ldots, p_n)$:*

$$u_n(t) = \sum_{k=1}^{n} V_{\mathcal{X},\mathcal{D},p_k}(t)\mathbf{1}_{A_k}(t) \ \forall t \in \mathcal{D}$$

then there exists a function $\mathcal{M}^{(n)}$, called aggregation operator, such that

$$\begin{aligned} \mathcal{F}_{\mathcal{X},\mathcal{D}}\left(u_n\right) &= \mathcal{M}^{(n)}\left(\mathcal{F}_{\mathcal{X},A_1}\left(u_n\right), \ldots, \mathcal{F}_{\mathcal{X},A_n}\left(u_n\right)\right) && (5.22)\\ &= \mathcal{M}^{(n)}\left(p_1, \ldots, p_n\right). && (5.23) \end{aligned}$$

Property 3. *For any u belonging to $\mathcal{I}^{\mathcal{D}}$, if $\{u_n\}$ denote a sequence of quantilized functions, such $\lim_{n\to\infty} u_n = u$, then*

$$\mathcal{F}_{X,\mathcal{D}}\left(u\right) = \lim_{n\to\infty} \mathcal{F}_{X,\mathcal{D}}\left(u_n\right). \tag{5.24}$$

Original contribution(s) 6. *In this section we define a new framework for* fcdfs*. The design of this framework is ruled by a principle, and three properties. The principle is that, an* frv *or a function must be always considered as whole, even if we only know it in a discretized version. The three properties set that, firstly, the* fcdf *of a functional quantile of order p ($p \in [0,1]$) can not depends of its evaluation's mode. If the* fcdf *is estimated on a discretized version of the functional quantile, then, the number of discretization points can not significantly influence the result. Secondly, the* fcdf *of a quantilized function must be based on an aggregation of the probability values of the quantiles which form the function. And thirdly, the* fcdf *calculated in a function u must be equal to the limit of a sequence of* fcdfs *of quantilized functions converging toward u.*

The fcdfs *which fulfill these three properties are called* fractal and quantile based distributions*, because two properties deals with functional quantiles, and because these properties must be fulfilled on any subset of the domain, and for any number of discretization points (the granularity).*

5.4 Aggregation operations

According to [Fodor and Roubens, 1994] an aggregation operation can be defined as follow.

Definition 5.4.1. *An aggregation operator $\mathcal{M}$ for the real interval $[a,b]$ is defined as follow:*

$$\mathcal{M}:\Lambda \ =\bigcup_{n=1}^{\infty}[a,b]^n \to [a,b],(x_1,\ldots,x_n)\mapsto \mathcal{M}^{(n)}(x_1,\ldots,x_n). \tag{5.25}$$

Let us give here the main properties desired for such an operator. First of all, the condition (5.21) must hold. For a functional quantile $V_{\mathcal{X},\mathcal{D},p}$, if we compute the *fcdf* over the quantilized approximation of $V_{\mathcal{X},\mathcal{D},p}$, we must have:

$$\begin{aligned}\mathcal{F}_{\mathcal{X},\mathcal{D}}\left(V_{\mathcal{X},\mathcal{D},p}\right) &= \mathcal{M}\left(\mathcal{F}_{\mathcal{X},[t_1,t_2[}\left(V_{\mathcal{X},\mathcal{D},p}\right),\ldots,\mathcal{F}_{\mathcal{X},[t_n,t_{n+1}[}\left(V_{\mathcal{X},\mathcal{D},p}\right)\right)\\ &= \mathcal{M}\left(p,\ldots,p\right) &(5.26)\\ &= p &(5.27)\end{aligned}$$

so the aggregation operator M must be *idempotent*.

Definition 5.4.2. *An aggregation operator $\mathcal{M}$ for the real interval $[a,b]$ is said* idempotent *if for all $n\in\mathbb{N}_0,\mathcal{M}$ satisfies*

$$\mathcal{M}^{(n)}(x,\ldots,x)=x,\ \forall x\in[a,b].$$

Now, let us take a look at the shared properties by aggregation operations and joint distributions.

Let u_i and v_i be two quantilized functions of $\mathcal{I}^{\mathcal{D}}$ that only differ on $[t_k,t_{k+1}[$, with $i\in\mathbb{N}_0$, and suppose that, for a given $k\in\{1,\ldots,n_i\}$, we have:

$$\begin{aligned}u_i(t_j) &= v_i(t_j) \quad if\ j\neq k,\\ &\geq v_i(t_j) \quad if\ j=k.\end{aligned}$$

Of course, using proposition 4.2.1 (p.84) we have

$$\begin{aligned}\mathbf{G}_{\mathcal{X},\mathcal{D}}[u_i;t_j] &= \mathbf{G}_{\mathcal{X},\mathcal{D}}[v_i;t_j] \quad if\ j\neq k,\\ &> \mathbf{G}_{\mathcal{X},\mathcal{D}}[v_i;t_j] \quad if\ j=k,\end{aligned}$$

then

$$\begin{aligned}\mathcal{F}_{\mathcal{X},[t_j,t_{j+1}[}(u_i) &= \mathcal{F}_{\mathcal{X},[t_j,t_{j+1}[}(v_i) \quad if\ j\neq k,\\ &> \mathcal{F}_{\mathcal{X},[t_j,t_{j+1}[}(v_i) \quad if\ j=k.\end{aligned}$$

If p_j denotes $\mathcal{F}_{\mathcal{X},[t_j,t_{j+1}[}(v_i)$ and p' denotes $\mathcal{F}_{\mathcal{X},[t_k,t_{k+1}[}(u_i)$ then, as for joint distributions we want to have

$$\mathcal{M}\left(p_{i,1},\ldots,p',\ldots,p_n\right) > \mathcal{M}\left(p_1,\ldots,p_k,\ldots,p_n\right), \tag{5.28}$$

that is to say that $\mathcal{M}$ must be *strictly increasing in each argument.*

Definition 5.4.3. *An aggregation operator $\mathcal{M}$ for the real interval $[a,b]$ is said*

- continuous *if for all $n \in \mathbb{N}_0, \mathcal{M}$ is a continuous function on $[a,b]^n$;*
- increasing *if for all $n \in \mathbb{N}_0$, $\mathcal{M}^{(n)}$ is increasing in each argument ($i \in \{1,\ldots,n\}$):*

$$x_i < x_i' \Rightarrow \mathcal{M}^{(n)}(x_1,\ldots,x_i,\ldots,x_n) \leq \mathcal{M}^{(n)}(x_1,\ldots,x_i',\ldots,x_n).$$

- strictly increasing *if for all $n \in \mathbb{N}_0$, $\mathcal{M}^{(n)}$ is strictly increasing in each argument ($i \in \{1,\ldots,n\}$):*

$$x_i < x_i' \Rightarrow \mathcal{M}^{(n)}(x_1,\ldots,x_i,\ldots,x_n) < \mathcal{M}^{(n)}(x_1,\ldots,x_i',\ldots,x_n).$$

Like $\mathcal{F}_{\mathcal{X},\mathcal{D}}$ must be a joint distribution of u_1 over $A_{i,1}$, and u_2 over $A_{i,2}$, ..., and u_{n_i} over A_{i,n_i}, we want also, as the first compatibility condition in the Kolmogorov's extension theorem (see p.50), the independence of the result with the "order" of the evaluation. Let a given partition $A_1,\ldots,A_n$:

$$\mathcal{M}\left(\mathcal{F}_{X,A_1}(u),\ldots,\mathcal{F}_{X,A_n}(u)\right) = \mathcal{M}\left(\mathcal{F}_{X,A_{\sigma(1)}}(u),\ldots,\mathcal{F}_{X,A_{\sigma(n)}}(u)\right) \tag{5.29}$$

where σ is a permutation of $\{1,\ldots,n\}$.

Definition 5.4.4. *An aggregation operator $\mathcal{M}$ for the real interval $[a,b]$ is said* symmetric *if for all $n \in \mathbb{N}_0, \mathcal{M}^{(n)}$:*

$$\mathcal{M}^{(n)}(x_1,\ldots,x_n) = \mathcal{M}^{(n)}(x_{\sigma(1)},\ldots,x_{\sigma(n)})$$

where σ is a permutation of $\{1,\ldots,n\}$ and $(x_1,\ldots,x_n) \in [a,b]^n$.

Now, let us suppose that $\mathcal{D} = [a,b]$, (with $a,b \in \mathbb{R}$), and suppose that, $\{A_1,\ldots,A_k\}$ is a partition of $\mathcal{D}$ in intervals of equal length. Let us denote $\mathcal{M}_k$ the "joint probability" of a quantilized function u_i over $\mathcal{D}$:

$$\mathcal{M}_k = \mathcal{M}^{(k)}\left(\mathcal{F}_{X,A_1}(u_i),\ldots,\mathcal{F}_{X,A_k}(u_i)\right).$$

If $\mathcal{X}$ is also defined on $\mathcal{D}' = [b,c]$, then it will be helpful, when we want to compute the probability over $[a,c] = \mathcal{D} \cup \mathcal{D}'$, to use the "information"

collected on $\mathcal{D}$, i.e. if $\{A_{k+1}, \ldots, A_n\}$ is a partition of $\mathcal{D}'$ in intervals of the same length that $\{A_1, \ldots, A_k\}$, then we want to have

$$\begin{aligned} \mathcal{M}_n &= \mathcal{M}^n\left(\mathcal{F}_{X,A_1}(u), \ldots, \mathcal{F}_{X,A_k}(u), \mathcal{F}_{X,A_{k+1}}(u), \ldots, \mathcal{F}_{X,A_n}(u)\right) \\ &= \mathcal{M}^n\left(\mathcal{M}_k, \ldots, \mathcal{M}_k, \mathcal{F}_{X,A_{k+1}}(u), \ldots, \mathcal{F}_{X,A_n}(u)\right). \end{aligned}$$

For the sake of illustration, this property could be interesting if, $\mathcal{X}$ is the result of a stream of information: if, at the time $t = b$ we can compute a probability about $\mathcal{X}$ over $[a, b]$, then at time $t = b + \delta = c$, we can compute the same probability about $\mathcal{X}$ over $[a, c]$ without having to recompute everything.

Definition 5.4.5. *An aggregation operator $\mathcal{M}$ for the real interval $[a, b]$ is said* decomposable *if for all $n \in \mathbb{N}_0$ and all $k \in \{1, \ldots, n\}$, the following equality holds:*

$$\mathcal{M}^{(n)}(x_1, \ldots, x_k, x_{k+1}, \ldots, x_n) = \mathcal{M}^{(n)}(\mathcal{M}_k, \ldots, \mathcal{M}_k, x_{k+1}, \ldots, x_n)$$

where $\mathcal{M}_k = \mathcal{M}^{(k)}(x_1, \ldots, x_k)$.

Thus we search for an aggregation function which is continuous, symmetric, strictly increasing, idempotent and decomposable. *Quasi-arithmetic means* have all these properties.

Original contribution(s) 7. *In this section we select the desired properties for aggregation operators to use in fractal and quantile based distributions. The properties chosen are the same than those of the classical joint distributions, and we chose also an additional property interesting for streaming data.*

5.5 Quasi-Arithmetic Means

[Kolmogorov, 1930] established the following result.

Theorem 5.5.1. *An aggregation operator $\mathcal{M}$, defined on Λ, is continuous, symmetric, strictly increasing, idempotent and decomposable if and only if for all $n \in \mathbb{N}_0$,*

$$\mathcal{M}_\phi^{(n)}(x_1, \ldots, x_n) = \phi^{-1}\left[\frac{1}{n}\sum_{i=1}^{n}\phi(x_i)\right] \tag{5.30}$$

where ϕ is any continuous strictly monotonic function on $[a, b]$.

Definition 5.5.1. *Let $[a, b]$ be a closed real interval, and $n \in \mathbb{N}_0$. A* quasi-arithmetic mean *is a function $\mathcal{M}_\phi^{(n)} : [a, b]^n \to [a, b]$ which satisfied* (5.30) *where ϕ is any continuous strictly monotonic function on $[a, b]$.*

$\phi(x)$	$\mathcal{M}_{\phi}^{(n)}(x_1, \ldots, x_n)$	Name
x	$\frac{1}{n}\sum_{i=1}^{n} x_i$	Arithmetic
x^2	$\sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2}$	Quadratic
$\log x$	$\left(\prod_{i=1}^{n} x_i\right)^{\frac{1}{n}}$	Geometric
x^{-1}	$\frac{1}{\frac{1}{n}\sum_{i=1}^{n} \frac{1}{x_i}}$	Harmonic
x^{α}, $\alpha \neq 0$ and finite	$\left(\frac{1}{n}\sum_{i=1}^{n} x_i^{\alpha}\right)^{\frac{1}{\alpha}}$	Root-power

Table 5.1: Special cases for *quasi-arithmetic means*.

Quasi-arithmetic means are an extension of the most common means, as it is shown in the table 5.1 [Fodor and Roubens, 1994]: if $\phi(x)$ is respectively x, x^2, $\log(x)$ and x^{-1}, expression (5.30) gives respectively the arithmetic, quadratic, geometric and harmonic means. Note that the root-power mean, is the more "general" in this table, it is possible to retrieve the classical means using only its expression: when $\alpha = 1$ it gives the arithmetic mean, when $\alpha = -1$, the harmonic mean, when $\alpha = 2$, the quadratic mean and when $\alpha \to 0$, the limit is the geometric mean.

Quasi-arithmetic means lead also to more "unusual" means as the radical mean and the trigonometric means.

Example 5.5.1. *Let $a \in \mathbb{R}$ such $a > 0$ and $a \neq 1$, then, setting $\phi(x) = a^{1/x}$ we obtain the radical mean [Bullen et al., 1988]:*

$$\mathcal{M}_{\phi}^{(n)}(x_1, \ldots, x_n) = \left\{ \log_a \left(\frac{1}{n} \sum_{i=1}^{n} a^{1/x_i} \right) \right\}^{-1}. \tag{5.31}$$

Example 5.5.2. *If $0 \leq x_i \leq \pi/2$, for $i = 1, \ldots, n$, then we obtain the trigonometric means, setting $\phi(x) = \sin(x)$, $\phi(x) = \cos(x)$, $\phi(x) = \tan(x)$ or $\phi(x) = \cot(x)$ ([Bonferroni, 1927], [Pratelli, 1940], and [Jecklin, 1953]):*

$$S^{(n)}(x_1, \ldots, x_n) = \arcsin \left(\frac{1}{n} \sum_{i=1}^{n} \sin(x_i) \right), \tag{5.32}$$

$$C^{(n)}(x_1, \ldots, x_n) = \arccos \left(\frac{1}{n} \sum_{i=1}^{n} \cos(x_i) \right), \tag{5.33}$$

$$T^{(n)}(x_1, \ldots, x_n) = \arctan \left(\frac{1}{n} \sum_{i=1}^{n} \tan(x_i) \right), \tag{5.34}$$

$$CT^{(n)}(x_1, \ldots, x_n) = \operatorname{arccot} \left(\frac{1}{n} \sum_{i=1}^{n} \cot(x_i) \right). \tag{5.35}$$

Example 5.5.3. *[Beliakov et al., 2007] gives trigonometric means defined in $\mathbb{R}$ setting $\phi(x) = \sin(\frac{\pi}{2}x)$, $\phi(x) = \cos(\frac{\pi}{2}x)$, $\phi(x) = \tan(\frac{\pi}{2}x)$ or $\phi(x) = \cot(\frac{\pi}{2}x)$:*

$$S^{(n)}(x_1, \ldots, x_n) = \frac{2}{\pi} \arcsin \left(\frac{1}{n} \sum_{i=1}^{n} \sin(\frac{\pi}{2} x_i) \right), \tag{5.36}$$

$$C^{(n)}(x_1, \ldots, x_n) = \frac{2}{\pi} \arccos \left(\frac{1}{n} \sum_{i=1}^{n} \cos(\frac{\pi}{2} x_i) \right), \tag{5.37}$$

$$T^{(n)}(x_1, \ldots, x_n) = \frac{2}{\pi} \arctan \left(\frac{1}{n} \sum_{i=1}^{n} \tan(\frac{\pi}{2} x_i) \right), \tag{5.38}$$

$$CT^{(n)}(x_1,\ldots,x_n) = \frac{2}{\pi}\text{arccot}\left(\frac{1}{n}\sum_{i=1}^{n}\cot(\frac{\pi}{2}x_i)\right). \tag{5.39}$$

In the following we will give only the properties of the means which are interesting for our purpose, and more precisely on the relations between the different means [Hardy et al., 1934] [Bullen et al., 1988].

First of all, two trivial, but important properties about the value of a *quasi-arithmetic mean* [Bullen et al., 1988].

Proposition 5.5.2. *Let $\mathcal{M}_\phi$ be a* quasi-arithmetic mean *defined on a closed real interval $[a,b]$, and $n \in \mathbb{N}$. If for all $1 \leq i \leq n$, we have $x_i \leq y_i$, then*

$$\mathcal{M}_\phi^{(n)}(x_1,\ldots,x_n) \leq \mathcal{M}_\phi^{(n)}(y_1,\ldots,y_n)$$

with equality iff $x_i = y_i$, $\forall i$.

Proposition 5.5.3. *Let $\mathcal{M}_\phi$ be a* quasi-arithmetic mean *defined on a closed real interval $[a,b]$, and $n \in \mathbb{N}$, then*

$$\min(x_1,\ldots,x_n) \leq \mathcal{M}_\phi^{(n)}(x_1,\ldots,x_n) \leq \max(x_1,\ldots,x_n).$$

Two *quasi-arithmetic means* are equivalent if their generators are linked by a linear transformation [Hardy et al., 1934].

Proposition 5.5.4. *Let $\mathcal{M}_\phi$, $\mathcal{M}_\varphi$ be two* quasi-arithmetic means*, then*

$$\mathcal{M}_\phi(x_1,\ldots,x_n) = \mathcal{M}_\varphi(x_1,\ldots,x_n)\ \forall(x_1,\ldots,x_n) \iff \phi = \alpha\varphi + \beta$$

where α and β are constants, and $\alpha \neq 0$.

If the arithmetic mean of a set $\{x_1,\ldots,x_n\}$ is m, then, the arithmetic mean of the set $\{\alpha x_1,\ldots,\alpha x_n\}$ is αm, i.e. the arithmetic mean is homogeneous. The following proposition shows that, only the root-power mean family (see table 5.1) has this property [Hardy et al., 1934].

Proposition 5.5.5. *Suppose that $\phi(x)$ is continuous in the open interval $]0,\infty[$, and that*

$$\mathcal{M}_\phi(\alpha x_1,\ldots,\alpha x_n) = \alpha\mathcal{M}_\phi(x_1,\ldots,x_n) \tag{5.40}$$

for all set of positives $\{x_1,\ldots,x_n\}$ and all $\alpha > 0$, then $\phi(x) = x^\alpha$, in other words, the root power means are the only homogeneous means.

Now, given two functions ϕ and ψ, an interesting question is: is it possible to compare the *quasi-arithmetic means* $\mathcal{M}_\phi$ and $\mathcal{M}_\psi$ in all cases [Hardy et al., 1934] and [Bullen et al., 1988]?

Definition 5.5.2. *The two quasi-arithmetic means generated by the two continuous and strictly monotonic functions ϕ and ψ defined on the real interval $[a, b]$, are* comparable *if one of the two following inequalities is always true:*

$$\mathcal{M}_{\phi}^{(n)}(x_1, \ldots, x_n) \leq \mathcal{M}_{\psi}^{(n)}(x_1, \ldots, x_n) \tag{5.41}$$

$$\mathcal{M}_{\phi}^{(n)}(x_1, \ldots, x_n) \geq \mathcal{M}_{\psi}^{(n)}(x_1, \ldots, x_n) \tag{5.42}$$

for all sets $\{x_1, \ldots, x_n\}$.

Proposition 5.5.6. *If ψ and ϕ are continuous and strictly monotonic, and ϕ is increasing, then a necessary and sufficient condition that*

$$\mathcal{M}_{\psi}^{(n)}(x_1, \ldots, x_n) \leq \mathcal{M}_{\phi}^{(n)}(x_1, \ldots, x_n)$$

for all sets $\{x_1, \ldots, x_n\}$ is that $\phi \circ \psi^{-1}$ should be convex.

Corollary 5.5.7. *If ψ is continuous, convex and strictly monotonic, and if $\mathbb{1}$ denotes the identity function $\mathbb{1}(x) = x$, then*

$$\mathcal{M}_{\psi}^{(n)}(x_1, \ldots, x_n) \leq \mathcal{M}_{\mathbb{1}}^{(n)}(x_1, \ldots, x_n) \tag{5.43}$$

for all sets $\{x_1, \ldots, x_n\}$. In other words the quasi-arithmetic mean $\mathcal{M}_{\psi}^{(n)}$ *is less or equal to the arithmetic mean, when ψ is convex.*

Proof. By proposition 4.4.2 (p.106) ψ^{-1} is convex, and then $\mathbb{1} \circ \psi^{-1}$ is convex. □

Intuitively it is easy to extend quasi-arithmetic means defined in the discrete case (see expression (5.30)) to the continuous case. Suppose there exists a function f defined on a real interval $\mathcal{D}$ such $\forall i \in \{1, \ldots, n\}\ f(t_i) = x_i$ then we can compute

$$\mathcal{M}_{\phi}^{(n)}(f(t_1), \ldots, f(t_n)) = \phi^{-1}\left[\frac{1}{n}\sum_{i=1}^{n}\phi(f(t_i))\right]. \tag{5.44}$$

If the t_i are equidistant, i.e. for all $i \in \{1, \ldots, n-1\}$:

$$t_{i+1}^{n} - t_i^{n} = \frac{|\mathcal{D}|}{n} = \Delta_t$$

then the following expression

$$\mathcal{M}_{\phi}^{(\mathcal{D})}(f) = \lim_{n\to\infty} \phi^{-1}\left[\frac{1}{|\mathcal{D}|}\sum_{i=1}^{n}\phi(f(t_i))\Delta_t\right]. \tag{5.45}$$

leads us directly to a Riemann integral which could define the continuous case of quasi-arithmetic means. However, we can directly define this continuous case using the Lebesgue integral [Hardy et al., 1934].

Definition 5.5.3. *Let $\mathcal{D}$ and $\mathcal{I}$ be two real closed intervals, f a real function defined on $\mathcal{D}$ with values in $\mathcal{I}$ (i.e. $f \in \mathcal{I}^{\mathcal{D}}$), ϕ a continuous strictly monotonic function defined on $\mathcal{I}$. Then, the* continuous quasi-arithmetic mean *of f is defined by*

$$\mathcal{M}_{\phi}^{(\mathcal{D})}(f) = \phi^{-1}\left[\frac{1}{|\mathcal{D}|}\int_{\mathcal{D}} \phi\left(f(t)\right) dt\right]. \tag{5.46}$$

Propositions 5.5.2, 5.5.3, 5.5.6 and corollary 5.5.7 are easily extended in the continuous case [Hardy et al., 1934].

Proposition 5.5.8. *Let $\mathcal{M}_{\phi}^{(\mathcal{D})}$ be a* quasi-arithmetic mean *defined on a closed real interval $[a, b]$. If for all $a \leq x \leq b$, we have $f(x) \leq g(x)$, then*

$$\mathcal{M}_{\phi}^{(\mathcal{D})}(f) \leq \mathcal{M}_{\phi}^{(\mathcal{D})}(g)$$

with equality iff $f = g$.

Proposition 5.5.9. *Let $\mathcal{M}_{\phi}^{(\mathcal{D})}$ be a* quasi-arithmetic mean *defined on a closed real interval $[a, b]$, then*

$$\min_{t\in\mathcal{D}}(f) \leq \mathcal{M}_{\phi}^{(\mathcal{D})}(f) \leq \max_{t\in\mathcal{D}}(f).$$

Proposition 5.5.10. *If ψ and ϕ are continuous and strictly monotonic, and ϕ is increasing, then a necessary and sufficient condition that*

$$\mathcal{M}_{\psi}^{(\mathcal{D})}(f) \leq \mathcal{M}_{\phi}^{(\mathcal{D})}(f)$$

for any function $f \in \mathcal{I}^{\mathcal{D}}$ is that $\phi \circ \psi^{-1}$ should be convex.

Corollary 5.5.11. *If ψ is continuous, convex and strictly monotonic, and if $\mathbb{1}$ denotes the identity function $\mathbb{1}(x) = x$, then*

$$\mathcal{M}_{\psi}^{(\mathcal{D})}(f) \leq \mathcal{M}_{\mathbb{1}}^{(\mathcal{D})}(f) \tag{5.47}$$

for any function $f \in \mathcal{I}^{\mathcal{D}}$. In other words the quasi-arithmetic mean $\mathcal{M}_{\psi}$ *is less or equal to the arithmetic mean, when ψ is convex.*

The theorem 5.5.1 leads directly to the definition 5.5.1 of *quasi-arithmetic means*, which lead naturally to definition 5.5.3 for the continuous case. These two definitions can be rewritten in a more general form. Thus, expression (5.30) and (5.46) can be rewritten, respectively

$$\mathcal{M}_{\phi,p}^{(n)}(x_1, \ldots, x_n) = \phi^{-1}\left[\sum_{i=1}^{n} p_i \phi(x_i)\right] \tag{5.48}$$

and

$$\mathcal{M}_{\phi,p}^{(\mathcal{D})}(f) = \phi^{-1}\left[\int_{\mathcal{D}} p(t)\phi\left(f(t)\right)dt\right]. \tag{5.49}$$

where $p_i = 1/n$ for all i, and $p(t) = 1/|\mathcal{D}|$ for all $t \in \mathcal{D}$. Of course, these two conditions imply that, respectively

$$\sum_{i=1}^{n} p_i = 1 \tag{5.50}$$

$$\int_{\mathcal{D}} p(t)dt = 1. \tag{5.51}$$

If we use, respectively, expressions (5.50) and (5.51) as conditions in definitions of *quasi-arithmetic means*, then we obtain the notion of *quasi-arithmetic weighted means*.

Definition 5.5.4. *Let $[a, b]$ be a closed real interval, and $n \in \mathbb{N}_0$. A* quasi-arithmetic weighted mean *is a function $\mathcal{M}_{\phi,p}^{(n)} : [a, b]^n \to [a, b]$ which satisfies* (5.48) *where ϕ is any continuous strictly monotonic function on $[a, b]$, and such the set $\{p_1, \ldots, p_n\}$ satisfying the expression* (5.50).

Definition 5.5.5. *Let $\mathcal{D}$ and $\mathcal{I}$ be two real closed intervals, f a real function defined on $\mathcal{D}$ with values in $\mathcal{I}$ (i.e. $f \in \mathcal{I}^{\mathcal{D}}$), ϕ a continuous strictly monotonic function defined on $\mathcal{I}$. Then, the* continuous quasi-arithmetic weighted mean of *f, denoted $\mathcal{M}_{\phi,p}^{(\mathcal{D})}(f)$, is defined by the expression* (5.49), *where $p(t)$ is the function defined on $\mathcal{D}$ such as the expression* (5.51) *is satisfied.*

It seems natural, for the values p_i in expression (5.50), and the function $p(t)$ in expression (5.51), to think, respectively, about discrete and continuous probabilities, even if it is not necessary.

Propositions 5.5.2, 5.5.3, 5.5.6 and 5.5.7 for the discrete case, and propositions 5.5.8, 5.5.9, 5.5.10 and 5.5.11 for the continuous case are easily extended [Hardy et al., 1934].

Proposition 5.5.12. *Let $\mathcal{M}_{\phi,p}$ be a* quasi-arithmetic weighted mean *defined on a closed real interval $[a, b]$, and $n \in \mathbb{N}$. If for all $1 \leq i \leq n$, we have $x_i \leq y_i$, then*

$$\mathcal{M}_{\phi,p}^{(n)}(x_1, \ldots, x_n) \leq \mathcal{M}_{\phi,p}^{(n)}(y_1, \ldots, y_n)$$

with equality iff $x_i = y_i$, $\forall i$.

Proposition 5.5.13. *Let $\mathcal{M}_{\phi,p}$ be a* quasi-arithmetic mean *defined on a closed real interval $[a, b]$, and $n \in \mathbb{N}$, then*

$$\min(x_1, \ldots, x_n) \leq \mathcal{M}_{\phi,p}^{(n)}(x_1, \ldots, x_n) \leq \max(x_1, \ldots, x_n).$$

Proposition 5.5.14. *If ψ and ϕ are continuous and strictly monotonic, and ϕ is increasing, then a necessary and sufficient condition that*

$$\mathcal{M}^{(n)}_{\psi,p}(x_1,\ldots,x_n) \leq \mathcal{M}^{(n)}_{\phi,p}(x_1,\ldots,x_n)$$

for all sets $\{x_1,\ldots,x_n\}$ is that $\phi\circ\psi^{-1}$ should be convex.

Corollary 5.5.15. *If ψ is continuous, convex and strictly monotonic, and if $\mathbb{1}$ denotes the identity function $\mathbb{1}(x)=x$, then*

$$\mathcal{M}^{(n)}_{\psi,p}(x_1,\ldots,x_n) \leq \mathcal{M}^{(n)}_{\mathbb{1},p}(x_1,\ldots,x_n) \tag{5.52}$$

for all set $\{x_1,\ldots,x_n\}$. In other words the quasi-arithmetic mean $\mathcal{M}^{(n)}_{\psi,p}$ *is less or equal to the arithmetic mean, when ψ is convex.*

Proposition 5.5.16. *Let $\mathcal{M}^{(\mathcal{D})}_{\phi,p}$ be a* continuous quasi-arithmetic mean *defined on a closed real interval $[a,b]$. If for all $a\leq x\leq b$, we have $f(x)\leq g(x)$, then*

$$\mathcal{M}^{(\mathcal{D})}_{\phi,p}(f) \leq \mathcal{M}^{(\mathcal{D})}_{\phi,p}(g)$$

with equality iff $f=g$.

Proposition 5.5.17. *Let $\mathcal{M}^{(\mathcal{D})}_{\phi,p}$ be a* continuous quasi-arithmetic mean *defined on a closed real interval $[a,b]$, then*

$$\min_{t\in\mathcal{D}}(f) \leq \mathcal{M}^{(\mathcal{D})}_{\phi,p}(f) \leq \max_{t\in\mathcal{D}}(f).$$

Proposition 5.5.18. *If ψ and ϕ are continuous and strictly monotonic, and ϕ is increasing, then a necessary and sufficient condition that*

$$\mathcal{M}^{(\mathcal{D})}_{\psi,p}(f) \leq \mathcal{M}^{(\mathcal{D})}_{\phi,p}(f)$$

for any function $f\in\mathcal{I}^{\mathcal{D}}$ is that $\phi\circ\psi^{-1}$ should be convex.

Corollary 5.5.19. *If ψ is continuous, convex and strictly monotonic, and if $\mathbb{1}$ denotes the identity function $\mathbb{1}(x)=x$, then*

$$\mathcal{M}^{(\mathcal{D})}_{\psi,p}(f) \leq \mathcal{M}^{(\mathcal{D})}_{\mathbb{1},p}(f) \tag{5.53}$$

for any function $f\in\mathcal{I}^{\mathcal{D}}$. In other words the quasi-arithmetic mean $\mathcal{M}_{\psi}$ *is less or equal to the arithmetic mean, when ψ is convex.*

5.6 QAMM: Joint Distributions

We have seen in section 5.5 that the **property 1** is fulfilled by the *quasi-arithmetic means*, because they are idempotent aggregation operators (definition 5.4.2, p.173).

And to fulfill **property 2**, it suffices that $\mathcal{M}_{\phi}^{(n)}\left(\mathbf{G}_{\mathcal{X},\mathcal{D}}[t_1;u],\ldots,\mathbf{G}_{\mathcal{X},\mathcal{D}}[t_n;u]\right)$ be a joint probability, then it is still a probability. But not a joint distribution in the classical meaning, i.e. without the Kolmogorov's Extension Theorem's second compatibility condition (see expression 3.48 p.50).

We have seen that *quasi-arithmetic means* share some properties with the classical joint distributions, they are

- continuous and strictly increasing in each argument (definition 5.4.3 p.174),
- symmetric (definition 5.4.4, p.174), which is equivalent to the first compatibility condition of the Kolmogorv Extension theorem, see expression (3.47), p.50.

Moreover the proposition 5.5.3 ensures us that the result of the aggregation is

$$\min_i \mathbf{G}_{\mathcal{X},\mathcal{D}}[t_i;u] \leq \mathcal{M}_{\phi}^{(n)}\left(\mathbf{G}_{\mathcal{X},\mathcal{D}}[t_1;u],\ldots,\mathbf{G}_{\mathcal{X},\mathcal{D}}[t_n;u]\right) \leq \max_i \mathbf{G}_{\mathcal{X},\mathcal{D}}[t_i;u]$$

and then the limit (5.24) will not always be zero.

Thus, the question is: under what condition the following expression is a joint distribution:

$$\phi^{-1}\left[\frac{1}{n}\sum_{i=1}^{n}\phi\left(G_{\mathcal{X},\mathcal{D}}\left[t_i;u\right]\right)\right]?$$

We show below that if we use a generator for Archimedean copulas to compute a *quasi-arithmetic mean*, then we define a cumulative distribution function built from one-dimensional distributions. It's easy to prove the following lemma.

Lemma 5.6.1. *Let $n \in \mathbb{N}_0$, X be a real random variable, and F_X its one dimensional* cdf. *If ϕ is a generator of Archimedean copula, then*

$$F_X^*(y) = \psi\left(\frac{1}{n}\phi\left(F_X(y)\right)\right) \tag{5.54}$$

is also a cdf. *And, if F_X is a strictly increasing* cdf, *then F_X^* is the* cdf *of the random variable Y given by:*

$$Y = \alpha_{\phi,X}(X) \tag{5.55}$$

where

$$\alpha_{\phi,X}(t) = Q_X[\psi(n\phi[F_X(t)])] \tag{5.56}$$

Moreover $\alpha_{\phi,X}(t)$ is a strictly increasing function.

Proof. As F_X is right continuous and ϕ is continuous, it is easy to see that $F_X^*(y)$ is right continuous.

For the increasing property we have

$$\begin{aligned} y_1 < y_2 &\Leftrightarrow F_X(y_1) \leq F_X(y_2) \\ &\Leftrightarrow \frac{1}{n}\phi\left(F_X(y_1)\right) \geq \frac{1}{n}\phi\left(F_X(y_2)\right) \\ &\Leftrightarrow F_X^*(y_1) \leq F_X^*(y_2). \end{aligned}$$

In the same way:

$$\begin{aligned} \lim_{y\to-\infty} F_X(y) = 0 &\Leftrightarrow \lim_{y\to-\infty} \frac{1}{n}\phi\left(F_X(y)\right) = +\infty \\ &\Leftrightarrow \lim_{y\to-\infty} F_X^*(y) = 0 \end{aligned}$$

and

$$\begin{aligned} \lim_{y\to-\infty} F_X(y) = 1 &\Leftrightarrow \lim_{y\to-\infty} \frac{1}{n}\phi\left(F_X(y)\right) = 0 \\ &\Leftrightarrow \lim_{y\to-\infty} F_X^*(y) = 1 \end{aligned}$$

So F^* is a *cdf*.

Now, suppose that F_X^* is the *cdf* of the real random variable $Y = \alpha(X)$, i.e.

$$\begin{aligned} F_X^*(y) &= P[Y \leq y] \\ &= P[\alpha(X) \leq y] \\ &= P[X \leq \alpha^{-1}(y)] \\ &= F_X[\alpha^{-1}(y)] \end{aligned}$$

then, using expression (5.54), we have

$$\psi\left(\frac{1}{n} \cdot \phi\left(F_X(y)\right)\right) = F_X[\alpha^{-1}(y)].$$

thus

$$\alpha^{-1}(y) = Q_X\left[\psi\left(\frac{1}{n} \cdot \phi\left(F_X(y)\right)\right)\right]$$

and finally

$$\alpha(t) = Q_X[\psi(n\phi[F_X(t)])].$$

Now, let us prove that α is strictly increasing:

$$\begin{aligned} t_1 < t_2 &\Leftrightarrow F_X(t_1) < F_X(t_2) \\ &\Leftrightarrow n\phi\left[F_X(t_1)\right] > n\phi\left[F_X(t_2)\right] \\ &\Leftrightarrow \psi\left(n\phi\left[F_X(t_1)\right]\right) < \psi\left(n\phi\left[F_X(t_2)\right]\right) \\ &\Leftrightarrow Q\left[\psi\left(n\phi\left[F_X(t_1)\right]\right)\right] < Q\left[\psi\left(n\phi\left[F_X(t_2)\right]\right)\right] \\ &\Leftrightarrow \alpha(t_1) < \alpha(t_2). \end{aligned}$$

□

In various situations one can apply transformations to the data without destroying the underlying dependence structure. The lemma 5.6.1 shows that there is an increasing transform $\alpha_{\phi,X}$ associated to the transform applied to the margins. Moreover, if the one dimensional distributions $F_{X_1}, \ldots, F_{X_n}$ are strictly increasing, then this lemma also proves that applications α_{ϕ,X_i} are strictly increasing, and so theorem 4.3.7 (p.93) shows that the copula of the real random variables $\{\alpha_{\phi,X_1}(X_1), \ldots, \alpha_{\phi,X_n}(X_n)\}$ is the same copula that $\{X_1, \ldots, X_n\}$, i.e. α_{ϕ,X_i} functions do not change the underlying dependence structure.

Let us show now, that a *quasi-arithmetic mean* using a generator of an Archimedean copula in conjunction with univariate CDFs gives us a joint distribution of the wanted type.

Proposition 5.6.2. *Let $n \in \mathbb{N}_0$, $\{F_i | 1 \leq i \leq n\}$ be a set of one dimensional* cdf*, and ϕ a generator of Archimedean copula, then*

$$H_{QAMM}(y_1, \ldots, y_n) = \psi\left(\frac{1}{n}\sum_{i=1}^{n}\phi\left(F_i(y_i)\right)\right) \tag{5.57}$$

is a multivariate cdf.
Moreover the underlying copula of H_{QAMM} is the Archimedean copula of generator ϕ.

Proof. By the above lemma we know that the functions $F_i^*(x)$ are *cdf*, and as ϕ is an "Archimedean generator" so:

$$\psi\left(\sum_{i=1}^{n}\phi(u_i)\right)$$

is a copula, and then

$$\psi\left(\sum_{i=1}^{n}\phi(F_i^*(y_i))\right) = \psi\left(\frac{1}{n}\sum_{i=1}^{n}\phi\left(F_i(y_i)\right)\right)$$

is a multivariate *cdf*. □

Definition 5.6.1. *Let $n \in \mathbb{N}_0$, $\{F_i | 1 \leq i \leq n\}$ be a set of one dimensional* cdf*, and ϕ a generator of Archimedean copula, then the multivariate* cdf *given by* (5.57) *is called a* Quasi-Arithmetic Mean of Margins (QAMM). *Moreover, if $X_1, \ldots, X_n$ are real random variables, then the associated QAMM will be denoted $H_{QAMM,X_1,\ldots,X_n}$.*

Proposition 5.6.2 shows that the two following expressions are two joint distributions sharing the same structure dependence:

$$\psi\left(\sum_{i=1}^{n}\phi(F_i(y_i))\right) \ \& \ \psi\left(\frac{1}{n}\sum_{i=1}^{n}\phi(F_i(y_i))\right).$$

But, the first one respects the second compatibility condition for the Daniell-Kolmogorov theorem, while the second one respects the conservation of functional quantiles[1].

Then we have found an aggregation operator which also gives a joint probability. And this operator fulfill **property 1** and **property 2** (p.172): if u_n is a quantilized function following expression (5.15), then if $U_{\mathcal{X},\mathcal{D},p}$ is a functional quantile of order p, then, for any $i \in \{1, \ldots, n\}$ we have

$$U_{\mathcal{X},\mathcal{D},p}(s) = p \ \forall s \in [t_i, t_{i+1}[$$

and then

$$\begin{aligned}
\mathcal{F}_{\mathcal{X},\mathcal{D}}(u_n) &= \mathcal{M}_{\phi}^{(n)}\left(\mathcal{F}_{\mathcal{X},[t_1,t_2[}(U_{\mathcal{X},\mathcal{D},p}), \ldots, \mathcal{F}_{\mathcal{X},[t_n,t_{n+1}[}(U_{\mathcal{X},\mathcal{D},p})\right) \\
&= \mathcal{M}_{\phi}^{(n)}\left(G_{\mathcal{X},\mathcal{D}}[t_1; u], \ldots, G_{\mathcal{X},\mathcal{D}}[t_n; U_{\mathcal{X},\mathcal{D},p}]\right) \\
&= \mathcal{M}_{\phi}^{(n)}(p, \ldots, p) \\
&= p
\end{aligned}$$

moreover $\mathcal{M}_{\phi}^{(n)}\left(\mathbf{G}_{\mathcal{X},\mathcal{D}}[t_1; U_{\mathcal{X},\mathcal{D},p}], \ldots, \mathbf{G}_{\mathcal{X},\mathcal{D}}[t_n; U_{\mathcal{X},\mathcal{D},p}]\right)$ is a QAMM distribution, and then a probability.

Definition 5.6.2. *Let $n \in I\!N$:*

- *$\mathcal{X}$ be an* frv *and u a function, both belonging to $\mathcal{I}^{\mathcal{D}}$,*
- *$\mathbf{G}_{\mathcal{X},\mathcal{D}}$ the surface of margins associated to $\mathcal{X}$,*
- *ϕ a generator of an Archimedean copulas (and ψ its inverse),*
- *$\{t_1, \ldots, t_{n+1}\} \subset \mathcal{D}$, $n+1$ equidistant points dividing $\mathcal{D}$ in n intervals of equal length, and, such $t_1 = \min(\mathcal{D})$ and $t_n = \max(\mathcal{D})$.*

[1] In fact the second type of distribution respects the second compatibility condition, but for the F_i^* margins.

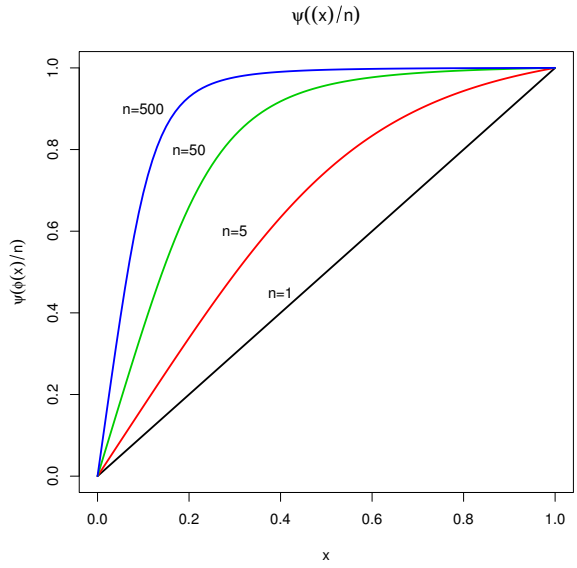


Figure 5.20: The transformation $f(u) = \psi\left(\frac{1}{n}\phi\left(u\right)\right)$.

If u_n is the following quantilized approximation of u:

$$u_n(t) = \sum_{i=1}^{n} V_{\mathcal{X},\mathcal{D},p_i}(t)\mathbf{1}_{[t_i,t_{i+1}[}(t) \tag{5.58}$$

where for all $i \in \{1,\ldots,n\}$, $V_{\mathcal{X},\mathcal{D},p_i}$ is a functional quantile of order $p_i = \mathbf{G}_{\mathcal{X},\mathcal{D}}[t_i;u]$.
Then, the Quasi-Arithmetic Mean of Margins *distribution of u_n is given by*

$$\begin{aligned} \mathcal{F}_{\mathcal{X},\mathcal{D}}(u_n) &= H_{QAMM,\mathcal{X}_{t_1},\ldots,\mathcal{X}_{t_n}}(u(t_1),\ldots,u(t_n)) & (5.59)\\ &= \psi\left[\frac{1}{n}\sum_{i=1}^{n}\phi\left(\mathbf{G}_{\mathcal{X},\mathcal{D}}[t_i;u]\right)\right]. & (5.60) \end{aligned}$$

The figure 5.20 gives an example of the transformation $\phi^{-1}(1/n\phi(.))$ using the *Clayton's generator*, with a parameter equal to 3 and several values of n. And the figure 5.21 shows the same transformation applied to the Gaussian cdf. We see clearly in these two figures, the "acceleration" toward 1 of the results, and this to compensate the volumetric behavior of the classical joint distributions.

Recall that a classical probability distribution of dimension n defines a probability measure on $\mathbb{R}^n$. If $(y_1,\ldots,y_n) \in \mathbb{R}^n$, then

$$H_{Classical}(y_1,\ldots,y_n) = P\left\{]-\infty,y_1] \times \ldots \times]-\infty,y_n]\right\}$$

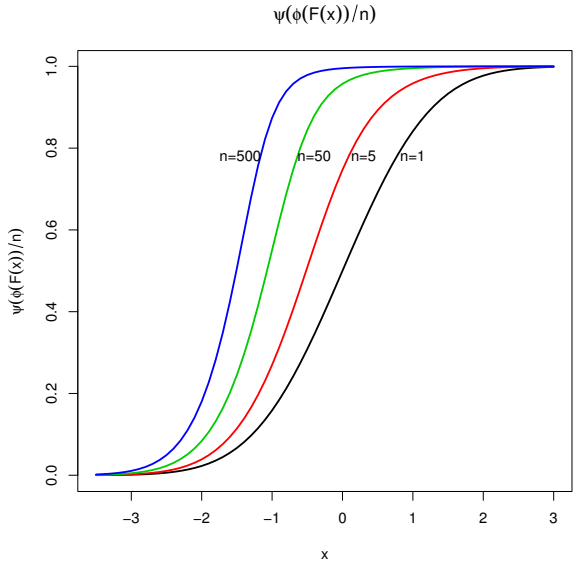


Figure 5.21: The transformation $\psi\left(\frac{1}{n}\phi\left(F(x)\right)\right)$ for the normal distribution.

i.e. the classical joint distribution computed in $(y_1, \ldots, y_n)$ is a probability measure of the n-cube $]-\infty, y_1] \times \ldots \times]-\infty, y_n]$.

While, if we work in the *functional data analysis* framework, then we suppose that there exists a set of n equidistant points $\{t_1, \ldots, t_{n+1}\} \subset \mathcal{D}$ and a surface of distributions $G_{\mathcal{X},\mathcal{D}}$, and a function u such $(y_1, \ldots, y_n) = (u(t_1), \ldots, u(t_n))$, and the following quantilized function is considered as an approximation for u:

$$u_n(t) = \sum_{i=1}^{n} V_{\mathcal{X},\mathcal{D},G_{\mathcal{X},\mathcal{D}}(t_i,y_i)}(t)\ \mathbf{1}_{[t_i,t_{i+1}[}(t)$$

where $G_{\mathcal{X},\mathcal{D}}(t_i, y_i)$ is the surface of distributions calculated for the point (t_i, y_i) (remember that $\mathbf{G}_{\mathcal{X},\mathcal{D}}[t; u] = G(t, u(t))$).

Then a QAMM distribution define the following probability measure on a portion of $\mathbb{R}^2$: $\mathcal{D} \times \mathbb{R}$:

$$\begin{aligned} H_{QAMM}(y_1, \ldots, y_n) &= \psi\left(\frac{1}{n}\sum_{i=1}^{n} G_{\mathcal{X},\mathcal{D}}(t_i; y_i)\right) \\ &= \psi\left(\frac{1}{n}\sum_{i=1}^{n} \mathbf{G}_{\mathcal{X},\mathcal{D}}[t_i; u_n]\right) \\ &= \mathcal{F}_{\mathcal{X},\mathcal{D}}\left(u_n\right) \\ &= P\left\{(t, y) \in \mathcal{D} \times \mathbb{R} : y \leq u_n(t)\right\}. \end{aligned}$$

Remark 5.6.1. Quasi-arithmetic means ***in conjunction with a surface of margins respect the conservation of functional quantiles, while it is not the case when we use a copula, and a fortiori, when we use an Archimedean copula.***

However the proposition 5.6.2 shows us that a QAMM distribution with a generator ϕ has the dependence structure of the Archimedean copula generated by ϕ.

This point is very important, because a QAMM distribution will share the same properties and weaknesses of the associated Archimedean copulas.

We are, now, able to show that, the condition of conservation of functional quantiles which has led us to the use of *quasi-arithmetic means*, leads us also to the "stability" of this kind of joint distribution for different values for n.

Proposition 5.6.3. *Suppose that for $n \in \mathbb{N}$, we divide $\mathcal{D}$ in $n+1$ intervals of equal length, i.e.*

- $\left\{t_1^n, \ldots, t_{n+1}^n\right\}$ *are equidistant points of* $\mathcal{D}$
- $t_1^n = \inf(\mathcal{D})$ *and* $t_{n+1}^n = \sup(\mathcal{D})$,
- $\forall i \in \{1, \ldots, n\}$ $\left|t_{i+1}^n - t_i^n\right| = \frac{|\mathcal{D}|}{n} = \Delta_t$.

Then for $u \in \mathcal{I}^{\mathcal{D}}$ and for any $n \in I\!N$, we have:

$$q \leq \psi\left[\frac{1}{n}\sum_{i=1}^{n}\phi\left(\mathbf{G}_{\mathcal{X},\mathcal{D}}\left[t_i^n; u\right]\right)\right] \leq p \tag{5.61}$$

with: $q = \min\left\{\mathbf{G}_{\mathcal{X},\mathcal{D}}[x;u] | x \in \mathcal{D}\right\}$ *and* $p = \max\left\{\mathbf{G}_{\mathcal{X},\mathcal{D}}[x;u] | x \in \mathcal{D}\right\}$.

Proof. It is a direct consequence of the proposition 5.5.3 (p.178). □

Then taking the limit of expression (5.60) does not always leads us to a zero limit like in the classical case.

The following result is about the comparison between the arithmetic mean and the quasi-arithmetic means.

Proposition 5.6.4. *Let $n \in \mathbb{N}_0$, $\{F_i | 1 \leq i \leq n\}$ be a set of one dimensional* cdf, *and ϕ a generator of Archimedean copula, then*

$$\psi\left(\frac{1}{n}\sum_{i=1}^{n}\phi\left(F_i(y_i)\right)\right) \leq \frac{1}{n}\sum_{i=1}^{n}F_i(y_i) \tag{5.62}$$

Proof. The proof is straightforward using the proposition 4.4.1 p.105 and the complete monotonicity of ϕ. □

Original contribution(s) 8. *In this section we show that, if the generator of a quasi-arithmetic mean is also a generator of Archimedean copulas, then the quasi-arithmetic mean of a set of univariate* cdfs*, gives a multivariate* cdf*. We called these distributions, QAMM distributions (Quasi-Arithmetic Mean of Margins). But this kind of joint distribution do not fulfill anymore the second compatibility condition of the Daniell-Kolmogorov's extension theorem.*

We show then, that the QAMM distribution calculated for a quantilized function is equal to the corresponding fcdf *calculated in this latter. We show also that, a QAMM distribution computed in any quantilized version of a function is always bounded, and its limits depend only of the function, and not of the number of discretization points. Then, the limit of a sequence of QAMM calculated in quantilized functions converging toward a chosen function, is not always equal to zero, like in the classical case.*

5.7 QAMML: Distributions for functional data

Definition 5.7.1. *Let $\mathcal{X}$ be an* frv*, and u a function, both belonging to $\mathcal{I}^{\mathcal{D}}$, $\mathbf{G}_{\mathcal{X},\mathcal{D}}$ the* Surface of Margins *associated to $\mathcal{X}$, and ϕ a generator of Archimedean Copulas. We define the* Quasi-Arithmetic Mean of Margins Limit (QAMML) *distribution of $\mathcal{X}$ by:*

$$P[\mathcal{X} \leq_{\mathcal{D}} u] = \mathcal{F}_{\mathcal{X},\mathcal{D},\phi}(u) = \psi\left[\frac{1}{|\mathcal{D}|}\int_{\mathcal{D}} \phi\left(\mathbf{G}_{\mathcal{X},\mathcal{D}}\left[t;u\right]\right) dt\right] \tag{5.63}$$

If θ is the parameter for $\mathbf{G}_{\mathcal{X},\mathcal{D}}$ and α is the parameter (numbers or vectors) for the function ϕ, then we will use the following notation when it is necessary for disambiguation: $\mathcal{F}_{\mathcal{X},\mathcal{D},\theta,\alpha}(u)$.

QAMML distributions are directly defined in the non denumerable set $\mathcal{D}$ and then permit to handle questions which are not treatable with the framework of finite dimensional distributions (see the Kolmogorov's objection p.51).

Proposition 5.7.1. *Let $\mathcal{X}$ be an* frv*, the* Quasi-Arithmetic Mean of Margins Limit *distribution of $\mathcal{X}$ is a* fractal and quantile based distribution*.*

Proof. Let us show that the three properties of the definition 5.3.1 are fulfilled.

Property 1:

Let S a closed interval of $\mathcal{D}$, and $V_{\mathcal{X},\mathcal{D},p}$ a functional quantile of order

$p \in [0,1]$:

$$\begin{aligned}
\mathcal{F}_{\mathcal{X},S,\phi}(V_{\mathcal{X},\mathcal{D},p}) &= \psi\left[\frac{1}{|S|}\int_S \phi\left(\mathbf{G}_{\mathcal{X},\mathcal{D}}\left[t; V_{\mathcal{X},\mathcal{D},p}\right]\right) dt\right] \\
&= \psi\left[\frac{1}{|S|}\int_S \phi\left(\mathbf{G}_{\mathcal{X},S}\left[t; V_{\mathcal{X},\mathcal{D},p}\right]\right) dt\right] \\
&= \psi\left[\frac{1}{|S|}\int_S \phi\left(p\right) dt\right] \\
&= \psi\left[\phi(p)\right] \\
&= p.
\end{aligned}$$

Property 2:

Let u_n a quantilized function belonging in $\mathcal{I}^{\mathcal{D}}$ defined by the following partition of $\mathcal{D}$: $\mathcal{A} = \{A_1, \ldots, A_n\}$ closed intervals of equal length, and a set of probabilities $\mathbf{p} = (p_1, \ldots, p_n)$:

$$u_n(t) = \sum_{k=1}^{n} V_{\mathcal{X},\mathcal{D},p_k}(t)\mathbf{1}_{A_k}(t) \ \forall t \in \mathcal{D}$$

then

$$\begin{aligned}
\mathcal{F}_{\mathcal{X},\mathcal{D},\phi}(u_n) &= \psi\left[\frac{1}{|\mathcal{D}|}\int_{\mathcal{D}} \phi\left(\mathbf{G}_{\mathcal{X},\mathcal{D}}\left[t; u_n\right]\right) dt\right] \\
&= \psi\left[\frac{1}{|n|}\sum_{i=1}^{n}\frac{n}{|\mathcal{D}|}\int_{A_i} \phi\left(\mathbf{G}_{\mathcal{X},\mathcal{D}}\left[t; u_n\right]\right) dt\right] \\
&= \psi\left[\frac{1}{|n|}\sum_{i=1}^{n}\frac{1}{|A_i|}\int_{A_i} \phi\left(\mathbf{G}_{\mathcal{X},A_i}\left[t; u_n\right]\right) dt\right] \\
&= \psi\left[\frac{1}{|n|}\sum_{i=1}^{n}\phi\left\{\psi\left[\frac{1}{|A_i|}\int_{A_i} \phi\left(\mathbf{G}_{\mathcal{X},A_i}\left[t; u_n\right]\right) dt\right]\right\}\right] \\
&= \psi\left[\frac{1}{|n|}\sum_{i=1}^{n}\phi\left\{\mathcal{F}_{\mathcal{X},A_i}(u_n)\right\}\right] \\
&= \mathcal{M}_{\phi}^{(n)}\left(\mathcal{F}_{\mathcal{X},A_1}(u_n), \ldots, \mathcal{F}_{\mathcal{X},A_n}(u_n)\right) \\
&= \mathcal{M}_{\phi}^{(n)}\left(\mathcal{F}_{\mathcal{X},A_1}(V_{\mathcal{X},\mathcal{D},p_1}), \ldots, \mathcal{F}_{\mathcal{X},A_n}(V_{\mathcal{X},\mathcal{D},p_n})\right) \\
&= \mathcal{M}_{\phi}^{(n)}\left(\mathcal{F}_{\mathcal{X},A_1}(V_{\mathcal{X},A_1,p_1}), \ldots, \mathcal{F}_{\mathcal{X},A_n}(V_{\mathcal{X},A_n,p_n})\right) \\
&= \mathcal{M}_{\phi}^{(n)}\left(p_1, \ldots, p_n\right).
\end{aligned}$$

Property 3: If, for a given function u belonging in $\mathcal{I}^{\mathcal{D}}$, then quantilized approximation u_n of u is given by the expression 5.58, using $n+1$ equidistant

points t_i in $\mathcal{D}$:

$$\begin{aligned}
\lim_{n\to\infty} \mathcal{F}_{\mathcal{X},\mathcal{D},\phi}(u_n) &= \lim_{n\to\infty} \psi\left[\frac{1}{n}\sum_{i=1}^{n} \phi\left(\mathbf{G}_{\mathcal{X},\mathcal{D}}\left[t_i;u\right]\right)\right] \\
&= \lim_{n\to\infty} \psi\left[\frac{1}{|\mathcal{D}|}\sum_{i=1}^{n} \frac{|\mathcal{D}|}{n}\cdot\phi\left(\mathbf{G}_{\mathcal{X},\mathcal{D}}\left[t_i;u\right]\right)\right] \\
&= \lim_{n\to\infty} \psi\left[\frac{1}{|\mathcal{D}|}\sum_{i=1}^{n} \Delta_t\cdot\phi\left(\mathbf{G}_{\mathcal{X},\mathcal{D}}\left[t_i;u\right]\right)\right] \\
&= \psi\left[\frac{1}{|\mathcal{D}|}\int_{\mathcal{D}} \phi\left(\mathbf{G}_{\mathcal{X},\mathcal{D}}\left[t;u\right]\right)dt\right] \\
&= \mathcal{F}_{\mathcal{X},\mathcal{D},\phi}(u).
\end{aligned}$$

□

Now, let us show that QAMML distributions have basic properties similar to the real case (see proposition 3.2.2, p.38).

Proposition 5.7.2. *Let $\mathcal{X}$ be an* frv*, $\mathcal{F}_{\mathcal{X},\mathcal{D},\phi}$ its* fcdf *and $u,v\in\mathcal{I}^{\mathcal{D}}$, then $\mathcal{F}_{\mathcal{X},\mathcal{D},\phi}$ has the following properties:*

1. *$\mathcal{F}_{\mathcal{X},\mathcal{D},\phi}$ is monotone increasing, i.e.*

$$u <_{\mathcal{D}} v \Rightarrow \mathcal{F}_{\mathcal{X},\mathcal{D},\phi}(u) \leq_{\mathcal{D}} \mathcal{F}_{\mathcal{X},\mathcal{D},\phi}(v),$$

2. *if ϕ is strictly monotone, then $\mathcal{F}_{\mathcal{X},\mathcal{D},\phi}$ is "right-continuous", i.e. $\forall u\in\mathcal{I}^{\mathcal{D}}$ we have*

$$\forall\varepsilon>0,\ \exists\delta\in\mathcal{I}^{\mathcal{D}}:\delta>_{\mathcal{D}}0:\ u<_{\mathcal{D}}v<_{\mathcal{D}}u+\delta\Rightarrow|\mathcal{F}_{\mathcal{X},\mathcal{D},\phi}(v)-\mathcal{F}_{\mathcal{X},\mathcal{D},\phi}(u)|<\varepsilon,$$

3. $\lim_{u\to-\infty} \mathcal{F}_{\mathcal{X},\mathcal{D},\phi}(u)=0$,

4. $\lim_{u\to+\infty} \mathcal{F}_{\mathcal{X},\mathcal{D},\phi}(u)=1$.

Proof. 1. If $u<_{\mathcal{D}}v$ then as *cdf* are monotone increasing (see proposition 3.2.2) we have

$$\forall t\in\mathcal{D}:\mathbf{G}_{\mathcal{X},\mathcal{D}}[t;u]\leq\mathbf{G}_{\mathcal{X},\mathcal{D}}[t;v],$$

and then using proposition 5.5.8 we have

$$\mathcal{F}_{\mathcal{X},\mathcal{D},\phi}(u)\leq_{\mathcal{D}}\mathcal{F}_{\mathcal{X},\mathcal{D},\phi}(v).$$

2. Let $\varepsilon > 0$, to build $\delta >_{\mathcal{D}} 0$ such $u <_{\mathcal{D}} v <_{\mathcal{D}} u + \delta \Rightarrow |\mathcal{F}_{\mathcal{X},\mathcal{D},\phi}(v) - \mathcal{F}_{\mathcal{X},\mathcal{D},\phi}(u)| < \varepsilon$ we are going to define a number η, used to build a function γ, and use this latter to obtain δ. Thus let us define

$$\eta = \psi\left\{\frac{1}{|\mathcal{D}|}\int_{\mathcal{D}} \phi\left(\mathbf{G}_{\mathcal{X},\mathcal{D}}[t;u]\right) dt - \phi\left(\mathcal{F}_{\mathcal{X},\mathcal{D},\phi}(u) + \varepsilon\right)\right\}. \tag{5.64}$$

Like

$$\begin{aligned} & \mathcal{F}_{\mathcal{X},\mathcal{D},\phi}(u) < \mathcal{F}_{\mathcal{X},\mathcal{D},\phi}(u) + \varepsilon \\ \Leftrightarrow \quad & \psi\left(\frac{1}{|\mathcal{D}|}\int_{\mathcal{D}} \phi\left(\mathbf{G}_{\mathcal{X},\mathcal{D}}[t;u]\right) dt\right) < \mathcal{F}_{\mathcal{X},\mathcal{D},\phi}(u) + \varepsilon \\ \Leftrightarrow \quad & \frac{1}{|\mathcal{D}|}\int_{\mathcal{D}} \phi\left(\mathbf{G}_{\mathcal{X},\mathcal{D}}[t;u]\right) dt > \phi\left(\mathcal{F}_{\mathcal{X},\mathcal{D},\phi}(u) + \varepsilon\right) \end{aligned}$$

and then

$$0 < \frac{1}{|\mathcal{D}|}\int_{\mathcal{D}} \phi\left(\mathbf{G}_{\mathcal{X},\mathcal{D}}[t;u]\right) dt - \phi\left(\mathcal{F}_{\mathcal{X},\mathcal{D},\phi}(u) + \varepsilon\right) \leq \max_{t\in[0,1]} \phi(t)$$

thus $0 \leq \eta < 1$.

Now, for any $t \in \mathcal{D}$, we define

$$\gamma(t) = \psi\left[\phi\left(\mathbf{G}_{\mathcal{X},\mathcal{D}}[t;u]\right) - \phi(\eta)\right] - \mathbf{G}_{\mathcal{X},\mathcal{D}}[t;u] \tag{5.65}$$

and like $\phi(\eta) > 0$

$$\begin{aligned} & \phi\left(\mathbf{G}_{\mathcal{X},\mathcal{D}}[t;u]\right) - \phi(\eta) < \phi\left(\mathbf{G}_{\mathcal{X},\mathcal{D}}[t;u]\right) \\ \Leftrightarrow \quad & \psi\left[\phi\left(\mathbf{G}_{\mathcal{X},\mathcal{D}}[t;u]\right) - \phi(\eta)\right] > \mathbf{G}_{\mathcal{X},\mathcal{D}}[t;u] \\ \Leftrightarrow \quad & \psi\left[\phi\left(\mathbf{G}_{\mathcal{X},\mathcal{D}}[t;u]\right) - \phi(\eta)\right] - \mathbf{G}_{\mathcal{X},\mathcal{D}}[t;u] > 0 \end{aligned}$$

and then, like $0 \leq \phi(t), \mathbf{G}_{\mathcal{X},\mathcal{D}}[t;u] \leq 1, \forall t$, we have that $0 < \gamma(t) < 1$.

Now, we set

$$\delta(t) = Q_{\mathcal{X},\mathcal{D}}\left(t, \gamma(t) + \mathbf{G}_{\mathcal{X},\mathcal{D}}[t,u]\right) - u(t) \tag{5.66}$$

where $Q_{\mathcal{X},\mathcal{D}}(t,.)$ is, for a fixed t the lowest functional quantile for $G_{\mathcal{X},\mathcal{D}}(t,.)$. Again like $\gamma(t) > 0$, we have

$$\begin{aligned} & \gamma(t) + \mathbf{G}_{\mathcal{X},\mathcal{D}}[t,u] > \mathbf{G}_{\mathcal{X},\mathcal{D}}[t;u] \\ \Leftrightarrow \quad & Q_{\mathcal{X},\mathcal{D}}\left(t, \gamma(t) + \mathbf{G}_{\mathcal{X},\mathcal{D}}[t,u]\right) > u(t) \end{aligned}$$

and then $\delta >_{\mathcal{D}} 0$.

Now we compute the following expression

$$\begin{aligned}
\mathcal{F}_{\mathcal{X},\mathcal{D},\phi}(u+\delta) &= \psi\left[\frac{1}{|\mathcal{D}|}\int_{\mathcal{D}}\phi\left(\mathbf{G}_{\mathcal{X},\mathcal{D}}\left[t;u+\delta\right]\right)dt\right] \\
&= \psi\left[\frac{1}{|\mathcal{D}|}\int_{\mathcal{D}}\phi\left(\gamma(t)+\mathbf{G}_{\mathcal{X},\mathcal{D}}[t,u]\right)dt\right] \textit{ by } (5.66) \\
&= \psi\left[\frac{1}{|\mathcal{D}|}\int_{\mathcal{D}}\left\{\phi\left(\mathbf{G}_{\mathcal{X},\mathcal{D}}[t,u]\right)-\phi(\eta)\right\}dt\right] \textit{ by } (5.65) \\
&= \psi\left[\frac{1}{|\mathcal{D}|}\int_{\mathcal{D}}\phi\left(\mathbf{G}_{\mathcal{X},\mathcal{D}}[t,u]\right)dt-\phi(\eta)\right] \\
&= \psi\left[\phi\left(\mathcal{F}_{\mathcal{X},\mathcal{D},\phi}(u)+\varepsilon\right)\right] \textit{ by } (5.64) \\
&= \varepsilon+\mathcal{F}_{\mathcal{X},\mathcal{D},\phi}(u)
\end{aligned}$$

and finally

$$\mathcal{F}_{\mathcal{X},\mathcal{D},\phi}(u+\delta)-\mathcal{F}_{\mathcal{X},\mathcal{D},\phi}(u)=\varepsilon$$

which implies

$$u<_{\mathcal{D}}v<_{\mathcal{D}}u+\delta\Rightarrow\left|\mathcal{F}_{\mathcal{X},\mathcal{D},\phi}(v)-\mathcal{F}_{\mathcal{X},\mathcal{D},\phi}(u)\right|<\varepsilon.$$

3. We have to prove that

$$\forall\varepsilon>0\exists\delta\in\mathcal{I}^{\mathcal{D}}:\ v\leq_{\mathcal{D}}\delta\Rightarrow\mathcal{F}_{\mathcal{X},\mathcal{D},\phi}(v)\leq\varepsilon.$$

If $0<\varepsilon<1$, then it suffices to set $\delta=Q_{\mathcal{X},\mathcal{D},\varepsilon}$.

4. We have to prove that

$$\forall\varepsilon>0\exists\delta\in\mathcal{I}^{\mathcal{D}},:\ v\geq_{\mathcal{D}}\delta\Rightarrow\left|1-\mathcal{F}_{\mathcal{X},\mathcal{D},\phi}(v)\right|\leq\varepsilon.$$

If $0<\varepsilon<1$, then it suffices to set $\delta=Q_{\mathcal{X},\mathcal{D},1-\varepsilon}$.

□

Remark 5.7.1. ***Let us note that in the expression (5.63) of the* QAMML *distribution we need to compute or estimate the surface of margins*** $G_{\mathcal{X},\mathcal{D}}(x,y)$***, and it is impossible if there is no dispersion. We can say that the domain of the* QAMML *model is the set of real numbers such that the statistical dispersion is greater then zero. Thus, from here, we suppose that*** $\mathcal{D}\subseteq\{x\in\mathbb{R}:\sigma(x)>0\}$ ***, and we call*** $\mathcal{D}$ ***the* model's domain.**

The corollary 5.5.11 implies that QAMML distributions have the same relation with the classical arithmetic mean that QAMM distributions.

Proposition 5.7.3. *Let* $\mathcal{X}$ *be an* frv*,* $\mathbf{G}_{\mathcal{X},\mathcal{D}}$ *its surface of margins and* $\mathcal{F}_{\mathcal{X},\mathcal{D}}$ *its QAMML distribution.*
If $u\in\mathcal{I}^{\mathcal{D}}$*, then*

$$\mathcal{F}_{\mathcal{X},\mathcal{D},\phi}(u)\leq\frac{1}{|\mathcal{D}|}\int_{\mathcal{D}}\mathbf{G}_{\mathcal{X},\mathcal{D}}\left[t;u\right]dt \tag{5.67}$$

The proposition 5.5.9 implies directly the following proposition, equivalent to the proposition 5.6.3 in the discrete case.

Proposition 5.7.4. *Let $\mathcal{X}$ be an* frv*, and u a function, both belonging to $\mathcal{I}^{\mathcal{D}}$, $\mathbf{G}_{\mathcal{X},\mathcal{D}}$ the* Surface of Margins *associated to $\mathcal{X}$, and ϕ a generator of Archimedean Copulas.*

Then, the QAMML of $\mathcal{X}$ is such that

$$q \leq \mathcal{F}_{\mathcal{X},\mathcal{D},\phi}(u) \leq p \tag{5.68}$$

with: $q = \min\{\mathbf{G}_{\mathcal{X},\mathcal{D}}[t;u] | t \in \mathcal{D}\}$ *and* $p = \max\{\mathbf{G}_{\mathcal{X},\mathcal{D}}[t;u] | t \in \mathcal{D}\}$.

At the end of the section 4.8 we have explained the main weaknesses of the classical approach using finite dimensional distributions:

1. the "volumetric behavior" which implies the instability of estimations depending of the dimension and the vanishing toward 0 when n tends to infinity,
2. the quasi-impossibility, in real situations, to share the same cylinder,
3. the precision issue in case of high dimensionality.

By definition the QAMML distributions are defined, not on cylinders like finite dimensional distributions, but directly in the infinite dimensional space $\mathcal{I}^{\mathcal{D}}$. Thus the problem of sharing the same cylinder is solved very simply: we avoid this kind of approximation.

The fulfillment of the **Property 1** and the propositions 5.6.3 and 5.7.4 ensure us the stability of the probability:

- for functional quantiles $V_{\mathcal{X},\mathcal{D},p}$, we know the $\mathcal{F}_{\mathcal{X},\mathcal{D},\phi}(V_{\mathcal{X},\mathcal{D},p}) = p$,
- for any other function $u \in \mathcal{I}^{\mathcal{D}}$ we know that the QAMML is always between $min\{\mathbf{G}_{\mathcal{X},\mathcal{D}}[x;u] | x \in \mathcal{D}\}$ and $\max\{\mathbf{G}_{\mathcal{X},\mathcal{D}}[t;u] | t \in \mathcal{D}\}$.

These latter points also solve the precision issue, because the value of the *fcdf* does not depends of any order of the estimation.

Moreover **Property 2** and **Property 3** allow us to use QAMM distributions as an approximation of the n^{th} order to compute the QAMML distribution of a function u.

Let us suppose that we have to deal with N functional data like the Tecator data (see section 2.3), directly given by n measures in n points of $\mathcal{D}$:

$$(u_k(t_1), \ldots, u_k(t_n)),\ 1 \leq k \leq N.$$

Then we can compute directly the following (5.60) approximation

$$\mathcal{F}_{\mathcal{X},\mathcal{D},\phi}(u_k) \approx \mathcal{F}_{\mathcal{X},\mathcal{D},\phi}(u_{k,n}) = \psi\left[\frac{1}{n}\sum_{i=1}^{n}\phi\left(\mathbf{G}_{\mathcal{X},\mathcal{D}}[t_i;u_k]\right)\right].$$

where expression (5.58) give a quantilized approximation for u_k:

$$u_{k,n}(t) = \sum_{i=1}^{n} V_{\mathcal{X},\mathcal{D},p_i}(t)\mathbf{1}_{[t_i,t_{i+1}[}(t)$$

with $p_i = \mathbf{G}_{\mathcal{X},\mathcal{D}}[t_i; u_k]$ for all $i \in \{1,\ldots,n\}$.

Thus, even if QAMML distributions are defined directly in the infinite dimensional functional space, it is possible to work with any discrete approximation of a function u, and in this way avoid using the numerical approximation of the integral included in definition 5.7.1.

The following proposition gives another kind of approximation for the QAMML of the *fcdf* of an *frv* $\mathcal{X}$, and shows how the variations between a function u and an associated functional quantile $V_{\mathcal{X},\mathcal{D},m}$, where m is the arithmetic mean of $\mathbf{G}_{\mathcal{X},\mathcal{D}}[t; u]$ along $\mathcal{D}$, link the arithmetic mean to the QAMML distribution.

Proposition 5.7.5. *Let u be a function defined on $\mathcal{D}$, $G_{\mathcal{X},\mathcal{D}}$ a surface of distribution for $\mathcal{D}$. If we define*

$$m = \frac{1}{|\mathcal{D}|}\int_{\mathcal{D}} \mathbf{G}_{\mathcal{X},\mathcal{D}}\left[t; u\right] dt \tag{5.69}$$

and

$$\epsilon_m\left[t; u\right] = \mathbf{G}_{\mathcal{X},\mathcal{D}}\left[t; u\right] - m \tag{5.70}$$

then

$$\mathcal{F}_{\mathcal{X},\mathcal{D},\phi}(u) = \psi\left[\phi\left(m\right) + \frac{\phi''\left(m\right)}{2} var\left(\mathbf{G}_{\mathcal{X},\mathcal{D}}\left[\ .\ ; u\right]\right) + \mathbb{E}\left(o(\epsilon_m^2)\right)\right] \tag{5.71}$$

Proof. We can use the following Taylor's approximation for all t (recall that $0 \leq \mathbf{G}_{\mathcal{X},\mathcal{D}}\left[t; u\right] \leq 1$):

$$\begin{aligned}
\phi\left(\mathbf{G}_{\mathcal{X},\mathcal{D}}\left[t; u\right]\right) &= \phi\left(m + \epsilon_m\left[t; u\right]\right) \\
&= \phi(m) + \phi'(m)\epsilon_m\left[t; u\right] + \frac{\phi''(m)}{2}\epsilon_m^2\left[t; u\right] + o\left(\epsilon_m^2\left[t; u\right]\right)
\end{aligned}$$

and then

$$\begin{aligned}
& \mathcal{F}_{\mathcal{X},\mathcal{D},\phi}(u) \\
= \ & \psi\left[\frac{1}{|\mathcal{D}|}\int_{\mathcal{D}} \phi\left(\mathbf{G}_{\mathcal{X},\mathcal{D}}\left[t; u\right]\right) dt\right] \\
= \ & \psi\left[\frac{1}{|\mathcal{D}|}\int_{\mathcal{D}} \left\{\phi(m) + \phi'(m)\epsilon_m\left[t; u\right] + \frac{\phi''(m)}{2}\epsilon_m^2\left[t; u\right] + o\left(\epsilon_m^2\left[t; u\right]\right)\right\} dt\right] \\
= \ & \psi\left[\phi\left(m\right) + \phi'\left(m\right)\mathbb{E}\left(\epsilon_p\right) + \frac{\phi''\left(m\right)}{2} var\left(\epsilon_m\right) + \mathbb{E}\left(o(\epsilon_m^2)\right)\right] \\
= \ & \psi\left[\phi\left(m\right) + \frac{\phi''\left(m\right)}{2} var\left(\mathbf{G}_{\mathcal{X},\mathcal{D}}\left[\ .\ ; u\right]\right) + \mathbb{E}\left(o(\epsilon_m^2)\right)\right]
\end{aligned}$$

□

The relation given by expression (5.71) gives us a formula for a quick approximation for $\mathcal{F}_{\mathcal{X},\mathcal{D},\phi}(u)$, and allows us to understand for a function u, the relation between the QAMML, m the arithmetic mean of $\mathbf{G}_{\mathcal{X},\mathcal{D}}[t;u]$ and functional quantiles $V_{\mathcal{X},\mathcal{D},m}$.

By the proposition 5.7.3 we know that

$$\mathcal{F}_{\mathcal{X},\mathcal{D},\phi}(u) \leq m = \frac{1}{|\mathcal{D}|}\int_{\mathcal{D}} \mathbf{G}_{\mathcal{X},\mathcal{D}}\left[t;u\right] dt$$

and then by expression (5.71) we see that $\mathcal{F}_{\mathcal{X},\mathcal{D},\phi}$ is equal to m if and only if $var\left(\mathbf{G}_{\mathcal{X},\mathcal{D}}\left[\ .\ ;u\right]\right) = 0$, i.e. only when u is a functional quantile.

Thus we can consider the four remarkable values for a given function u:

$$p = \max\left\{\mathbf{G}_{\mathcal{X},\mathcal{D}}[t;u] | t \in \mathcal{D}\right\} \tag{5.72}$$

$$m = \frac{1}{|\mathcal{D}|}\int_{\mathcal{D}} \mathbf{G}_{\mathcal{X},\mathcal{D}}\left[t;u\right] dt \tag{5.73}$$

$$\mathcal{M} = \mathcal{F}_{\mathcal{X},\mathcal{D},\phi}(u) \tag{5.74}$$

$$q = \min\left\{\mathbf{G}_{\mathcal{X},\mathcal{D}}[t;u] | t \in \mathcal{D}\right\} \tag{5.75}$$

And we have the following relation between these values

$$q \leq \mathcal{M} \leq m \leq p \tag{5.76}$$

which implies the following relation between corresponding functional quantiles

$$V_{\mathcal{X},\mathcal{D},q} \leq_{\mathcal{D}} V_{\mathcal{X},\mathcal{D},\mathcal{M}} \leq_{\mathcal{D}} V_{\mathcal{X},\mathcal{D},m} \leq_{\mathcal{D}} V_{\mathcal{X},\mathcal{D},p}. \tag{5.77}$$

And, of course, we have the following relation

$$V_{\mathcal{X},\mathcal{D},q} \leq_{\mathcal{D}} u \leq_{\mathcal{D}} V_{\mathcal{X},\mathcal{D},p}. \tag{5.78}$$

To illustrate this relation, we have chosen a function u of the Tecator dataset (section 2.3, figure 2.11), and we have computed the four following values: q, m, $\mathcal{M}$ and p (expressions (5.72), (5.73), (5.74) and (5.75)), when the surface of margins $G_{\mathcal{X},\mathcal{D}}$ is given by a normal distribution: $G_{\mathcal{X},\mathcal{D}}(t,.) = F_{\mathcal{X}_t}(.) = F_{\mathcal{N}(\mu(t),\sigma(t))}(.)$. These values are shown in the top left of the figure 5.22, and figure shows the three following associated functional quantiles: $Q_{\mathcal{X},\mathcal{D},p}$ (top dotted line), $Q_{\mathcal{X},\mathcal{D},m}$ (dashed line) and $Q_{\mathcal{X},\mathcal{D},q}$ (bottom dotted line).

An interesting implication of the formula given by the expression (5.71) is that $\mathcal{F}_{\mathcal{X},\mathcal{D},\phi}(u)$ decreases when $var\left(\mathbf{G}_{\mathcal{X},\mathcal{D}}\left[\ .\ ;u\right]\right)$ increases, and it is due to the fact that ϕ and ϕ'' are always positive functions, and ψ is a decreasing function. Therefore the value of the QAMML distribution computed for a function u, is directly influenced by the "variance" of u around any functional quantile $V_{\mathcal{X},\mathcal{D},m}$. To illustrate this fact, using the normal surface of margins

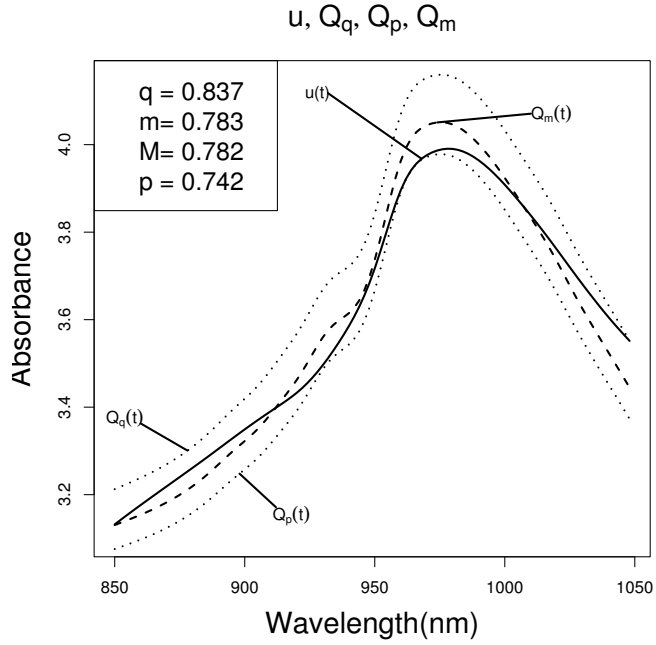


Figure 5.22: A functional data and 3 associated functional quantiles: $Q_{\mathcal{X},\mathcal{D},p}$, $Q_{\mathcal{X},\mathcal{D},m}$ and $Q_{\mathcal{X},\mathcal{D},q}$.

for the Tecator data, we have taken the functional quantile $Q_{\mathcal{X},\mathcal{D},m}$ shown in figure 5.22 with $m = 0.7834597$, and then we have added to this function several artificial perturbations $\delta(t)$, to define $u(t) = Q_{\mathcal{X},\mathcal{D},m}(t) + \delta(t)$. And, in each case, we have chosen $\delta(t)$ such

$$\frac{1}{|\mathcal{D}|}\int_{\mathcal{D}} \mathbf{G}_{\mathcal{X},\mathcal{D}}\left[t; Q_{\mathcal{X},\mathcal{D},m} + \delta\right] dt \ = m,$$

but with different values for $\sigma\left(\mathbf{G}_{\mathcal{X},\mathcal{D}}\left[.; Q_{\mathcal{X},\mathcal{D},m} + \delta\right]\right) > 0$.

Figures from 5.23 to 5.26 show in the bottom parts, graphs of the chosen δ, and in the top parts, solid lines show the graph of $u(t) = Q_{\mathcal{X},\mathcal{D},m}(t) + \delta(t)$, and dashed lines show the original functional quantile: $Q_{\mathcal{X},\mathcal{D},m}(t)$.

The table 5.2 shows the values of the QAMML distribution computed in $u(t) = Q_{\mathcal{X},\mathcal{D},m}(t) + \delta(t)$. The first column gives the standard deviation of u around its functional quantile $Q_{\mathcal{X},\mathcal{D},m}$: $\sigma\left(\mathbf{G}_{\mathcal{X},\mathcal{D}}\left[.; u\right]\right)$, and the first row gives the different values of the parameter of the Archimedean copulas used for the QAMML distribution, here the Clayton copula (see example 4.4.1, p.108).

Firstly, we have here an illustration of the fact that the QAMML is always less or equal to the arithmetic mean of margins (see proposition 5.67, p.194), because all values of this table are less than $m = 0.7834597$.

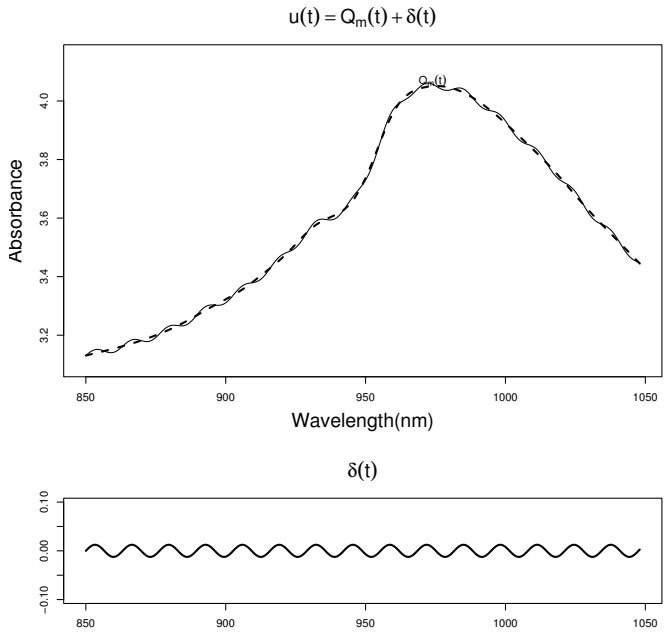


Figure 5.23: Variations around $Q_{\mathcal{X},\mathcal{D},m}$, when $m = 0.7834597$, with $\sigma\left(\mathbf{G}_{\mathcal{X},\mathcal{D}}\left[.;u\right]\right) = 0.005$.

Secondly we can see how the increasing of $\sigma\left(\mathbf{G}_{\mathcal{X},\mathcal{D}}\left[.;u\right]\right)$ implies a decreasing of the QAMML value (see proposition 5.71 p.196).

And finally, we can see also, the influence of the parameter. We have not yet studied the interpretation of the value of this latter, but we have an interpretation in the copula case (see example 4.4.1, p.108). At this point it is very important to recall the strong link between Archimedean copulas and QAMM/QAMML distributions.

In the remark 5.6.1 (p.189) we have emphasized how proposition 5.6.2 (p.185) implies the fact that the QAMM distributions share dependence structures of Archimedean copulas. Proposition 5.7.1 shows that for any n the QAMM distribution is an approximation of the QAMML distribution. Then the two kinds of distributions will capture the same type of dependence structure.

Let us recall that the three important copulas W (expressions (4.18) and (4.76), pp.85 and 125), Π (expressions (4.20) and (4.77)) and M (expressions (4.19) and (4.78)) give a direct interpretation of the dependence structure (see remark 4.3.1 p.95 for the bidimensional case and theorem 4.5.6 p.125 for higher dimensions).

In the case of the Clayton copula, small values of θ correspond to weak dependence structures between the real random variables, while higher val-

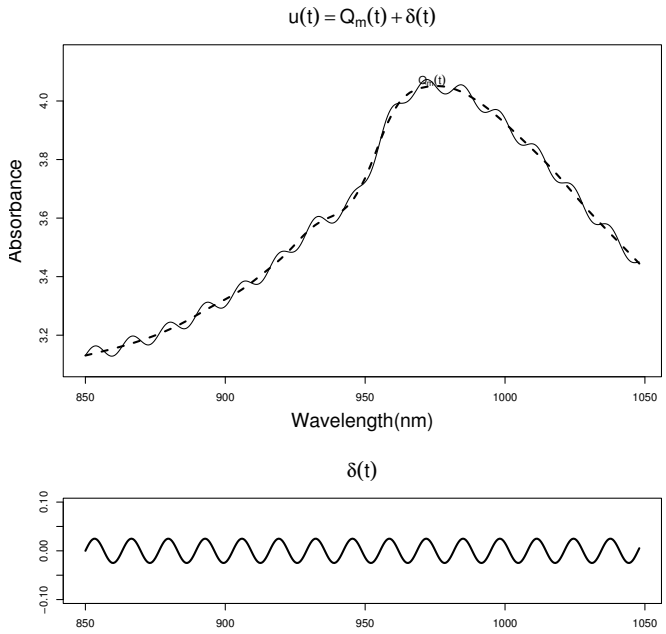


Figure 5.24: Variations around $Q_{\mathcal{X},\mathcal{D},m}$, when $m = 0.7834597$, with $\sigma\left(\mathbf{G}_{\mathcal{X},\mathcal{D}}\left[.;u\right]\right) = 0.010$.

ues correspond to strong positive dependence between the real random variables. In the first case the copula is close to Π and in the second case the copula tends toward $M = \min$. For the sake of illustration the last column of table 5.2 gives the value of $\min_{t\in\mathcal{D}} \mathbf{G}_{\mathcal{X},\mathcal{D}}[t,u]$.

$\sigma \setminus \theta$	0.5	2	10	100	1000	∞
0.005	0.78343	0.78341	0.78327	0.78186	0.77747	0.77472
0.010	0.78336	0.78326	0.78273	0.77773	0.76920	0.76585
0.020	0.78306	0.78266	0.78056	0.76573	0.75120	0.74776
0.040	0.78187	0.78028	0.77197	0.73552	0.71349	0.71021

Table 5.2: Influence on the QAMML value of $\sigma\left(\mathbf{G}_{\mathcal{X},\mathcal{D}}\left[.;u\right]\right)$ and of the parameter θ of the Archimedean copula (here Clayton).

If QAMM distributions share dependence structures of Archimedean copulas, then they also share the limitation of these latter ones.

The theorem 4.6.3 (p.129) states that for dimensions higher than 2, Archimedean copulas can only capture dependence structures comprised between independence (Π copula) and completely positive dependence (M copula).

Then, if there exists a partition $\{A, B\}$ of $\mathcal{D}$ such for:

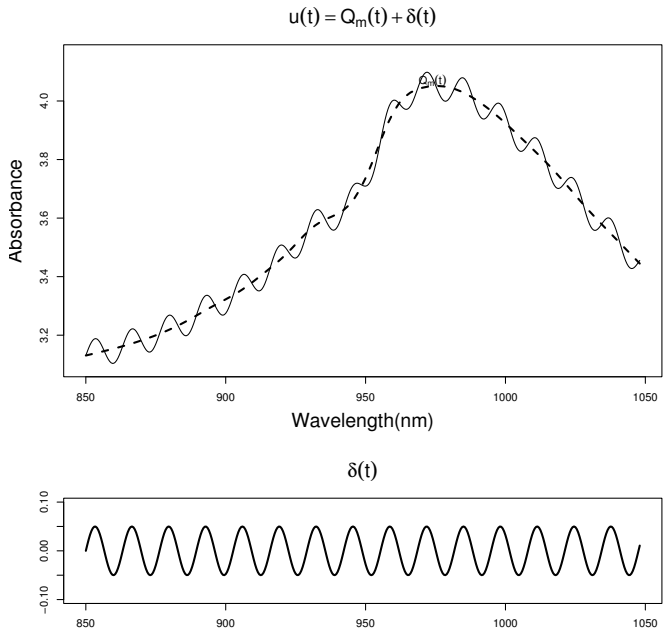


Figure 5.25: Variations around $Q_{\mathcal{X},\mathcal{D},m}$, when $m = 0.7834597$, with $\sigma\left(\mathbf{G}_{\mathcal{X},\mathcal{D}}\left[.;u\right]\right) = 0.020$.

- $\forall s,t \in A$, there is a positive dependence between $\mathcal{X}_s$ and $\mathcal{X}_t$ (comonotonic real random variables),
- $\forall s,t \in B$, there is a positive dependence between $\mathcal{X}_s$ and $\mathcal{X}_t$ (comonotonic real random variables),
- $\forall s \in A$ and $\forall t \in B$, there is a negative dependence between $\mathcal{X}_s$ and $\mathcal{X}_t$ (anti-comonotonic real random variables),

the QAMM and QAMML distributions will not be able to capture this latter kind of dependence.

When we use the QAMM approximation it is easy to see if there is this kind of dependence, it suffices to take a look at the sign of the Kendall's tau τ (definition 4.3.11, p.97) between t_1 and the other t_i. The Kendall's tau is a measure of association (definition 4.3.10, p.97) and is directly linked to the generator ϕ of an Archimedean copula (theorem 4.4.7 p.110).

Figures 5.27, 5.28 and 5.29 show the different types of situation. In the original data, fig. 5.27, all the measures of association made with τ are positives, then there is only dependence structures using copulas "greater" than the independence.

For the first derivatives, fig. 5.28, we have is two main zones: $A =$

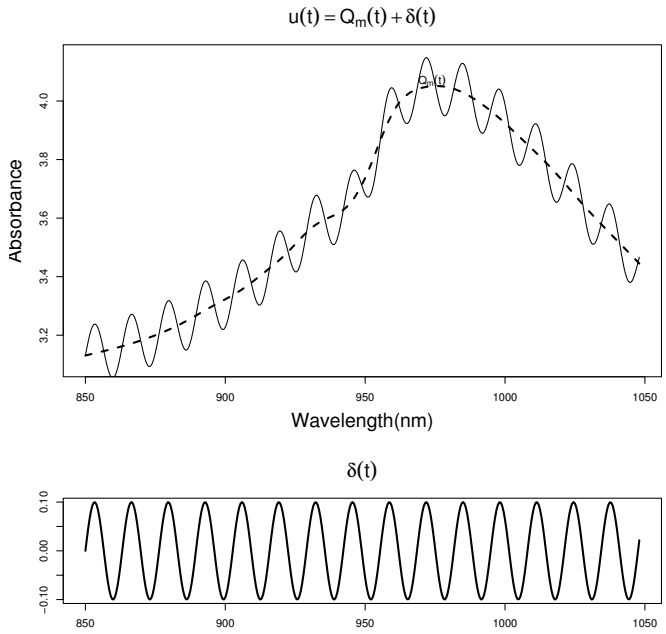


Figure 5.26: Variations around $Q_{\mathcal{X},\mathcal{D},m}$, when $m = 0.7834597$, with $\sigma\left(\mathbf{G}_{\mathcal{X},\mathcal{D}}\left[.;u\right]\right) = 0.040$.

$[850, 930] \cup [968, 1038]$ and $B = [932, 966] \cup [1040, 1048]$[2]

For the second derivative, fig. 5.29, there are more subdivisions: the first zone is $A = [850, 886] \cup [896, 910] \cup [936, 952] \cup [984, 1008] \cup [1022, 1030]$ and the second zone is given by $B = [888, 894] \cup [912, 934] \cup [954, 982] \cup [1010, 1020] \cup [1032, 1048]$.

In these two latter cases, generators of Archimedean copulas of dimensions higher than two are not able to capture the dependence structure. In practice the best settings will be the lowest possible bound for Archimedean copulas, when $n > 2$: the independence copula Π.

To solve this problem, let us remark that bidimensional Archimedean copulas do not have this limitation: some of these copulas can take values from W to M, passing by Π (see examples 4.4.1 and 4.4.2).

If $\mathcal{F}_{\mathcal{X},\mathcal{D},\phi}$ denotes the *fcdf* using the generator ϕ, then in any case we can always divide $\mathcal{D}$ in two arbitrary sub-domains A and B, and do the following

[2]We have computed the Kendall's tau on discrete values, that is why we do not really know what it happen in $[930, 932]$, $[966, 968]$ and $[1038, 1040]$.

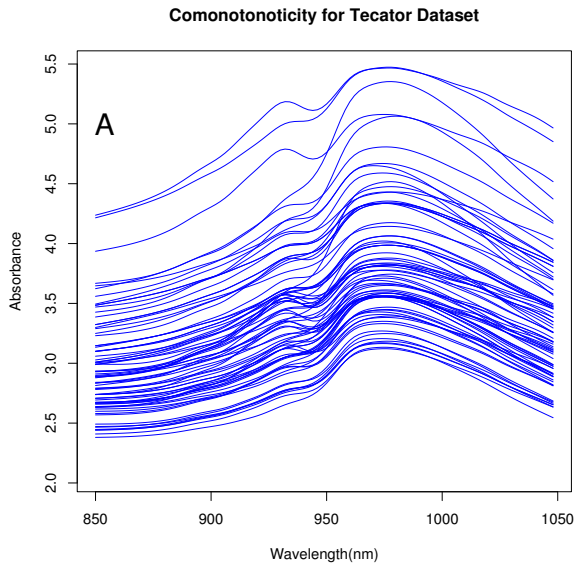


Figure 5.27: Set of comonotonoticity for the Tecator Dataset

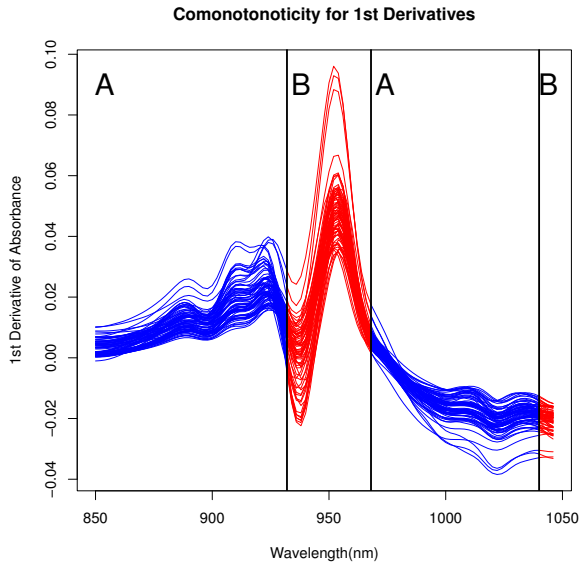


Figure 5.28: Sets of comonotonoticity for first derivatives of the Tecator Dataset

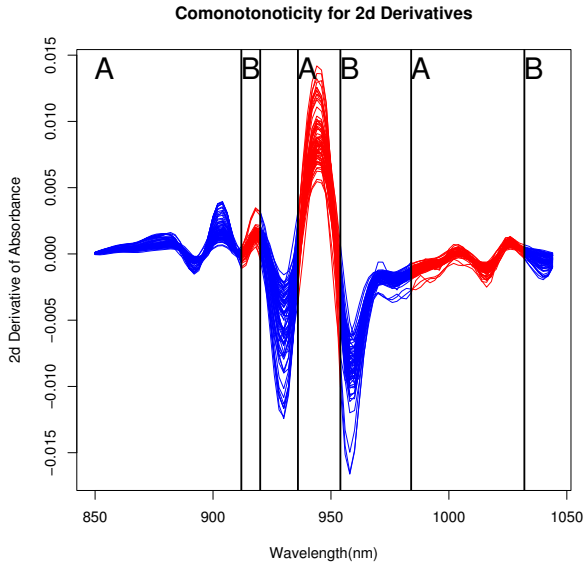


Figure 5.29: Sets of comonotonoticity for second derivatives of the Tecator Dataset

transformation:

$$\begin{aligned}
\mathcal{F}_{\mathcal{X},\mathcal{D},\phi}(u) &= \psi\left[\frac{1}{|\mathcal{D}|}\int_{\mathcal{D}}\phi\left(\mathbf{G}_{\mathcal{X},\mathcal{D}}[u;t]\right)dt\right] \\
&= \psi\left[\frac{1}{|\mathcal{D}|}\int_{A}\phi\left(\mathbf{G}_{\mathcal{X},\mathcal{D}}[u;t]\right)dt + \frac{1}{|\mathcal{D}|}\int_{B}\phi\left(\mathbf{G}_{\mathcal{X},\mathcal{D}}[u;t]\right)dt\right] \\
&= \psi\left[\frac{|A|}{|\mathcal{D}|}\psi\circ\phi\left\{\frac{1}{|A|}\int_{A}\phi\left(\mathbf{G}_{\mathcal{X},\mathcal{D}}[u;t]\right)dt\right\}\right. \\
&\quad \left. + \frac{|B|}{|\mathcal{D}|}\psi\circ\phi\left\{\frac{1}{|B|}\int_{B}\phi\left(\mathbf{G}_{\mathcal{X},\mathcal{D}}[u;t]\right)dt\right\}\right] \\
&= \psi\left(\frac{|A|}{|\mathcal{D}|}\phi\left[\mathcal{F}_{\mathcal{X},A,\phi}(u)\right] + \frac{|B|}{|\mathcal{D}|}\phi\left[\mathcal{F}_{\mathcal{X},B,\phi}(u)\right]\right). \qquad (5.79)
\end{aligned}$$

Then $\mathcal{F}_{\mathcal{X},\mathcal{D},\phi}(u)$ is the weighted mean of $\mathcal{F}_{\mathcal{X},A,\phi}(u)$ and $\mathcal{F}_{\mathcal{X},B,\phi}(u)$.

In the above expression the three *fcdf*: $\mathcal{F}_{\mathcal{X},\mathcal{D},\phi}(u)$, $\mathcal{F}_{\mathcal{X},A,\phi}(u)$ and $\mathcal{F}_{\mathcal{X},B,\phi}(u)$ use the same generator ϕ, but if we use two generators of Archimedean copulas of higher dimensions ϕ_A and ϕ_B for the *fcdfs* in A and in B, and a generator ϕ_2 of a bidimensional Archimedean copula for the weighted mean of $\mathcal{F}_{\mathcal{X},A,\phi_A}(u)$ and $\mathcal{F}_{\mathcal{X},B,\phi_B}(u)$, and then we will be able to capture a negative dependence structure between $\mathcal{F}_{\mathcal{X},A,\phi_A}(u)$ and $\mathcal{F}_{\mathcal{X},B,\phi_B}(u)$. That's how we define the generalization of the QAMML.

Definition 5.7.2. *Let an* frv $\mathcal{X}$ *and* u *both belonging to* $\mathcal{I}^{\mathcal{D}}$. *If there exists a partition* $\{A, B\}$ *of* $\mathcal{D}$ *such for:*

- $\forall s, t \in A$, *there is a positive dependence between* $\mathcal{X}_s$ *and* $\mathcal{X}_t$ *(comonotonic real random variables),*

- $\forall s, t \in B$, *there is a positive dependence between* $\mathcal{X}_s$ *and* $\mathcal{X}_t$ *(comonotonic real random variables),*

- $\forall s \in A$ *and* $\forall t \in B$, *there is a negative dependence between* $\mathcal{X}_s$ *and* $\mathcal{X}_t$ *(anti-comonotonic real random variables),*

then we define the Generalized Quasi-Arithmetic Mean of Margins Limit (GQAMML) *by*

$$\mathbb{F}_{\mathcal{X},\mathcal{D}}(u) = \psi_{\mathcal{D}}\left(\frac{|A|}{|\mathcal{D}|}\phi_{\mathcal{D}}\left[\mathcal{F}_{\mathcal{X},A,\phi_A}(u)\right] + \frac{|B|}{|\mathcal{D}|}\phi_{\mathcal{D}}\left[\mathcal{F}_{\mathcal{X},B,\phi_B}(u)\right]\right) \tag{5.80}$$

where

- $\mathcal{F}_{\mathcal{X},A,\phi_A}(u)$ *is the* fcdf *of* $\mathcal{X}$ *computed in* u *on the subset* A, *using* ϕ_A *a generator of Archimedean n-dimensional* ($n > 2$) *copula,*

- $\mathcal{F}_{\mathcal{X},B,\phi_B}(u)$ *is the* fcdf *of* $\mathcal{X}$ *computed in* u *on the subset* B, *using* ϕ_B *a generator of Archimedean n-dimensional* ($n > 2$) *copula,*

- $\phi_{\mathcal{D}}$ *is a generator of an bidimensional Archimedean copula, and* $\psi_{\mathcal{D}}$ *its inverse.*

GQAMML distributions are a generalization of QAMML distributions, which are able to capture positive and negative dependence structures of an *frv*. Moreover expression (5.79), shows that GQAMML are equal to QAMML when there is only positive dependence structures for an *frv*.

Original contribution(s) 9. *In this section we define the QAMML distributions (Quasi-Arithmetic Mean of Margins Limit) which are* fcdfs *directly defined in the infinite dimensional space of functions. We show then, that QAMML distributions are fractal and quantile based distributions, have similar properties to univariate* cdfs *and are always less or equal to the arithmetic mean of margins. We show also that, the QAMML distribution calculated in a function* u *depends of the standard deviation of* u *compared to the its own functional quantile (functional quantile based on the arithmetic mean of margins for* u*).*

We propose also a generalized version of QAMML distributions, the GQAMML distributions. These latter can capture comonotoniticity and anticomonotoniticity, which is not the case for QAMML distributions.

5.8 Gateaux Density

Even if a probability distribution is a precious tool for statistical applications, an *fcdf* is an incomplete tool for data analysis without an associate density. Then in this section we propose to deal with the density of QAMML distributions.

As long as we use finite dimensional distributions (4.122) or QAMM distributions (5.59), we can use the classical multivariate density function (4.83):

$$h(x_1, ..., x_n) = \frac{\partial^n}{\partial x_1 \dots \partial x_n} H(x_1, \dots, x_n)$$

But this definition can not be used when $n \to \infty$.

Let us recall that our definition of an *fcdf* (3.30) looks like the classical definition for univariate distributions, and when we compare the proposition 5.7.2 (p.192) we can see that the properties of the QAMML distributions are closer to the classical univariate case than to the classical multivariate case.

Let us recall that, for continuous real random variable (univariate case), the density f_X is used because it is not possible to use $P[X = i]$ as in the discrete case, and this because, in the continuous case, for a given *rrv* X and a given real r, we always have

$$\begin{aligned} P[X = r] &= \lim_{\delta \to 0} P[r - \delta \leq X \leq r + \delta] && (5.81) \\ &= \lim_{\delta \to 0} \{F_X(r + \delta) - F_X(r - \delta)\} && (5.82) \\ &= 0. && (5.83) \end{aligned}$$

And, it suffices to divide the above expression by 2δ to define the density

$$\begin{aligned} f_X(r) &= \lim_{\delta \to 0} \frac{P[r - \delta \leq X \leq r + \delta]}{2\delta} && (5.84) \\ &= \lim_{\delta \to 0} \frac{F_X(r + \delta) - F_X(r - \delta)}{2\delta}. && (5.85) \end{aligned}$$

Which can be re-written as follows

$$\begin{aligned} f_X(r) &= \lim_{\delta \to 0} \frac{P[r - \delta \leq X \leq r + \delta]}{d(r - \delta, r + \delta)} && (5.86) \\ &= \lim_{\delta \to 0} \frac{F_X(r + \delta) - F_X(r - \delta)}{d(r - \delta, r + \delta)} && (5.87) \end{aligned}$$

where $d(x, y) = |x - y|$.

In the same way, we propose to search for an adapted derivative for $\mathcal{F}_{\mathcal{X},\mathcal{D},\phi}$ to define an adapted density function. For this, let us, firstly recall the norm in the L^1 space [Lusternik and Sobolev, 1974].

Definition 5.8.1. *The set $L^1(\mathcal{D})$, is the set of functions u belonging to $\mathcal{I}^{\mathcal{D}}$ such*

$$||u||_1 = \left(\int_{\mathcal{D}} |u(t)| \ dt \right) < \infty \tag{5.88}$$

where $||u||_1$ is the norm of u.

In the set $L^1(\mathcal{D})$, the distance between two functions $u, v \in L^1(\mathcal{D})$ is given by

$$d_1(u, v) = ||u - v||_1. \tag{5.89}$$

Definition 5.8.2. *Let $\mathcal{X}$ be an* frv *belonging to $\mathcal{I}^{\mathcal{D}}$, and $\mathcal{F}_{\mathcal{X},\mathcal{D}}$ its associated* fcdf.

If for a given $v \in \mathcal{I}^{\mathcal{D}}$, u_ε and w_ε are two continuous functions with a parameter $\varepsilon > 0$ such:

$$\lim_{\varepsilon \to 0} u_\varepsilon = v \ \& \ \lim_{\varepsilon \to 0} w_\varepsilon = v \tag{5.90}$$

$$\forall \varepsilon \in I\!R, \ \varepsilon > 0 \ \ u_\varepsilon \leq_{\mathcal{D}} v \ \& \ v \leq_{\mathcal{D}} w_\varepsilon \tag{5.91}$$

then we define the functional density of $\mathcal{X}$ surrounded by $\{u_\varepsilon\}$ and $\{w_\varepsilon\}$ *by*

$$f_{\mathcal{X},\mathcal{D},\{u_\varepsilon\},\{w_\varepsilon\}}(v) = \lim_{\varepsilon \to 0} \frac{P[u_\varepsilon \leq_{\mathcal{D}} \mathcal{X} \leq_{\mathcal{D}} w_\varepsilon]}{d_1(u_\varepsilon, w_\varepsilon)} \tag{5.92}$$

$$= \lim_{\varepsilon \to 0} \frac{\mathcal{F}_{\mathcal{X},\mathcal{D}}(w_\varepsilon) - \mathcal{F}_{\mathcal{X},\mathcal{D}}(u_\varepsilon)}{d_1(u_\varepsilon, w_\varepsilon)}. \tag{5.93}$$

The proposition 5.7.2 and the condition expressed by (5.91) imply that the above expression (5.93) is always positive or equal to zero, but,of course, for a single function u there is an infinity of possible sequences $\{u_\varepsilon\}$ and $\{w_\varepsilon\}$.

We have already seen how, the functional quantile is a central notion in the building of our new type of distribution. Then we propose to build the two required sequences u_ε and w_ε using the functions $Q_{\mathcal{X},\mathcal{D}}$.

Lemma 5.8.1. *Let $\mathcal{X}$ be an* frv *belonging to $\mathcal{I}^{\mathcal{D}}$, $\mathbf{G}_{\mathcal{X},\mathcal{D}}$ its surface of margins and $\mathcal{F}_{\mathcal{X},\mathcal{D}}$ its associated* fcdf.

If, for a given $v \in L^1(\mathcal{D})$, we define

$$m_v = \min_{t \in \mathcal{D}} \mathbf{G}_{\mathcal{X},\mathcal{D}}[t; v] \tag{5.94}$$

and

$$M_v = \max_{t \in \mathcal{D}} \mathbf{G}_{\mathcal{X},\mathcal{D}}[t; v] \tag{5.95}$$

and then for any numbers q and r of $[0, 1]$ such

$$0 < p \leq m_v \tag{5.96}$$

$$M_v \leq r < 1 \tag{5.97}$$

and, if we define the two sequences

$$u_\varepsilon = (1-\varepsilon)\cdot v + \varepsilon \cdot Q_{\mathcal{X},\mathcal{D},p} \tag{5.98}$$

$$w_\varepsilon = (1-\varepsilon)\cdot v + \varepsilon \cdot Q_{\mathcal{X},\mathcal{D},r}. \tag{5.99}$$

Then $\forall\ 0 < \varepsilon < 1$, $u_\varepsilon, w_\varepsilon \in L^1(\mathcal{D})$, *and*

$$f_{\mathcal{X},\mathcal{D},\{u_\varepsilon\},\{w_\varepsilon\}}(v) = \lim_{\varepsilon\to 0} \frac{\mathcal{F}_{\mathcal{X},\mathcal{D}}(v+\varepsilon\Delta_r) - \mathcal{F}_{\mathcal{X},\mathcal{D}}(v)}{d_1(u_\varepsilon, w_\varepsilon)} \tag{5.100}$$

$$+ \lim_{\varepsilon\to 0} \frac{\mathcal{F}_{\mathcal{X},\mathcal{D}}(v) - \mathcal{F}_{\mathcal{X},\mathcal{D}}(v-\varepsilon\Delta_p)}{d_1(u_\varepsilon, w_\varepsilon)} \tag{5.101}$$

where

$$\Delta_p = v - Q_{\mathcal{X},\mathcal{D},p} \tag{5.102}$$

$$\Delta_r = Q_{\mathcal{X},\mathcal{D},r} - v \tag{5.103}$$

and

$$d_1(u_\varepsilon, w_\varepsilon) = ||Q_{\mathcal{X},\mathcal{D},r} - Q_{\mathcal{X},\mathcal{D},p}||_1. \tag{5.104}$$

Proof. Firstly, to show that $u_\varepsilon, w_\varepsilon \in L^1(\mathcal{D})$, we have to recall that $L^1(\mathcal{D})$ is a vector space, then if v, $Q_{\mathcal{X},\mathcal{D},p}$ and $Q_{\mathcal{X},\mathcal{D},r}$ belong to $L^1(\mathcal{D})$, then it is also true for the two sequences. Then it remains to prove that $Q_{\mathcal{X},\mathcal{D},p}$ and $Q_{\mathcal{X},\mathcal{D},r}$ belong to $L^1(\mathcal{D})$. And, like $m_v \leq M_v$, then like $0 < r < 1$, i.e. $0 < \mathbf{G}_{\mathcal{X},\mathcal{D}}[t; Q_{\mathcal{X},\mathcal{D},r}] < 1$, thus, like $\mathbf{G}_{\mathcal{X},\mathcal{D}}$ is continuous, we must have $-\infty < Q_{\mathcal{X},\mathcal{D},r}(t) < \infty$, then $||Q_{\mathcal{X},\mathcal{D},r}||_1 < \infty$. The same reasoning can be held for $Q_{\mathcal{X},\mathcal{D},p}$.

Let us verify that u_ε and w_ε fulfill the conditions expressed by the definition 5.8.2. By expressions (5.95) and (5.97) we have

$$\begin{aligned} Q_{\mathcal{X},\mathcal{D},p} \leq_{\mathcal{D}} v &\Leftrightarrow Q_{\mathcal{X},\mathcal{D},p} - v \leq_{\mathcal{D}} 0 \\ &\Leftrightarrow \varepsilon\cdot Q_{\mathcal{X},\mathcal{D},p} - \varepsilon\cdot v \leq_{\mathcal{D}} 0 \\ &\Leftrightarrow \varepsilon\cdot Q_{\mathcal{X},\mathcal{D},p} - \varepsilon\cdot v + v \leq_{\mathcal{D}} v \\ &\Leftrightarrow \varepsilon\cdot Q_{\mathcal{X},\mathcal{D},p} + (1-\varepsilon)\cdot v \leq_{\mathcal{D}} v \\ &\Leftrightarrow u_\varepsilon \leq_{\mathcal{D}} v. \end{aligned}$$

Obviously $\lim_{\varepsilon\to 0} u_\varepsilon = v$, and, the same reasoning can be done for the sequence w_ε. And both sequences can be re-written as follow

$$u_\varepsilon = v + \varepsilon \underbrace{(Q_{\mathcal{X},\mathcal{D},p} - v)}_{\leq_{\mathcal{D}} 0} \tag{5.105}$$

$$= v - \varepsilon \underbrace{(v - Q_{\mathcal{X},\mathcal{D},p})}_{\geq_{\mathcal{D}} 0} \tag{5.106}$$

$$= v - \varepsilon \underbrace{\Delta_p}_{\geq_{\mathcal{D}} 0} \tag{5.107}$$

$$w_\varepsilon = v + \varepsilon \underbrace{(Q_{\mathcal{X},\mathcal{D},r} - v)}_{\geq_{\mathcal{D}} 0} \tag{5.108}$$

$$= v + \varepsilon \underbrace{\Delta_r}_{\geq_{\mathcal{D}} 0}. \tag{5.109}$$

Then *functional density surrounded by* $\{u_\varepsilon\}$ *and* $\{w_\varepsilon\}$ become

$$\begin{aligned} f_{\mathcal{X},\mathcal{D},\{u_\varepsilon\},\{w_\varepsilon\}}(v) &= \lim_{\varepsilon\to 0} \frac{\mathcal{F}_{\mathcal{X},\mathcal{D}}(w_\varepsilon) - \mathcal{F}_{\mathcal{X},\mathcal{D}}(u_\varepsilon)}{d_1(u_\varepsilon, w_\varepsilon)} \\ &= \lim_{\varepsilon\to 0} \frac{\mathcal{F}_{\mathcal{X},\mathcal{D}}(w_\varepsilon) - \mathcal{F}_{\mathcal{X},\mathcal{D}}(v) + \mathcal{F}_{\mathcal{X},\mathcal{D}}(v) - \mathcal{F}_{\mathcal{X},\mathcal{D}}(u_\varepsilon)}{d_1(u_\varepsilon, w_\varepsilon)} \\ &= \lim_{\varepsilon\to 0} \frac{\mathcal{F}_{\mathcal{X},\mathcal{D}}(w_\varepsilon) - \mathcal{F}_{\mathcal{X},\mathcal{D}}(v)}{d_1(u_\varepsilon, w_\varepsilon)} \\ &\quad + \lim_{\varepsilon\to 0} \frac{\mathcal{F}_{\mathcal{X},\mathcal{D}}(v) - \mathcal{F}_{\mathcal{X},\mathcal{D}}(u_\varepsilon)}{d_1(u_\varepsilon, w_\varepsilon)} \\ &= \lim_{\varepsilon\to 0} \frac{\mathcal{F}_{\mathcal{X},\mathcal{D}}(v + \varepsilon\Delta_r) - \mathcal{F}_{\mathcal{X},\mathcal{D}}(v)}{d_1(u_\varepsilon, w_\varepsilon)} \qquad (5.110) \\ &\quad + \lim_{\varepsilon\to 0} \frac{\mathcal{F}_{\mathcal{X},\mathcal{D}}(v) - \mathcal{F}_{\mathcal{X},\mathcal{D}}(v - \varepsilon\Delta_p)}{d_1(u_\varepsilon, w_\varepsilon)} \qquad (5.111) \end{aligned}$$

where the distance between u_ε and w_ε is equal to

$$\begin{aligned} d_1(u_\varepsilon, w_\varepsilon) &= ||u_\varepsilon - w_\varepsilon||_1 \\ &= \int_{\mathcal{D}} |u_\varepsilon(t) - w_\varepsilon(t)|\, dt \\ &= \varepsilon \left(\int_{\mathcal{D}} |\Delta_p(t) + \Delta_r(t)|\, dt \right) \\ &= \varepsilon \left(\int_{\mathcal{D}} |Q_{\mathcal{X},\mathcal{D},r}(t) - Q_{\mathcal{X},\mathcal{D},p}(t)|\, dt \right) \\ &= \varepsilon ||Q_{\mathcal{X},\mathcal{D},r} - Q_{\mathcal{X},\mathcal{D},p}||_1. \qquad (5.112) \end{aligned}$$

□

Expressions (5.100) and (5.101) can be seen as the directional derivatives of $\mathcal{F}_{\mathcal{X},\mathcal{D},\phi}$ computed in v with, respectively, the directions Δ_u and Δ_w.

The concept of the derivative of a functional F computed in a function u, in direction of another function h was introduced in functional analysis by René Gâteaux in two posthumous papers: [Gâteaux, 1919a] and [Gâteaux, 1922], and is a generalization of directional derivative (see also [Atkinson and Han, 2001]).

Definition 5.8.3. *Suppose V and W are normed vector spaces, and F an operator from V to W. The* Gâteaux differential *$DF(u; s)$ of F at u in the*

direction $h \in V$ is given by:

$$DF(u;s) = \lim_{\epsilon \to 0} \frac{F(u+\epsilon \cdot h) - F(u)}{\epsilon} \tag{5.113}$$

$$= F'(u) \cdot s \tag{5.114}$$

If (5.113) exists $\forall h \in V$ then F is Gâteaux differentiable *and the map $F'(u)$ is the* Gâteaux derivative *of F at u.*

The following proposition is obvious.

Proposition 5.8.2. *Let $\mathcal{X}$ be an* frv*, $\mathcal{F}_{\mathcal{X},\mathcal{D}}$ its* fcdf*, v a function of $L^1(\mathcal{D})$, $\{u_\varepsilon\}$ and $\{w_\varepsilon\}$ two continuously indexed sequences of functions defined by expressions* (5.98) *and* (5.99)*.*

If $\mathcal{F}_{\mathcal{X},\mathcal{D}}$ is continuous, then, the functional density surrounded by $\{u_\varepsilon\}$ and $\{w_\varepsilon\}$ *is equal to*

$$f_{\mathcal{X},\mathcal{D},\{u_\varepsilon\},\{w_\varepsilon\}}(v) = \frac{D\mathcal{F}_{\mathcal{X},\mathcal{D}}(v;\Delta_r) + D\mathcal{F}_{\mathcal{X},\mathcal{D}}(v;\Delta_p)}{||Q_{\mathcal{X},\mathcal{D},r} - Q_{\mathcal{X},\mathcal{D},p}||_1} \tag{5.115}$$

where Δ_p and Δ_r are given by expressions (5.102) *and* (5.103)*.*

To give the *Gâteaux density* for the *QAMML* distribution we need a result from functional analysis [Lusternik and Sobolev, 1974].

Proposition 5.8.3. *Let the following integral transform:*

$$\mathcal{T}(f) = \int_a^b K(t,s) \cdot g[s, f(s)] \; ds \tag{5.116}$$

where the kernel $K(s,t)$ is continuous on $[a,b]^2$, and $g(s,t)$ is a function of two variables, defined and continuous on $[a,b] \times]-\infty, +\infty[$. Then for any function $h \in C[a,b]$ we have

$$D\mathcal{T}(f,h) = \int_a^b K(t,s) \cdot g'_v[s, f(s)] \cdot h(s) \; ds \tag{5.117}$$

Where $D\mathcal{T}(f,h)$ is the Gâteaux differential *of $\mathcal{F}$ at f in the direction h.*

Theorem 5.8.4. *Let $\mathcal{X}$ be an* frv*, $\mathcal{F}_{\mathcal{X},\mathcal{D},\phi}$ its QAMML* fcdf*, v a function of $L^1(\mathcal{D})$. If $h \in L^1(\mathcal{D})$, then the* Gâteaux differential of $\mathcal{F}_{\mathcal{X},\mathcal{D},\phi}$ in v and in direction of h *is given by:*

$$D\mathcal{F}_{\mathcal{X},\mathcal{D},\phi}(v;h) = \psi'\left[\frac{1}{|\mathcal{D}|}\int_{\mathcal{D}} \phi\left(\mathbf{G}_{\mathcal{X},\mathcal{D}}[t;v]\right) \; dt\right] \cdot \left\{\frac{1}{|\mathcal{D}|}\int_{\mathcal{D}} \phi'\left(\mathbf{G}_{\mathcal{X},\mathcal{D}}[t;v]\right) \mathbf{g}_{\mathcal{X},\mathcal{D}}[t;v]\, h(t) \; dt\right\} \tag{5.118}$$

Proof. It is sufficient to use (5.117) with $K(t,s) = \frac{1}{|\mathcal{D}|}$ and $g(s,t) = \phi(\mathbf{G}_{\mathcal{X},\mathcal{D}}(s,t))$ then

$$\mathcal{F}_{\mathcal{X},\mathcal{D},\phi}(v) = \psi[\mathcal{T}(u)] = (\psi \circ \mathcal{T})(u)$$

with the following composition schema

$$\mathcal{F}_{\mathcal{X},\mathcal{D},\phi} : L^1(\mathcal{D}) \xrightarrow{\mathcal{T}} \mathbb{R} \xrightarrow{\psi} [0,1]$$

thus

$$D\mathcal{F}_{\mathcal{X},\mathcal{D},\phi}(v;h) = \psi[\mathcal{T}(v)]' \cdot D\mathcal{T}(v;h).$$

□

The *Gâteaux differential* is not always linear, but the theorem 5.8.4 implies the linearity of $D\mathcal{F}_{\mathcal{X},\mathcal{D},\phi}(v;h)$.

Corollary 5.8.5. *Let $\mathcal{X}$ be an* frv*, $\mathcal{F}_{\mathcal{X},\mathcal{D},\phi}$ its QAMML* fcdf*, then* Gâteaux differential *of a $\mathcal{F}_{\mathcal{X},\mathcal{D},\phi}$ is linear for the direction h.*

Proof. Let $v,h,l \in L^1(\mathcal{D})$ and $\alpha,\beta \in \mathbb{R}$, then

$$\begin{aligned}
D\mathcal{F}_{\mathcal{X},\mathcal{D},\phi}(v;\alpha h+\beta l) &= \psi'\left[\frac{1}{|\mathcal{D}|}\int_{\mathcal{D}} \phi(\mathbf{G}_{\mathcal{X},\mathcal{D}}[t;v])\ dt\right] \\
&\quad \left\{\frac{1}{|\mathcal{D}|}\int_{\mathcal{D}} \phi'(\mathbf{G}_{\mathcal{X},\mathcal{D}}[t;v])\,\mathbf{g}_{\mathcal{X},\mathcal{D}}[t;u]\,(\alpha h+\beta l)(t)\ dt\right\} \\
&= \psi'\left[\frac{1}{|\mathcal{D}|}\int_{\mathcal{D}} \phi(\mathbf{G}_{\mathcal{X},\mathcal{D}}[t;v])\ dt\right] \\
&\quad \cdot\left\{\frac{\alpha}{|\mathcal{D}|}\int_{\mathcal{D}} \phi'(\mathbf{G}_{\mathcal{X},\mathcal{D}}[t;v])\,\mathbf{g}_{\mathcal{X},\mathcal{D}}[t;u]\,h(t)\ dt\right. \\
&\quad \left.+\frac{\beta}{|\mathcal{D}|}\int_{\mathcal{D}} \phi'(\mathbf{G}_{\mathcal{X},\mathcal{D}}[t;v])\,\mathbf{g}_{\mathcal{X},\mathcal{D}}[t;u]\,l(t)\ dt\right\} \\
&= \psi'\left[\frac{1}{|\mathcal{D}|}\int_{\mathcal{D}} \phi(\mathbf{G}_{\mathcal{X},\mathcal{D}}[t;v])\ dt\right] \\
&\quad \cdot\frac{\alpha}{|\mathcal{D}|}\int_{\mathcal{D}} \phi'(\mathbf{G}_{\mathcal{X},\mathcal{D}}[t;v])\,\mathbf{g}_{\mathcal{X},\mathcal{D}}[t;u]\,h(t)\ dt \\
&\quad +\psi'\left[\frac{1}{|\mathcal{D}|}\int_{\mathcal{D}} \phi(\mathbf{G}_{\mathcal{X},\mathcal{D}}[t;v])\ dt\right] \\
&\quad \frac{\beta}{|\mathcal{D}|}\int_{\mathcal{D}} \phi'(\mathbf{G}_{\mathcal{X},\mathcal{D}}[t;v])\,\mathbf{g}_{\mathcal{X},\mathcal{D}}[t;u]\,l(t)\ dt \\
&= \alpha D\mathcal{F}_{\mathcal{X},\mathcal{D},\phi}(v;h) + \beta D\mathcal{F}_{\mathcal{X},\mathcal{D},\phi}(v;l)
\end{aligned}$$

□

Now let us suppose that the surface of margins $G_{\mathcal{X},\mathcal{D}}$ is a known theoretical distribution F with a scale parameter s and a location parameter[3] l [NIST/SEMATECH, 2008], then we can always express F, its density f and its quantile function Q using the standard form, i.e. when $l = 0$ and $s = 1$:

$$F(y; l, s) = F\left(\frac{y-l}{s}; 0, 1\right) \tag{5.119}$$

$$f(y; l, s) = \frac{1}{s} \cdot f\left(\frac{y-l}{s}; 0, 1\right) \tag{5.120}$$

$$Q(\alpha; l, s) = s \cdot Q(\alpha; 0, 1) + l. \tag{5.121}$$

Table 5.3 show six well known distributions with their location and scale parameters.

In this case the surface of margins $G_{\mathcal{X},\mathcal{D}}$, the surface of densities $g_{\mathcal{X},\mathcal{D}}$ and the functional quantiles $Q_{\mathcal{X},\mathcal{D}}$ can be written using the standard form of the functions F, f and Q:

$$\begin{aligned} G_{\mathcal{X},\mathcal{D}}(t, y) &= F(y; l(t), s(t)) && (5.122)\\ &= F\left(\frac{y-l(t)}{s(t)}; 0, 1\right) && (5.123) \end{aligned}$$

$$\begin{aligned} g_{\mathcal{X},\mathcal{D}}(t, y) &= f(y; l(t), s(t)) && (5.124)\\ &= \frac{1}{s(t)} f\left(\frac{y-l(t)}{s(t)}; 0, 1\right) && (5.125) \end{aligned}$$

$$\begin{aligned} Q_{\mathcal{X},\mathcal{D},\alpha}(t) &= Q(\alpha; s(t), l(t)) && (5.126)\\ &= s(t) \cdot Q(\alpha; 0, 1) + l(t) && (5.127) \end{aligned}$$

where $s, l \in \mathcal{I}^{\mathcal{D}}$ are the two functional parameters which give the location and the scale for any $t \in \mathcal{D}$.

Then the *functional density* defined by expression (5.93) is independent of the randomly chosen values p and r.

Theorem 5.8.6. *Let $\mathcal{X}$ be an* frv*, $\mathcal{F}_{\mathcal{X},\mathcal{D},\phi}$ its QAMML* fcdf*, v a function of $L^1(\mathcal{D})$.*

If, $\mathbf{G}_{\mathcal{X},\mathcal{D}}[t; .]$, the surface of margins of $\mathcal{X}$ is a continuous and strictly increasing distribution $\forall t \in \mathcal{D}$ and has a location parameter and a scale parameter, given $\forall t \in \mathcal{D}$ respectively by l and s both in $L^1(\mathcal{D})$, then for any continuously indexed sequences $\{u_\varepsilon\}$ and $\{w_\varepsilon\}$ fulfilling expressions (5.98) *and* (5.99) *the* functional density surrounded by $\{u_\varepsilon\}$ and $\{w_\varepsilon\}$ *is unique*

[3]The location parameter is not necessarily the mean of the associate *rrv*, but more a translation parameter, which permits to "shift" the distribution along the real line.

Name	density: $f(t)$	l	s
Normal	$f(t) = \frac{\exp\left(-\frac{(t-\mu)^2}{2\sigma^2}\right)}{\sigma\sqrt{2\pi}}$	μ	σ
Cauchy	$f(t) = \frac{1}{s\pi(1+((t-l)/s)^2)}$	l	s
Exponential[a]	$f(t) = \frac{1}{\beta}\exp\left(-\frac{t-\mu}{\beta}\right)$	μ	β
Weibull[b]	$f(t) = \frac{\gamma}{\alpha}\left(\frac{t-\mu}{\alpha}\right)^{\gamma-1}\exp-\left(\left(\frac{t-\mu}{\alpha}\right)^{\gamma}\right)$	μ	α
Gamma[c]	$f(t) = \frac{\left(\frac{t-\mu}{\beta}\right)^{\gamma-1}\exp\left(-\frac{t-\mu}{\beta}\right)}{\beta\Gamma(\gamma))}$	μ	β
Beta[d]	$f(t) = \frac{(t-a)^{p-1}(b-t)^{q-1}}{B(p,q)(b-a)^{p+q-1}}$	a	$b-a$

[a] for $t \geq \mu; \beta > 0$
[b] for $t \geq \mu; \alpha > 0$
[c] $t \geq \mu; \gamma, \beta > 0$ and Γ is the gamma function: $\Gamma(a) = \int_0^\infty t^{a-1}e^{-t}dt$.
[d] for $a \leq t \leq b; p, q > 0$, and $B(p,q)$ is the beta function: $B(p,q) = \int_0^1 t^{p-1}(1-t)^{q-1}dt$.

Table 5.3: Densities of theoretical distributions with their location l and scale s parameters.

(independent of the value p and r) and then simply called the Gâteaux density *and denoted* $f_{\mathcal{X},\mathcal{D},\phi}(v)$*:*

$$f_{\mathcal{X},\mathcal{D},\phi}(v) = \frac{D\mathcal{F}_{\mathcal{X},\mathcal{D},\phi}(v;s)}{||s||_1}. \tag{5.128}$$

Proof. If $G_{\mathcal{X},\mathcal{D}}$ is a distribution with a location parameter and a scale parameter, then, in this case $\forall t \in \mathcal{D}$ using (5.127), we have the following relation for any $\alpha, \beta \in [0,1]$:

$$Q_{\mathcal{X},\mathcal{D},\alpha}(t) - Q_{\mathcal{X},\mathcal{D},\beta}(t) = s(t)(Q(\alpha;0,1) - Q(\beta;0,1)) \tag{5.129}$$

and thus

$$Q_{\mathcal{X},\mathcal{D},\alpha}(t) = Q_{\mathcal{X},\mathcal{D},\beta}(t) + s(t)K_{\alpha,\beta} \tag{5.130}$$

where $K_{\alpha,\beta} = Q(\alpha;0,1) - Q(\beta;0,1)$ is a constant which depends only of α and β and not of the value of t. The sign of $K_{\alpha,\beta}$ depends of the order between α and β.

If we denotes q for the arithmetic means of $\mathbf{G}_{\mathcal{X},\mathcal{D}}[t;u]$ over $\mathcal{D}$:

$$q = \frac{1}{|\mathcal{D}|}\int_{\mathcal{D}} \mathbf{G}_{\mathcal{X},\mathcal{D}}[t;u]dt \tag{5.131}$$

and if m_v, M_v, p and q are defined by expressions (5.94), (5.96), (5.97) and (5.95), then by proposition 5.5.9 we have

$$0 < p \leq m_v \leq q \leq M_v \leq r < 1 \tag{5.132}$$

and thus

$$Q_{\mathcal{X},\mathcal{D},p} \leq_{\mathcal{D}} Q_{\mathcal{X},\mathcal{D},q} \leq_{\mathcal{D}} Q_{\mathcal{X},\mathcal{D},r}. \tag{5.133}$$

Then using the relation 5.129 we have

$$Q_{\mathcal{X},\mathcal{D},q}(t) - Q_{\mathcal{X},\mathcal{D},p}(t) = s(t)\underbrace{K_{q,p}}_{\geq_{\mathcal{D}}0} \tag{5.134}$$

$$Q_{\mathcal{X},\mathcal{D},r}(t) - Q_{\mathcal{X},\mathcal{D},q}(t) = s(t)\underbrace{K_{r,q}}_{\geq_{\mathcal{D}}0} \tag{5.135}$$

and thus

$$Q_{\mathcal{X},\mathcal{D},p}(t) = Q_{\mathcal{X},\mathcal{D},q}(t) - s(t)\underbrace{K_{q,p}}_{\geq_{\mathcal{D}}0}. \tag{5.136}$$

$$Q_{\mathcal{X},\mathcal{D},r}(t) = Q_{\mathcal{X},\mathcal{D},q}(t) + s(t)\underbrace{K_{r,q}}_{\geq_{\mathcal{D}}0}. \tag{5.137}$$

Then the functions Δ_p and Δ_r given by expressions (5.102) and (5.103), can be written as follow:

$$\Delta_p = v + Q_{\mathcal{X},\mathcal{D},q} - sK_{q,p} \tag{5.138}$$

$$\Delta_r = Q_{\mathcal{X},\mathcal{D},q} + sK_{r,q} - v. \tag{5.139}$$

Let us remember that, by the proposition 5.8.2 the *functional density of* $\mathcal{X}$ *surrounded by* $\{u_\varepsilon\}$ *and* $\{w_\varepsilon\}$ can be written (expression (5.115)):

$$f_{\mathcal{X},\mathcal{D},\{u_\varepsilon\},\{w_\varepsilon\}}(v) = \frac{D\mathcal{F}_{\mathcal{X},\mathcal{D},\phi}(v;\Delta_r) + D\mathcal{F}_{\mathcal{X},\mathcal{D},\phi}(v;\Delta_p)}{||Q_{\mathcal{X},\mathcal{D},r} - Q_{\mathcal{X},\mathcal{D},p}||_1}.$$

But, by the corollary 5.8.5 we know that $D\mathcal{F}_{\mathcal{X},\mathcal{D},\phi}$ is linear for the direction, and then we can write

$$D\mathcal{F}_{\mathcal{X},\mathcal{D},\phi}(v;\Delta_p) = D\mathcal{F}_{\mathcal{X},\mathcal{D},\phi}(v;v) - D\mathcal{F}_{\mathcal{X},\mathcal{D},\phi}(v;Q_{\mathcal{X},\mathcal{D},q}) - K_{q,p}D\mathcal{F}_{\mathcal{X},\mathcal{D},\phi}(v;s)$$

$$D\mathcal{F}_{\mathcal{X},\mathcal{D},\phi}(v;\Delta_r) = D\mathcal{F}_{\mathcal{X},\mathcal{D},\phi}(v;Q_{\mathcal{X},\mathcal{D},q}) + K_{r,q}D\mathcal{F}_{\mathcal{X},\mathcal{D},\phi}(v;s) - D\mathcal{F}_{\mathcal{X},\mathcal{D},\phi}(v;v)$$

and then

$$D\mathcal{F}_{\mathcal{X},\mathcal{D},\phi}(v;\Delta_p) + D\mathcal{F}_{\mathcal{X},\mathcal{D},\phi}(v;\Delta_p) = (K_{r,q} + K_{q,p})D\mathcal{F}_{\mathcal{X},\mathcal{D},\phi}(v;s).$$

Then it remains to compute the denominator of the expression (5.115):

$$\begin{aligned} ||Q_{\mathcal{X},\mathcal{D},r} - Q_{\mathcal{X},\mathcal{D},p}||_1 &= \int_{\mathcal{D}} |Q_{\mathcal{X},\mathcal{D},r}(t) - Q_{\mathcal{X},\mathcal{D},p}(t)|\, dt \\ &= \int_{\mathcal{D}} |(K_{r,q} + K_{q,p})s(t)|\, dt \\ &= (K_{r,q} + K_{q,p})||s||_1 \end{aligned}$$

and then

$$f_{\mathcal{X},\mathcal{D},\phi}(v) = \frac{D\mathcal{F}_{\mathcal{X},\mathcal{D},\phi}(v;s)}{||s||_1}.$$

□

Thus, when the surface of margins is given by probability distribution with location and scale parameters, then the Gâteaux density in $v \in L^1(\mathcal{D})$, is given by the Gâteaux differential of $\mathcal{F}_{\mathcal{X},\mathcal{D},\phi}$ computed in v, in direction of the scale parameter $s \in L^1(\mathcal{D})$, is unique and depends only of v and s.

For the sake of illustration we give in the top part of figures 5.30, 5.31, 5.32 and 5.33, the graph of v and $v \pm \varepsilon\sigma$(for a constant $v(t) = 3,\ \forall t$), expressions used in the Gâteaux differential to define the Gâteau density. We show also the decile functions of the Tecator data (see 2.11). The bottom parts show the direction used: the standard deviation of the Tecator data.

If the surface of margins is given by probability distribution with location and scale parameters,then, similarly the fact that the QAMML *fcdf* in a functional quantile $Q_{\mathcal{X},\mathcal{D},p}$ is always equal to p, we can show that the Gâteaux density of this *fcdf* is always known for functional quantiles.

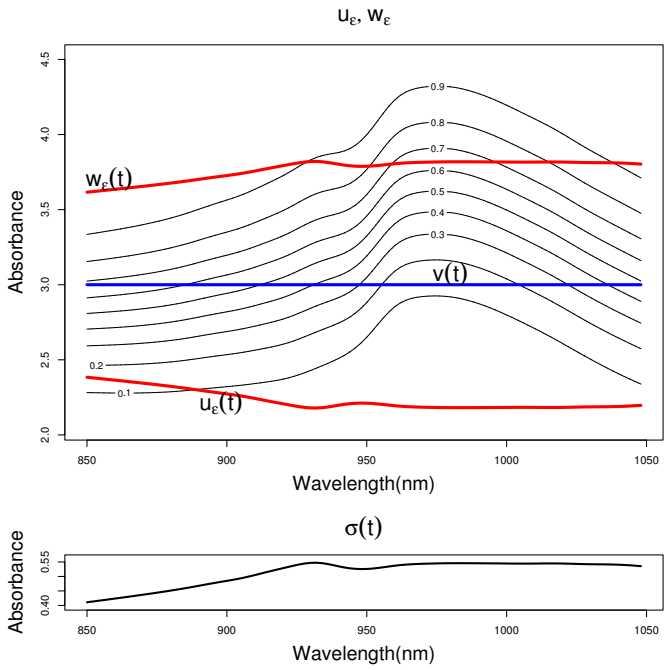


Figure 5.30: Inside the Gâteaux differential: graph of $v \pm \varepsilon\sigma$, with $\varepsilon = 1.5$.

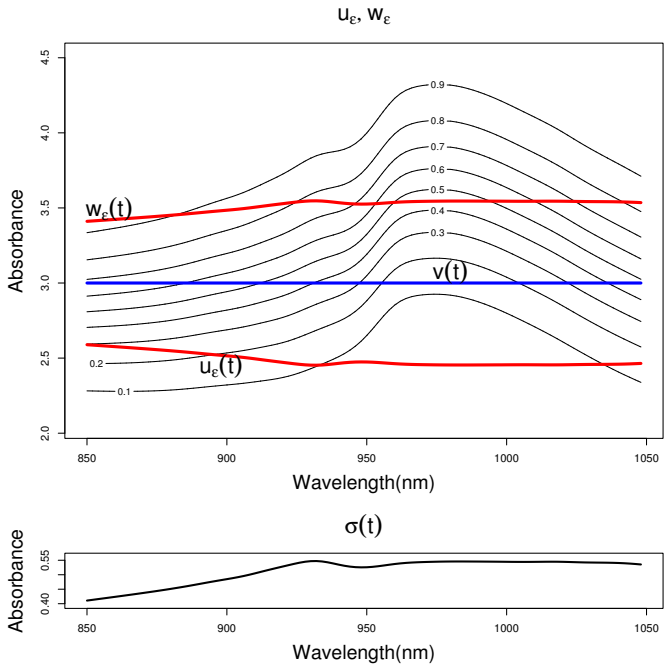


Figure 5.31: Inside the Gâteaux differential: graph of $u \pm \varepsilon\sigma$, with $\varepsilon = 1$.

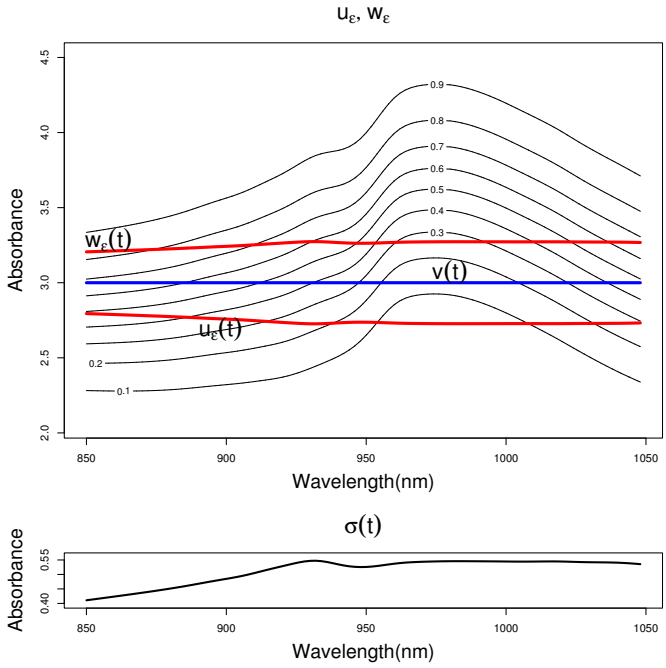


Figure 5.32: Inside the Gâteaux differential: graph of $u \pm \varepsilon\sigma$, with $\varepsilon = 0.5$.

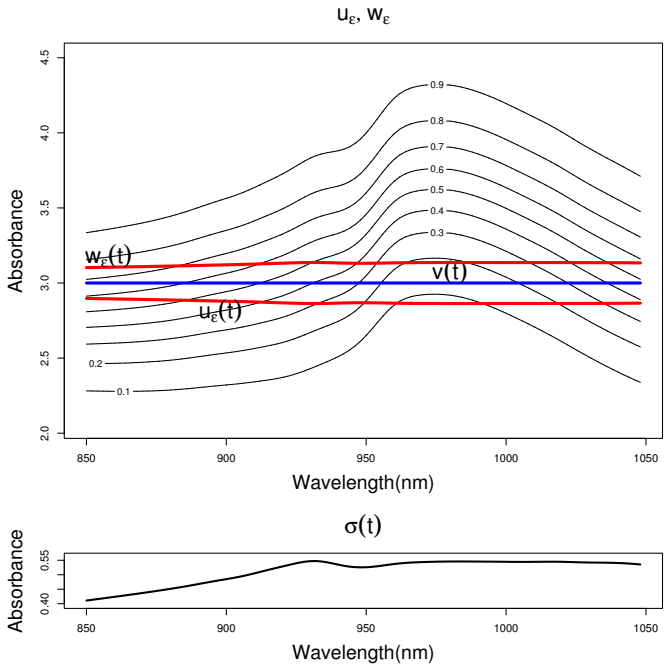


Figure 5.33: Inside the Gâteaux differential: graph of $u \pm \varepsilon\sigma$, with $\varepsilon = 0.25$.

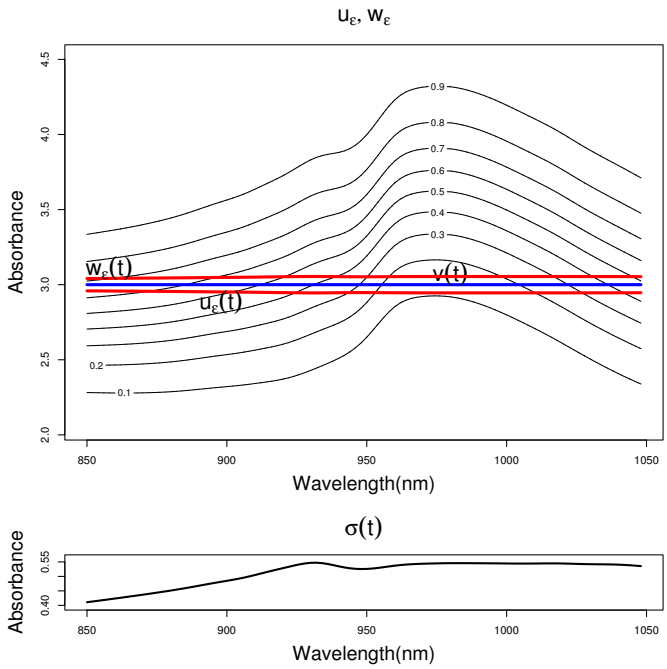


Figure 5.34: Inside the Gâteaux differential: graph of $u \pm \varepsilon\sigma$, with $\varepsilon = 0.1$.

Lemma 5.8.7. *Let $\mathcal{X}$ be an* frv*, $\mathcal{F}_{\mathcal{X},\mathcal{D},\phi}$ its QAMML* fcdf*. If, $G_{\mathcal{X},\mathcal{D}}$, the surface of margins of $\mathcal{X}$ is continuous $\forall t \in \mathcal{D}$ and is given be the expression* (5.123)*, and then has a location parameter and a scale parameter, given $\forall t \in \mathcal{D}$ respectively by l and s both in $L^1(\mathcal{D})$, and if $Q_{\mathcal{X},\mathcal{D},p} \in L^1(\mathcal{D})$ is a* functional quantile *of order p then $\forall t \in \mathcal{D}$*

$$g_{\mathcal{X},\mathcal{D}}\left[t; Q_{\mathcal{X},\mathcal{D},p}\right] \cdot s(t) = \pi \tag{5.140}$$

with

$$\pi = f\left(Q(p;0,1);0,1\right) \tag{5.141}$$

where f and Q are given by expressions (5.125) *and* (5.127)*.*

Proof. By hypothesis, and using relations (5.125) and (5.127), we can write

$$\begin{aligned}
\mathbf{g}_{\mathcal{X},\mathcal{D}}[t; Q_{\mathcal{X},\mathcal{D},p}] &= f\left(Q_{\mathcal{X},\mathcal{D},p}(t); l(t), s(t)\right) \\
&= \frac{1}{s(t)} f\left(\frac{Q_{\mathcal{X},\mathcal{D},p}(t) - l(t)}{s(t)}; 0, 1\right) \\
&= \frac{1}{s(t)} f\left(\frac{Q\left(p; l(t), s(t)\right) - l(t)}{s(t)}; 0, 1\right) \\
&= \frac{1}{s(t)} f\left(\frac{s(t)Q(p;0,1) + l(t) - l(t)}{s(t)}; 0, 1\right) \\
&= \frac{1}{s(t)} f\left(Q(p;0,1);0,1\right).
\end{aligned}$$

□

Figures from 5.35 to 5.43 illustrate the lemma 5.8.7 in the normal case and give the surfaces $g_{\mathcal{X},\mathcal{D}}(t,y) * \sigma(t)$.

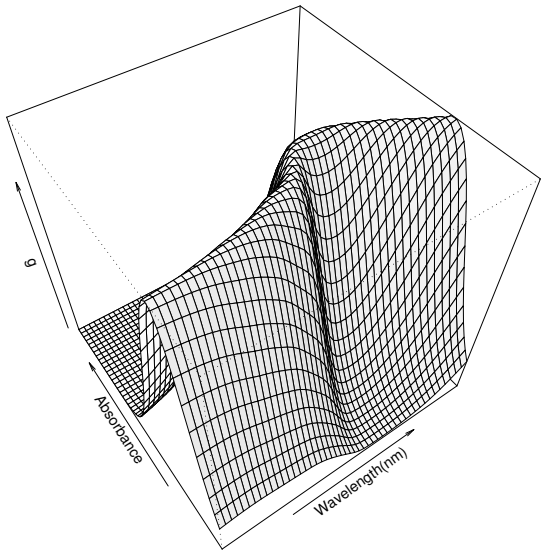


Figure 5.35: The normal surface of densities $g_{\mathcal{X},\mathcal{D}}(t,y) * \sigma(t)$ for the Tecator data (see fig. 2.11, p.21).

Proposition 5.8.8. *Let $\mathcal{X}$ be an* frv*, $\mathcal{F}_{\mathcal{X},\mathcal{D},\phi}$ its QAMML* fcdf *and $f_{\mathcal{X},\mathcal{D},\phi}$ its Gâteaux density. If, $\mathbf{G}_{\mathcal{X},\mathcal{D}}[t,.]$, the surface of margins of $\mathcal{X}$ is a continuous cdf $\forall t \in \mathcal{D}$ and has a location parameter and a scale parameter, given $\forall t \in \mathcal{D}$ respectively by l and s both in $L^1(\mathcal{D})$, and if $Q_{\mathcal{X},\mathcal{D},p} \in L^1(\mathcal{D})$ is a* functional quantile *of order p then*

$$f_{\mathcal{X},\mathcal{D},\phi}(Q_{\mathcal{X},\mathcal{D},p}) = \frac{\pi}{\|s\|_2} \tag{5.142}$$

where π is given by expression (5.141).

Proof. From the hypothesis, and using the lemma 5.8.7, it comes easily that:

$$\begin{aligned} D\mathcal{F}_{\mathcal{X},\mathcal{D},\phi}(Q_{\mathcal{X},\mathcal{D},p};s) &= \psi'\left[\frac{1}{|\mathcal{D}|}\int_{\mathcal{D}} \phi\left(\mathbf{G}_{\mathcal{X},\mathcal{D}}\left[t;Q_{\mathcal{X},\mathcal{D},p}\right]\right)\, dt\right] \\ &\quad \left\{\frac{1}{|\mathcal{D}|}\int_{\mathcal{D}} \phi'\left(\mathbf{G}_{\mathcal{X},\mathcal{D}}\left[t;Q_{\mathcal{X},\mathcal{D},p}\right]\right)\mathbf{g}_{\mathcal{X},\mathcal{D}}\left[t;Q_{\mathcal{X},\mathcal{D},p}\right]s(t)dt\right\} \\ &= \psi'\left[\frac{1}{|\mathcal{D}|}\int_{\mathcal{D}} \phi\left(p\right)\, dt\right]\left\{\frac{1}{|\mathcal{D}|}\int_{\mathcal{D}} \phi'\left(p\right)\pi dt\right\} \\ &= \psi'\left[\phi\left(p\right)\right]\cdot\left\{\phi'\left(p\right)\right)\cdot\pi\right\} \end{aligned}$$

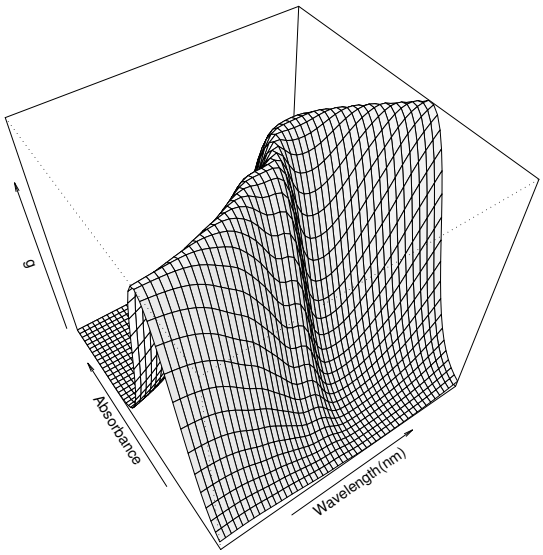


Figure 5.36: The normal surface of densities $g_{\mathcal{X},\mathcal{D}}(t,y) * \sigma(t)$ for the Tecator data with high fat content (see fig. 2.22, p.29).

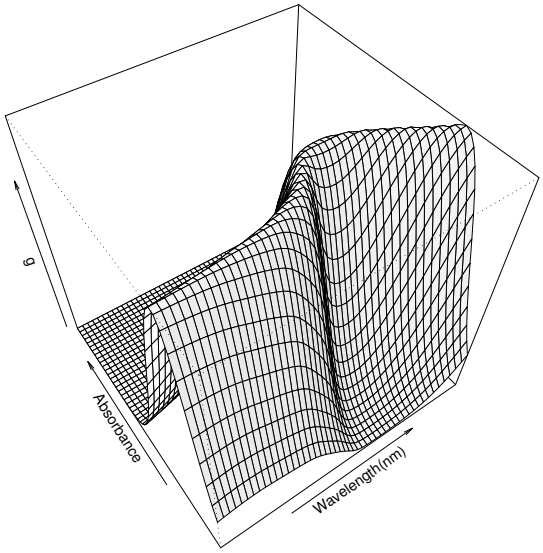


Figure 5.37: The normal surface of densities $g_{\mathcal{X},\mathcal{D}}(t,y) * \sigma(t)$ for the Tecator data with low fat content (see fig. 2.23, p.30).

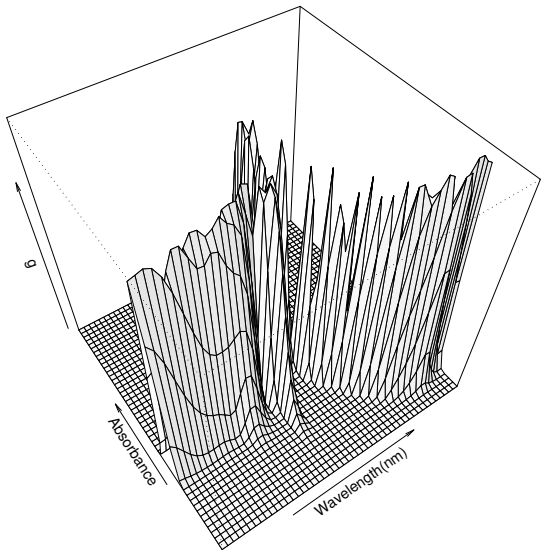


Figure 5.38: The normal surface of densities $g_{\mathcal{X},\mathcal{D}}(t,y) * \sigma(t)$ for the first derivative of the Tecator data (see fig. 2.20, p.28).

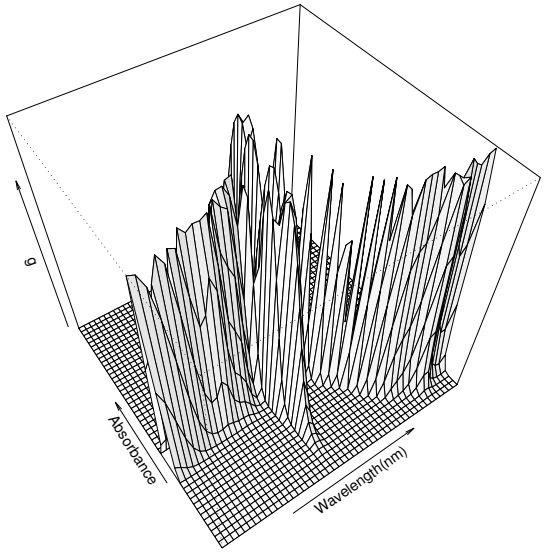


Figure 5.39: The normal surface of densities $g_{\mathcal{X},\mathcal{D}}(t,y) * \sigma(t)$ for the first derivative of the Tecator data with high fat content (see fig. 2.24, p.30).

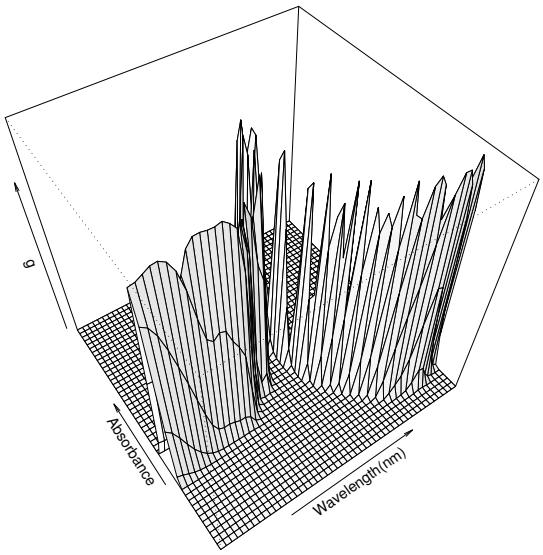


Figure 5.40: The normal surface of densities $g_{\mathcal{X},\mathcal{D}}(t,y) * \sigma(t)$ for the first derivative of the Tecator data with low fat content (see fig. 2.25, p.31).

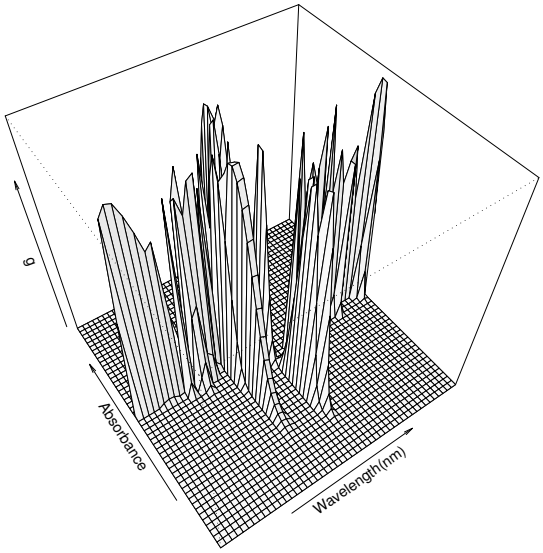


Figure 5.41: The normal surface of densities $g_{\mathcal{X},\mathcal{D}}(t,y) * \sigma(t)$ for the first derivative of the Tecator data (see fig. 2.21, p.28).

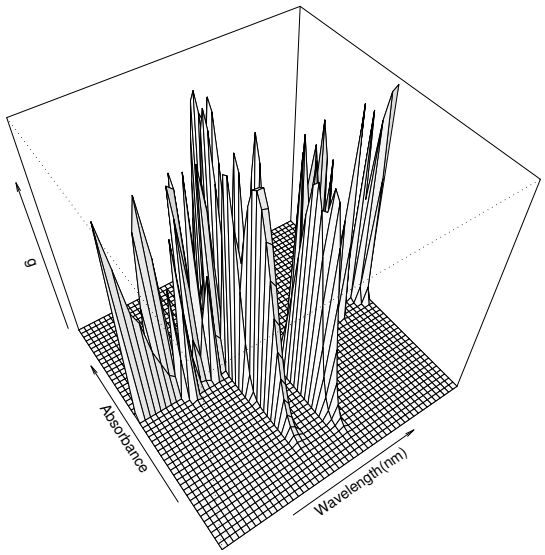


Figure 5.42: The normal surface of densities $g_{\mathcal{X},\mathcal{D}}(t,y) * \sigma(t)$ for the first derivative of the Tecator data with high fat content (see fig. 2.21, p.28).

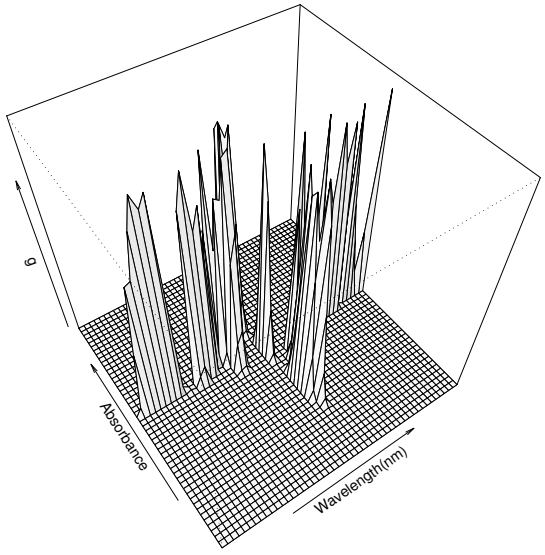


Figure 5.43: The normal surface of densities $g_{\mathcal{X},\mathcal{D}}(t,y) * \sigma(t)$ for the first derivative of the Tecator data with low fat content (see fig. 2.21, p.28).

and then

$$f_{\underline{X},\mathcal{D},\phi}(Q_p) = \frac{\psi'\left[\phi\left(p\right)\right] \cdot \phi'\left(p\right)) \cdot \pi}{\|s\|_1}$$

and like $\psi\left(\phi\left(p\right)\right) = p$, then $\psi'\left[\phi(p)\right] \cdot \phi'(p) = 1$ □

Thus, when $G_{\mathcal{X},\mathcal{D}}$ is given by a parametric distribution with location and scale parameters, the relation expressed by (5.127) implies the theorem (5.8.6), and then assures us the uniqueness of the Gâteaux density.

But when the univariate distributions of $G_{\mathcal{X},\mathcal{D}}$ are estimated in an empirical way, then we have recalled in section 4.2, how to proceed to the estimation of the density using the classical kernel estimation (see expression(4.7)), and using the empirical cumulative distribution (see expression (4.6)) for the *cdf*. In these case we not longer have a relation as in the expression (5.127), and thus no "direction" to define the Gâteaux density. We show below that, if, when using a kernel k for the density estimation, we use its primitive K, then under some conditions, the functional bandwidth h can be used for direction in the functional density, and give a Gâteaux density.

Theorem 5.8.9. *Let $\mathcal{X}$ be an* frv*, and $(\mathcal{X}_1(t), ..., \mathcal{X}_N(t))$ the sample over which the estimation is made.*

If the surface of densities is estimated by the kernel estimator $\widehat{g}_{\mathcal{X},\mathcal{D}}$ given in expression (4.7)*, and if for any $p \in [0,1]$ there exists a unique $\pi \in I\!R^+$ such $\forall t \in \mathcal{D}$*

$$\widehat{g}_{\mathcal{X},\mathcal{D}}[t; Q_{\mathcal{X},\mathcal{D},p}] \cdot h(t) = \pi \tag{5.143}$$

where $h \in L^1(\mathcal{D})$ is the functional bandwidth, and moreover, if the surface of margins is estimated as follow

$$\widehat{G}_{\mathcal{X},\mathcal{D}}\left(t, y\right) = \frac{1}{N} \sum_{i=1}^{N} K\left(\frac{y - \mathcal{X}_i(t)}{h(t)}\right) \tag{5.144}$$

where K is given by

$$K(x) = \int_{-\infty}^{x} k(t)dt. \tag{5.145}$$

Then for any $p, q \in [0,1]$ we have the following relation

$$Q_{\mathcal{X},\mathcal{D},q}(t) = Q_{\mathcal{X},\mathcal{D},p}(t) + K_{q,p} \cdot h(t) \tag{5.146}$$

where $K_{q,p}$ does not depend of $t \in \mathcal{D}$, and is solution of an equation of the following form

$$K_{q,p} + o(K_{q,p}^2) = \frac{q - p}{\pi}. \tag{5.147}$$

Proof. Let us suppose that a real $K_{q,p}$ which fulfills (5.146) exists, then by definition of a functional quantile we can write $\forall t \in \mathcal{D}$

$$\begin{aligned} q &= \widehat{G}_{\mathcal{X},\mathcal{D}}\left[t; Q_{\mathcal{X},\mathcal{D},q}\right] \\ &= \frac{1}{N}\sum_{i=1}^{N} K\left(\frac{Q_{\mathcal{X},\mathcal{D},q}(t) - \mathcal{X}_i(t)}{h(t)}\right) \\ &= \frac{1}{N}\sum_{i=1}^{N} K\left(\frac{Q_{\mathcal{X},\mathcal{D},p}(t) + K_{q,p}\cdot h(t) - \mathcal{X}_i(t)}{h(t)}\right) \\ &= \frac{1}{N}\sum_{i=1}^{N} K\left(\frac{Q_{\mathcal{X},\mathcal{D},p}(t) - \mathcal{X}_i(t)}{h(t)} + K_{q,p}\right). \end{aligned}$$

But the first order of Taylors approximation of the kernel K is given by

$$K\left(x+\Delta\right) = K\left(x\right) + \Delta\cdot k\left(x\right) + o(\Delta^2)$$

and then we have

$$\begin{aligned} q &= \frac{1}{N}\sum_{i=1}^{N}\left\{K\left(\frac{Q_{\mathcal{X},\mathcal{D},p}(t) - \mathcal{X}_i(t)}{h(t)}\right)\right. \\ &\quad \left. + K_{q,p}\cdot k\left(\frac{Q_{\mathcal{X},\mathcal{D},p}(t) - \mathcal{X}_i(t)}{h(t)}\right) + o(K_{q,p}^2)\right\} \\ &= \underbrace{\frac{1}{N}\sum_{i=1}^{N}\left\{K\left(\frac{Q_{\mathcal{X},\mathcal{D},p}(t) - \mathcal{X}_i(t)}{h(t)}\right)\right\}}_{=\widehat{G}_{\mathcal{X},\mathcal{D}}[t;Q_{\mathcal{X},\mathcal{D},p}]} \\ &\quad + \underbrace{\frac{K_{q,p}}{N}\sum_{i=1}^{N}\left\{k\left(\frac{Q_{\mathcal{X},\mathcal{D},p}(t) - \mathcal{X}_i(t)}{h(t)}\right)\right\}}_{=K_{q,p}\cdot h(t)\cdot\widehat{g}_{\mathcal{X},\mathcal{D}}[t;Qxp]} \\ &\quad + \frac{1}{N}\sum_{i=1}^{N} o(K_{q,p}^2) \\ &= \widehat{G}_{\mathcal{X},\mathcal{D}}[t; Q_{\mathcal{X},\mathcal{D},p}] + K_{q,p}\cdot h(t)\cdot\widehat{g}_{\mathcal{X},\mathcal{D}}[t;Qxp] + o(K_{q,p}^2) \\ &= p + K_{q,p}\cdot\pi + o(K_{q,p}^2). \end{aligned}$$

Thus

$$K_{q,p} + o(K_{q,p}^2) = \frac{q-p}{\pi}.$$

□

The use of the expression (5.144) to estimate the cdf $G_{\mathcal{X},\mathcal{D}}$ is, a priori, unusual, but make sense in this case. The non-use of the kernel estimation

of the cdf has several reasons, and one of these is the fact that the optimal bandwidth for the kernel density estimation is $h \approx O(N^{1/5})$ (see expression (4.8)) (where N is the sample size), while the optimal bandwidth for the kernel cdf estimation is $h \approx O(N^{1/3})$ [Bowman et al., 1998]. However [Hjort and Walker, 2001] have shown that using a bandwidth $h \approx O(N^{1/4})$ corrects this point.

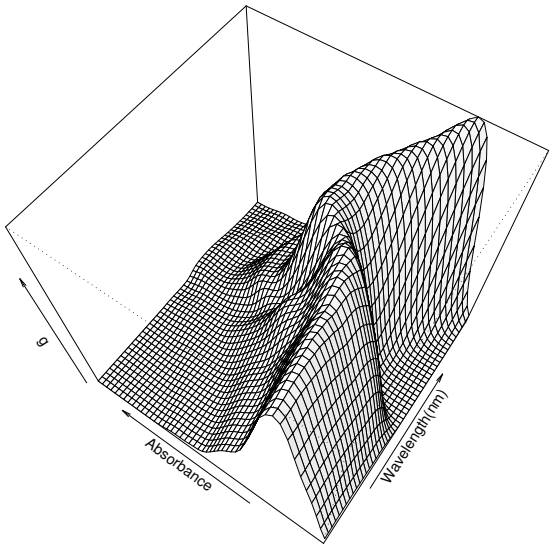


Figure 5.44: The estimated surface of densities $g_{\mathcal{X},\mathcal{D}}(t,y) * h(t)$ for the Tecator data (see fig. 2.11, p.21).

Then, under the condition expressed by (5.143) the direction of the Gâteaux differential is equivalent to the standard deviation of the data $\widehat{\sigma}$ when we use the optimal bandwidth (see expression (4.8))

$$h_{optimal}(t) = 1.06 \cdot \widehat{\sigma}(t) N^{-1/5}. \tag{5.148}$$

If $G_{\mathcal{X},\mathcal{D}}$ is expressed using a normal distribution, then the scale parameter s is also equal to the standard deviation σ, which can be, then, considered like a preferential direction to determine the direction of the Gâteaux density.

Figures from 5.44 to 5.52 illustrate the condition (5.143) and show the surfaces $g_{\mathcal{X},\mathcal{D}}(t,y) * h(t)$ in the kernel density estimation case. It is interesting to compare these graphs and the graphs show in figures 4.10 to 4.18, how the "factor" $h(t)$ "softens" the surfaces.

Let us remark that the condition expressed by (5.143) can be seen as "constant shape" condition, because in this case, for a fixed $p \in [0,1]$ the values of $g_{\mathcal{X},\mathcal{D}}$ depends only of the value of $h(t)$.

If the condition expressed by (5.143) is not fulfilled, then it will be interesting to envisage research on statistical tests on the value of the standard

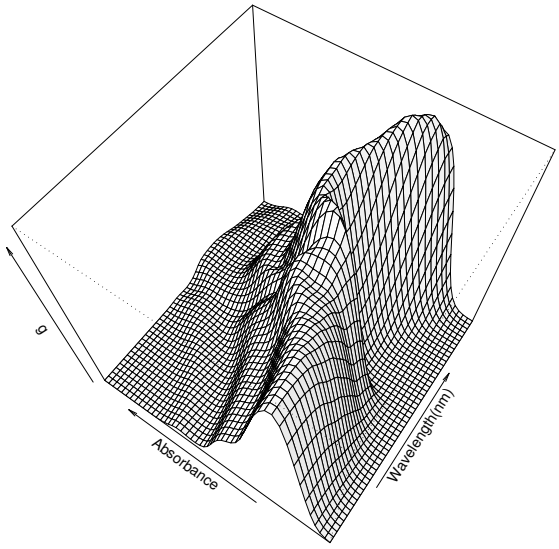


Figure 5.45: The estimated surface of densities $g_{\mathcal{X},\mathcal{D}}(t,y) * h(t)$ for the Tecator data with high fat content (see fig. 2.22, p.29).

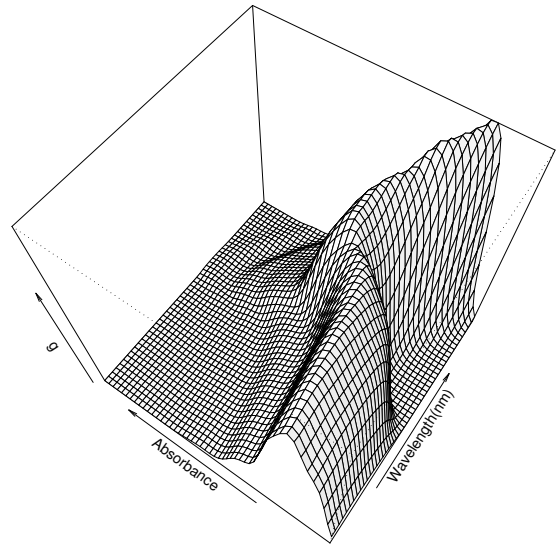


Figure 5.46: The estimated surface of densities $g_{\mathcal{X},\mathcal{D}}(t,y) * h(t)$ for the Tecator data with low fat content (see fig. 2.23, p.30).

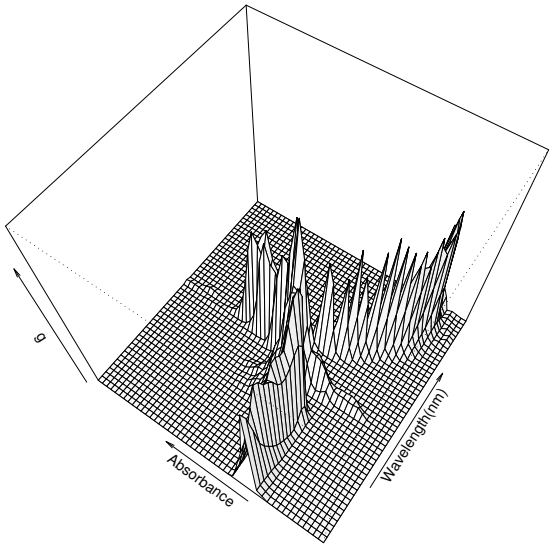


Figure 5.47: The estimated surface of densities $g_{\mathcal{X},\mathcal{D}}(t,y) * h(t)$ for the first derivative of the Tecator data (see fig. 2.20, p.28).

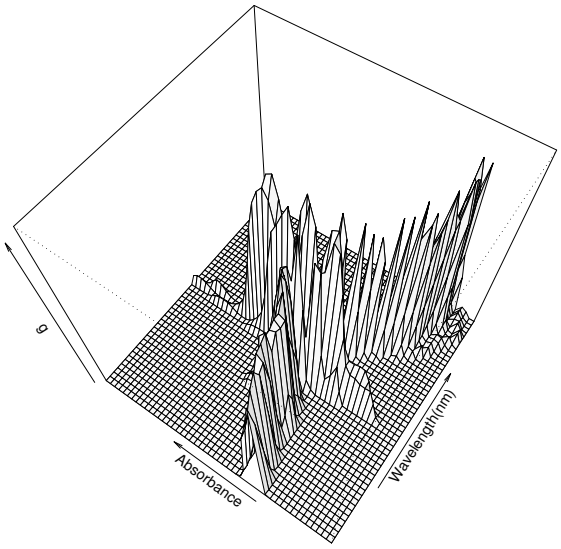


Figure 5.48: The estimated surface of densities $g_{\mathcal{X},\mathcal{D}}(t,y) * h(t)$ for the first derivative of the Tecator data with high fat content (see fig. 2.24, p.30).

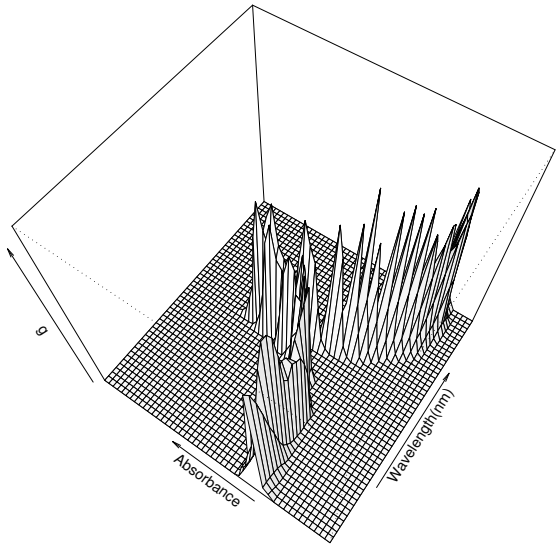


Figure 5.49: The estimated surface of densities $g_{\mathcal{X},\mathcal{D}}(t,y) * h(t)$ for the first derivative of the Tecator data with low fat content (see fig. 2.25, p.31).

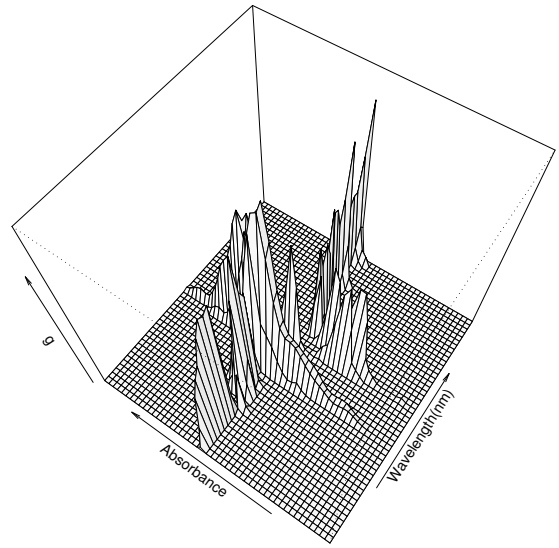


Figure 5.50: The estimated surface of densities $g_{\mathcal{X},\mathcal{D}}(t,y) * h(t)$ for the first derivative of the Tecator data (see fig. 2.21, p.28).

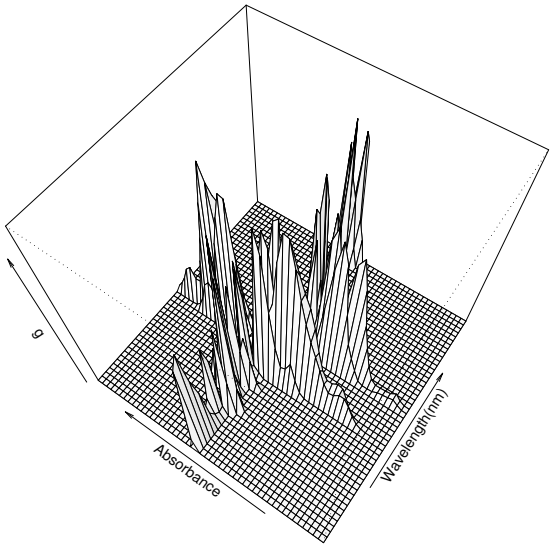


Figure 5.51: The estimated surface of densities $g_{\mathcal{X},\mathcal{D}}(t,y) * h(t)$ for the first derivative of the Tecator data with high fat content (see fig. 2.21, p.28).

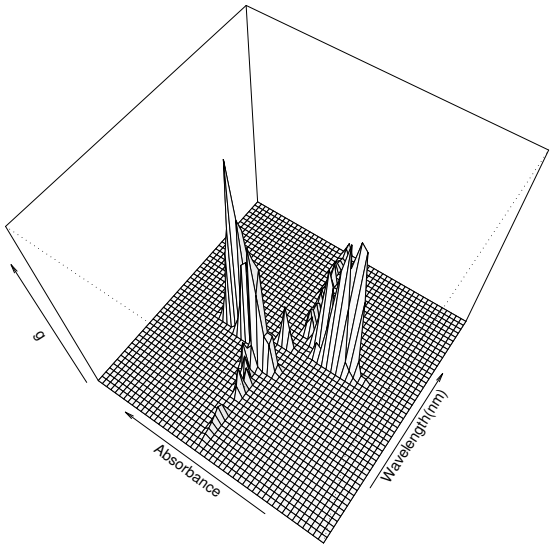


Figure 5.52: The estimated surface of densities $g_{\mathcal{X},\mathcal{D}}(t,y) * h(t)$ for the first derivative of the Tecator data with low fat content (see fig. 2.21, p.28).

deviation of $g_{\mathcal{X},\mathcal{D}}[t; Q_{\mathcal{X},\mathcal{D},p}] \cdot h(t)$:

$$\sqrt{\frac{1}{|\mathcal{D}|}\int_{\mathcal{D}} \left(g_{\mathcal{X},\mathcal{D}}[t; Q_{\mathcal{X},\mathcal{D},p}] \cdot h(t) - \pi\right)^2 dt} \tag{5.149}$$

where

$$\pi = \frac{1}{|\mathcal{D}|}\int_{\mathcal{D}} g_{\mathcal{X},\mathcal{D}}[t; Qxp] \cdot h(t)dt. \tag{5.150}$$

We have defined the Gâteaux density for the QAMML distributions, but we need to extend it to the case of GQAMML distribution, using the chain rule.

Definition 5.8.4. *Let $\mathcal{X}$ be an* frv*, and u a function, both belonging to $L^1(\mathcal{D})$. If it there exists a partition $\{A, B\}$ of $\mathcal{D}$ such the GQAMML distribution $\mathbb{F}_{\mathcal{X},\mathcal{D}}(u)$ is given by expression* (5.80)*, and if $\mathcal{F}_{\mathcal{X},A,\phi_A}$ and $\mathcal{F}_{\mathcal{X},B,\phi_B}(u)$ have Gâteaux densities with respective directions s_A and s_B, then the Gâteaux density of $\mathbb{F}_{\mathcal{X},\mathcal{D}}$ computed in u is given by the following expression*

$$f_{\mathcal{X},\mathcal{D},\Phi}(u) = \psi'_{\mathcal{D}}\left(\frac{|A|}{|\mathcal{D}|}\phi_{\mathcal{D}}\left[\mathcal{F}_{\mathcal{X},A,\phi_A}(u)\right] + \frac{|B|}{|\mathcal{D}|}\phi_{\mathcal{D}}\left[\mathcal{F}_{\mathcal{X},B,\phi_B}(u)\right]\right)$$
$$\left\{\frac{|A|}{|\mathcal{D}|}\phi'_{\mathcal{D}}\left[\mathcal{F}_{\mathcal{X},A,\phi_A}(u)\right] f_{\mathcal{X},A,\phi_A}(u) + \frac{|B|}{|\mathcal{D}|}\phi'_{\mathcal{D}}\left[\mathcal{F}_{\mathcal{X},B,\phi_B}(u)\right] f_{\mathcal{X},B,\phi_B}(u)\right\} \tag{5.151}$$

with

- $\Phi = (\phi_{\mathcal{D}}, \phi_A, \phi_B)$,
- $f_{\mathcal{X},A,\phi_A}(u)$ *and* $f_{\mathcal{X},B,\phi_B}(u)$ *are the respective Gâteaux densities of* $\mathcal{F}_{\mathcal{X},A,\phi_A}$ *and* $\mathcal{F}_{\mathcal{X},B,\phi_B}$.

Original contribution(s) 10. *In this section we give an original solution to the difficult problem of define and compute a density function in the infinite dimensional space of functions.*

For this we define a functional density similar to the density function in the univariate case: the functional density of $\mathcal{X}$ surrounded by two parametric functions which converge toward the considered function. We show also that, this density function is linked to the Gâteaux derivative, which is a directional derivative. We give the general expression of the Gâteaux derivative for QAMML distributions.

If the distribution used for the surface of margins has a scale parameter, then we prove that, the previously defined density for QAMML distributions is given by the Gâteaux derivative of QAMML, in direction of the scale parameter, divided by the norm of this latter. In this case we call the density: the Gâteaux density. This result in conjunction with the Gâteaux derivative

for QAMML distributions permits to give an expression for the Gâteaux density.

We give also an expression for the Gâteaux density computed for a functional quantile. And when the surface of margins is estimated using a kernel with a scale parameter, then we show that, under some conditions, this scale parameter is the direction to choose for the Gâteaux derivative.

We end this section in given the general expression of the Gâteaux density of GQAMML distributions.

5.9 Parameters Estimation

Before concluding this chapter, let us consider the the parameter estimation issue for the QAMML (and GQAMML) distributions. In section 4.7 we have explained the two steps estimation method called IFM for copulas. We propose to use a similar approach, because QAMM, QAMML and GQAMML distributions are built on existing margins (i.e. the surface of margins), and that, it is necessary, to estimate firstly these in the first place. Thus, firstly we estimate the margins parameters, and then estimate the (G)QAMM(L) parameter(s).

To estimate the parameters of the surface of margins, we suppose that the N data are measured in n values of $\mathcal{D}$ as shown in table 5.4. Then, firstly the parameters of $\mathbf{G}_{\mathcal{X},\mathcal{D}}[t_i;.]$ are estimated for each $i \in \{1,\ldots,n\}$ and then we obtain a vector of parameters: $(\theta_1,\ldots,\theta_n)$, where each θ_j could be a single real number, a vector or a function[4].

$j \setminus i$	t_1	$\ldots$	t_n
u_1	$u_1(t_1)$	$\ldots$	$u_1(t_n)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
u_N	$u_N(t_1)$	$\ldots$	$u_N(t_n)$

Table 5.4: Structure of a discrete functional dataset

For the sake of illustration:

- if $G_{\mathcal{X},\mathcal{D}}$ and $g_{\mathcal{X},\mathcal{D}}$ are estimated using the normal distribution, then for each i we have $\widehat{\theta}_i = (\widehat{\mu}_i, \widehat{\sigma}_i)$, and these parameters can be used both for the surface of margins and for the surface of densities;
- if $G_{\mathcal{X},\mathcal{D}}$ is estimated using the empirical distribution, then $\mathbf{G}_{\mathcal{X},\mathcal{D}}[t_i;.]$ itself is the parameter, i.e. $\widehat{\theta}_i = (\mathbf{G}_{\mathcal{X},\mathcal{D}}[t_i;y_1],\ldots,\mathbf{G}_{\mathcal{X},\mathcal{D}}[t_i;y_m])$, where $(y_1,\ldots,y_m)$ are m equally distant points of $\mathcal{I}$, or $\widehat{\theta}_i = \mathbf{G}_{\mathcal{X},\mathcal{D}}[t_i;.]$ where $\mathbf{G}_{\mathcal{X},\mathcal{D}}[t_i;.]$ is stored in a functional way (cf.section 2.3);

[4]The functions "parameters" can be stored in a functional way, or using a sampling of the function.

- if $G_{\mathcal{X},\mathcal{D}}$ and $g_{\mathcal{X},\mathcal{D}}$ are estimated using the kernel estimation, then $\mathbf{G}_{\mathcal{X},\mathcal{D}}[t_i;.]$ and $\mathbf{g}_{\mathcal{X},\mathcal{D}}[t_i;.]$ themselves are the parameters, i.e the parameter for $\mathbf{G}_{\mathcal{X},\mathcal{D}}[t_i;.]$ is equal to $\widehat{\theta}_{G,i} = (\mathbf{G}_{\mathcal{X},\mathcal{D}}[t_i;y_1],\ldots,\mathbf{G}_{\mathcal{X},\mathcal{D}}[t_i;y_m])$, where $(y_1,\ldots,y_m)$ are m equally distant points of $\mathcal{I}$, or $\widehat{\theta}_{G,i} = \mathbf{G}_{\mathcal{X},\mathcal{D}}[t_i;.]$ where $\mathbf{G}_{\mathcal{X},\mathcal{D}}[t_i;.]$ is stored in a functional way, and the parameters for $\mathbf{g}_{\mathcal{X},\mathcal{D}}[t_i;.]$ are equal to $\widehat{\theta}_{g,i} = (\mathbf{g}_{\mathcal{X},\mathcal{D}}[t_i;y_1],\ldots,\mathbf{g}_{\mathcal{X},\mathcal{D}}[t_i;y_m])$, or $\widehat{\theta}_{g,i} = \mathbf{g}_{\mathcal{X},\mathcal{D}}[t_i;.]$ where $\mathbf{g}_{\mathcal{X},\mathcal{D}}[t_i;.]$ is stored in a functional way.

In the case of QAMML distributions, once parameters of the surface of margins are estimated, we propose to determine the parameters of ϕ using the classical maximum likelihood method. The likelihood is naturally given by the following expression

$$\mathcal{L}(\theta) = \prod_{i=1}^{N} f_{\mathcal{X},\mathcal{D},\phi_\theta}(\mathbf{x}_i) \tag{5.152}$$

where θ are the parameters of ϕ_θ. And the log-likelihood is given by

$$\mathcal{L}^*(\theta) = \sum_{i=1}^{N} \log f_{\mathcal{X},\mathcal{D},\phi_\theta}(\mathbf{x}_i). \tag{5.153}$$

The estimation of θ is then given by

$$\widehat{\theta} = \arg\max_{\theta} \mathcal{L}^*(\theta). \tag{5.154}$$

For the GQAMML distribution we must have a three steps estimation methods, because GQAMMML is mean of QAMMML.

Following is a summary for the estimation of parameters:

1. estimate the (functional) parameters $\widehat{\theta}_{G,g}$ of surfaces of margins and densities,

2. estimate the parameters of $\widehat{\theta}_{\phi}$ of the QAMM(L) distribution, based upon $\widehat{\theta}_{G,g}$,

3. estimate the parameters of $\widehat{\Theta}$ of the GQAMML distribution, based upon $\widehat{\theta}_{G,g}$ and $\widehat{\theta}_{\phi}$.

Original contribution(s) 11. *In this section we give a multi-steps estimation methods, based on the IFM method for copulas. This method, estimate firstly the functional parameters of the surfaces of margins and densities, and then use the Gâteaux density in a maximum likelihood method to estimate parameters of QAMML and GQAMML distributions.*

5.10 Conclusions

In this chapter we first showed that any function can be approximated by quantilized functions, which are parts of quantile functions. Quantilized functions suppose that, if we know a function u_n only in n points $(t_1, \ldots, t_n)$: $(u_n(t_1), \ldots, u_n(t_n))$, then we suppose that the value in $[t_i, t_{i+1}[$ has the same univariate probability that in t_i, i.e. u is such that, for any $t \in [t_i, t_{i+1}[$ we have: $\mathbf{G}_{\mathcal{X},\mathcal{D}}[t; u_n] = \mathbf{G}_{\mathcal{X},\mathcal{D}}[t_i; u_n]$. The concept is, for a quantilized function, when we know the function in $(t_1, \ldots, t_n)$, then we know the probability for all $t \in \mathcal{D}$.

Adding to this the condition[5] that a functional quantile of order p must have an *fcdf* equal to p, we were naturally led to define a new kind of distribution as a *quasi-arithmetic mean* of margins. And these *Quasi-Arithmetic Mean of Margins* are joint probability distributions if the generator of the *quasi-arithmetic mean* is a generator of an Archimedean copula. The continuous case, the *Quasi-Arithmetic Mean of Margins Limit* is then obtained immediately in taking the limit of the QAMM of a sequence of quantilized functions tending to the considered function. Thereafter an extended version of QAMML is given as a *quasi-arithmetic mean* of QAMML, and this because generators of Archimedean copulas higher than 2 are not able to capture negative dependence structures.

Finally we had to deal with the density issue in the infinite dimensional space of functions. Like the definition of the *fcdf* of an *frv* is very similar to the cdf of a single real random variable, we have proposed to define the functional density using a single derivative. This latter being the Gâteaux derivative which is a functional directional derivative. Using this directional derivative when the surface of margins has a scale parameter or when this surface is estimated using the classical kernel estimation, then using the functional scale parameter or the functional standard deviation as directions in the Gâteaux derivative, we have defined an adapted kind of density for the functional case: the Gâteaux density. Therefore we have, proposed to use this latter in a several steps maximum likelihood estimation method,inspired by the IFM method for copulas, for the estimation of parameters in QAMMLS distributions.

It is interesting to remark that, in the univariate case, the *cdf* and the density are linked by derivative, and conversely by integration. In the functional case we define the Gâteaux density as a directional derivative of a *fcdf*, but do not yet have a formula which expresses the *fcdf* as an functional integral of the Gâteaux density. However, René Gâteaux, young promising mathematician, dead too early during the first world war [Mazliak, 2007], worked on integrals in an infinite number of dimensions [Gâteaux, 1919b], and this, independently of Percy John Daniell

[5] And some other additional conditions.

[Daniell, 1919b]. After his death, Jacques Hadamard gave the unpublished papers of Gâteaux to Paul Lévy who published them posthumously: [Gâteaux, 1919a], [Gâteaux, 1919b] and [Gâteaux, 1922]. In [Mazliak, 2007] , Malziak says: *"His ideas were extensively developed later by Lévy. Among other things, Lévy interpreted Gâteaux's integral in a probabilistic framework that later led to the construction of Wiener measure"*. It will be interesting to see, in the the future, how to connect, the Gâteaux density and the *fcdf* using an integration in an infinite number of dimensions.

Chapter 6

Applications

All models are wrong, but some are useful.

George E. P. Box

A large part of mathematics which becomes useful were developed with absolutely no desire to be useful, and in a situation where nobody could possibly know in what area it would become useful

John Von Neumann

La science qui n'aurait en vue que les applications ne serait plus de la science, elle ne serait plus que de la cuisine.

Henri Poincaré

6.1 Introduction

The goal of this chapter is, firstly, to show that QAMML distributions and the Gâteau density permit to apply directly, in symbolic data analysis and in functional data analysis, classical methods of univariate or multivariate analysis. And, conversely, these results are also a validation of the utility of all the concepts developed all along this thesis.

We have chosen three classical applications built upon probability distributions and probability densities: unsupervised classification by mixture decomposition, bayesian supervised classification and construction of functional confidence intervals. For these three applications, we have just

"plugged" QAMML distributions and the Gâteaux density in these existing methods. Then, more than the acuity of the performance, the main result is the fact that QAMML distributions and the Gâteaux density can be used directly in existing methods, with very acceptable results.

We begin in the first section with the unsupervised classification of probability distribution in the Symbolic Data Analysis framework. In the following section we perform a Bayesian supervised classification on the functional Tecator dataset. And in the third section we propose to build a solution to the problem of finding confidence bounds in the functional context.

6.2 Unsupervised Classification by Mixture Decomposition

The task of unsupervised classification[1] is to divide a set of data into groups (or clusters) in such a way that objects from the same cluster are more similar to each other than objects from other clusters. Among the different existing techniques, those based on mixture decompositions are the more interesting in our case, because they are based on probability distributions.

In the mixture decomposition classification, we suppose that we have K random variables $\mathcal{X}_1, \ldots, \mathcal{X}_K$, and N given observations $u_1, \ldots, u_N$. The problem is, for each individual u_i, to retrieve from which random variable $\mathcal{X}_k$, u_i is a realization, and thus what is the best group for u_i. Two main algorithms to solve this problem, are the EM algorithm (Expectation-Maximization) [Dempster et al., 1977] (and more precisely CEM its classification version) and the Dynamical Clustering algorithm [Diday et al., 1974]. In both algorithms, the main idea is to find a repartition of individuals u_i, between groups associated to random variables, which maximizes a likelihood.

The classification version of the EM algorithm is due to [Celeux and Govaert, 1992] and add a classification step to the EM algorithm which is devoted to the estimation of the parameters of a mixture. The mixture is defined as follows: given a realization u of one of the random variables, the probability of u is given, using the law of total probability, by the following expression

$$\begin{aligned} P(u) &= \sum_{k=1}^{K} P(u|\mathcal{X}_k = u) \cdot P(\mathcal{X}_k = u) \\ &= \sum_{k=1}^{K} f_{\mathcal{X}_k}(u) \cdot \pi_{\mathcal{X}_k} \end{aligned} \tag{6.1}$$

[1] The unsupervised classification is also called data clustering (or just clustering),cluster analysis, automatic classification, numerical taxonomy or unsupervised learning!

where $\pi_{\mathcal{X}_k} = P(\mathcal{X}_k) = u$ and $f_{\mathcal{X}_k}(u) = P(u|\mathcal{X}_k = u)$. In the expression (6.1), we suppose that the densities $f_{\mathcal{X}_1}, \ldots, f_{\mathcal{X}_K}$ follow known probability laws, with for each $k \in \{1, \ldots, K\}$ a parameter θ_k:

$$P(u) \quad = \quad \sum_{k=1}^{K} f_{\mathcal{X}_k}(u|\theta_k) \cdot \pi_{\mathcal{X}_k}. \tag{6.2}$$

The likelihood of the mixture is defined as follows:

$$L(\mathbf{u}, \Theta, \Pi) = \prod_{i=1}^{N} \sum_{k=1}^{K} f_{\mathcal{X}_k}(u_i|\theta_k) \cdot \pi_{\mathcal{X}_k} \tag{6.3}$$

where $\mathbf{u} = \{u_1, \ldots, u_N\}$, $\Theta = (\theta_1, \ldots, \theta_K)$ and $\Pi = (\pi_{\mathcal{X}_1}, \ldots, \pi_{\mathcal{X}_K})$.

A logarithmic version, the log-likelihood, is often used for computational reasons:

$$L(\mathbf{u}, \Theta, \Pi) = \sum_{i=1}^{N} \log \left(\sum_{k=1}^{K} f_{\mathcal{X}_k}(u_i|\theta_k) \cdot \pi_{\mathcal{X}_k} \right). \tag{6.4}$$

The algorithm starts with a random initial solution $\Theta^0 = (\theta_1^0, \ldots, \theta_K^0)$ and $\Pi^0 = (\pi_{\mathcal{X}_1}^0, \ldots, \pi_{\mathcal{X}_K}^0)$, then the following steps are repeated until the convergence of the algorithm:

Iteration n $(n \geq 0)$**:**

E step (Estimation): compute the following posterior probabilities

$$t_k^n(u_i) = \frac{\pi_{\mathcal{X}_k}^n f_{\mathcal{X}_k}(u_i|\theta_k^n)}{\sum_{\ell=1}^{K} \pi_{\mathcal{X}_\ell}^n f_{\mathcal{X}_\ell}(u_i|\theta_\ell^n)} \tag{6.5}$$

for $1 \leq i \leq N$ and $1 \leq k \leq K$.

C step (Classification): Assign each individual u_j to the cluster P_k^n for which the probability $t_k^n(u_j)$ is maximal:

$$k^n(u_i) = \arg \max_{1 \leq k \leq K} t_k^n(u_i) \tag{6.6}$$

.

M step (Maximization): Compute the estimators $(\pi_{\mathcal{X}_k}^n, \theta_k^n)$ in the following way

$$\pi_{\mathcal{X}_k}^{n+1} = \frac{|P_k^n|}{N}, \quad k = 1, \ldots, K \tag{6.7}$$

$$\Theta^{n+1} = \arg \max_{\Theta} L(\mathbf{u}, \Theta, \Pi^n). \tag{6.8}$$

In the second algorithm, the Dynamical Clustering algorithm, the idea is to find a partition $P = (P_1, \ldots, P_K)$ of the sample $\{u_1, \ldots, u_N\}$ such each cluster P_i can be considered as a sample of the random variable $\mathcal{X}_k$, following the probability law $f_{\mathcal{X}_k}(.|\theta_k)$. We suppose that for a given partition $P = (P_1, \ldots, P_K)$, each set P_k has an associated parameter θ_k. We firstly define the likelihood of the set P_k as follows:

$$L(P_k, \theta_k) = \prod_{u_i \in P_k} f_{\mathcal{X}_k}(u_i|\theta_k). \tag{6.9}$$

Then, we define the classifying likelihood by

$$cl(P, \Theta) = \prod_{k=1}^{K} L(P_k, \theta_k). \tag{6.10}$$

Of course the logarithmic version of this latter is often usen

$$lcl(P, \Theta) = \sum_{k=1}^{K} \sum_{u_i \in P_k} \log\left(f_{\mathcal{X}_k}(u_i|\theta_k)\right). \tag{6.11}$$

The best partition P is the partition which maximize the lcl criterion. To find this partition, the classification starts with a random solution, and then the following steps are repeated until the stabilization of the partition:
Iteration n ($n \geq 1$)**:**

Parameter Estimation Step: Find the vector $(\theta_1, \ldots, \theta_K)$ which maximizes the lcl criterion;

Affectation Step: Build the new partition P^n with the parameters found in the preceding step:

$$P_k^n = \left\{ u_j : k = \arg\max_{\ell} f_{\mathcal{X}_\ell}(u_j|\theta_\ell) \right\}. \tag{6.12}$$

The Dynamical Clustering has already been used in the Symbolic Data Analysis in conjunction with the Archimedean copulas, i.e. with finite dimensional distributions given by expression (4.123), and was firstly proposed in [Diday, 2001] and [Diday, 2002]. Then it was used on climatological data in [Vrac et al., 2001], [Vrac, 2002], [Vrac et al., 2004] and [Diday and Vrac, 2005], using the Frank copula of dimension 2. In [Cuvelier and Noirhomme-Fraiture, 2003] we have proposed the use of the Clayton copula (cf. formula (4.102)) for any $n \geq 2$, and in [Cuvelier and Noirhomme-Fraiture, 2005] and [Cuvelier and Noirhomme-Fraiture, 2008a] we have applied it with success. In [Cuvelier and Noirhomme-Fraiture, 2006] and

Table 6.1: Range of the parameters for the 140 synthetic cdfs.

Name	**Size**	**Location**	**Scale**
Normal	45	$[-1.52, -0.86]$	$[0.79, 1.85]$
Beta	50	$[0.26, 1.2]$	$[8.99, 11.05]$
Exp	45	$[-0.56, 1.16]$	$[1.15, 2.15]$

[Cuvelier and Noirhomme-Fraiture, 2007] we have introduced QAMML distributions used in Dynamical Clustering of synthetic functional data in the Symbolic Data Analysis framework. That is why, we have decided to use the Dynamical Clustering to test the QAMML distributions in a clustering task. In the following we give the results of two clusterings. In both cases functional data are distributions functions. The first dataset is a synthetic dataset of cdfs, and the second one contains cdfs coming from real data.

To build the synthetic dataset of cdfs in a way to be closer of the case of real data, we do not choose directly the parameters of the distribution, but generate them "randomly", and then we work on an empirical estimation of the cdfs. Here is the complete generation steps:

1. we have chosen three well known families of distributions: Normal distributions, Beta distributions and Exponential distribution (cf. table 5.3),
2. for each family of distributions, we have chosen a range for each location and scale parameters, which are given in table 6.1 (for Beta distributions, additional shape parameters p and q were chosen respectively in $[1.70, 2.92]$ and $[1.83, 2.93]$),
3. to generate one distribution of a given family we have generated:
   (a) each parameter in the chosen range,
   (b) 500 random numbers with the chosen parameters,
4. we have estimated the empirical cdf of the distribution over the 500 generated numbers.

The resulting set is shown in fig. 6.1. We have used QAMML distributions with the following parameters:

- Surface of margins G: empirical distribution,
- Surface of densities G: kernel density distribution,
- QAMML and GQAMML generators: Clayton generator.

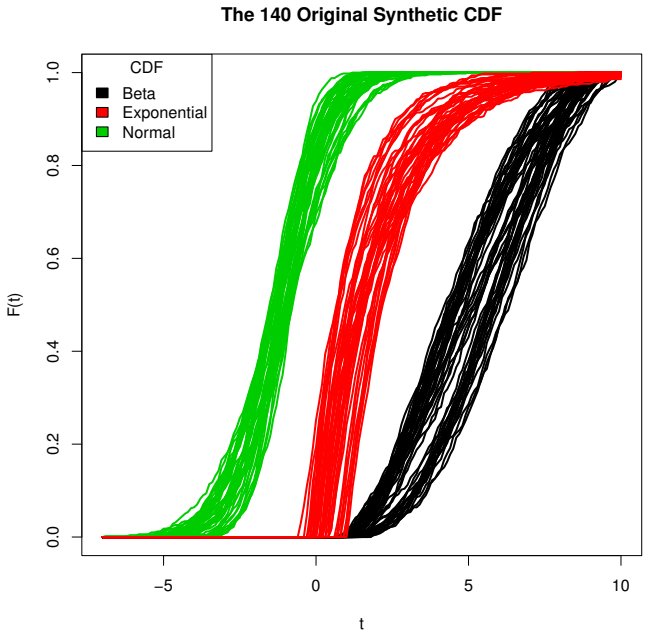


Figure 6.1: The 140 original synthetic cdfs.

Then, the Dynamical Clustering algorithm was used with the classifying log-likelihood criterion (expression (6.11)). The algorithm was applied 20 times and we have retained the result which gives the best criterion: 107.6513.

The figure 6.2 shows resulting clusters, and we see that almost all distributions, except in one case, belong to a cluster with the same type of distributions. The 6.3 figure shows surfaces of margins of the three clusters, using the level curves, i.e the functional quantiles, of values 0.2, 0.4, 0.6, 0.8 and 1.0. We see clearly that these functional quantiles have shapes close to the distributions of the corresponding cluster, excepted for the functional quantile of order 1 in the cluster number 1. Then we can conclude that the Dynamical Clustering in conjunction with QAMML distributions is able to create homogeneous clusters of distinct cdfs.

However, in case of real data, clusters are not necessarily so well separated, and then it is interesting to test the behavior of the Dynamical Clustering in conjunction with QAMML distributions on real data.

We have chosen to work on the Long-Term Instrumental Climatic Data Base of the People's Republic of China (`http://dss.ucar.edu/datasets/ds578.5/data/`). This set of data consists of monthly means, extremes, or totals of 14 meteorological variables observed at 60 stations in China during several years. The 60 stations are shown in the figure 6.4, and the 14 variables in the table 6.2. These climatological

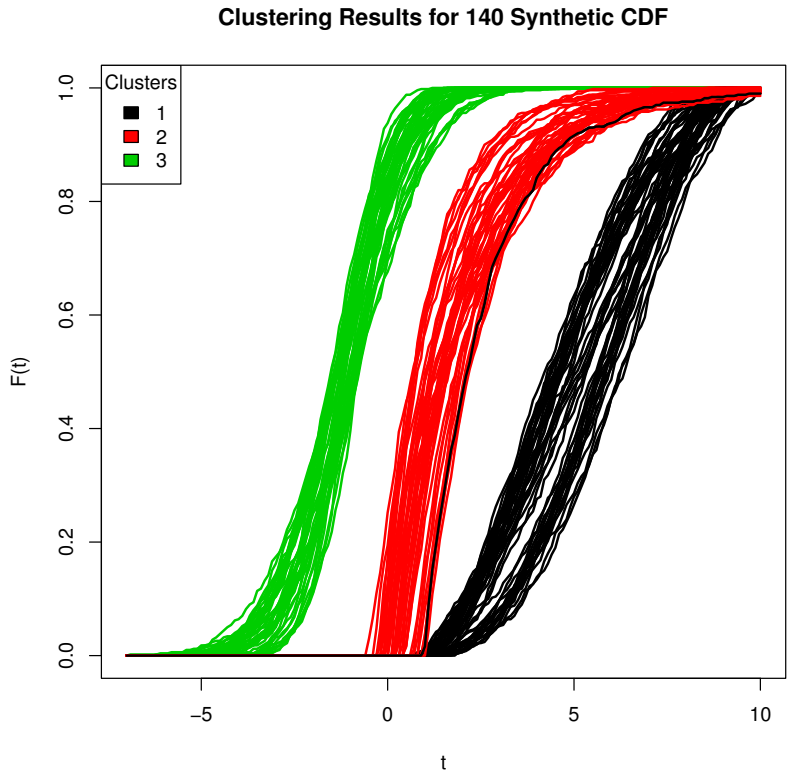


Figure 6.2: Clustering results of the mixture decomposition by dynamical clustering on synthetic empirical cdf.

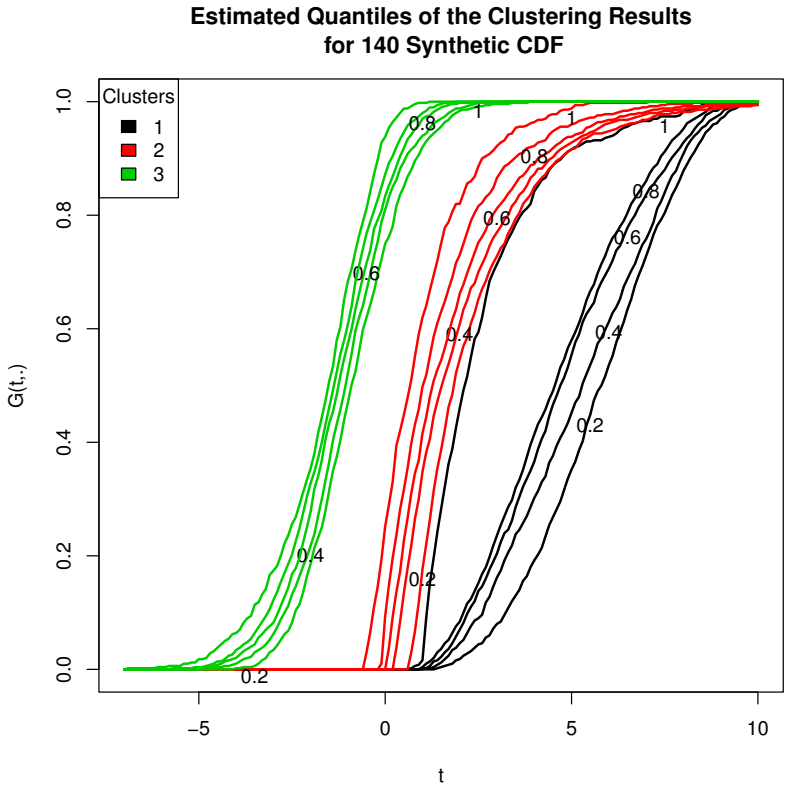


Figure 6.3: Functional Quantile/Surfaces of Margins of the clustering results of the mixture decomposition by dynamical clustering on synthetic empirical cdf.

Variables
Mean Station Pressure (mb)
Mean Temperature (C)
Mean Maximum Temperature (C)
Mean Minimum Temperature (C)
Total Precipitation (mm)
Sunshine Duration (hours)
Mean Cloud Amount (percentage of sky cover)
Mean Relative Humidity (percent)
Snow Days (days with snow cover)
Dominant Wind Direction (degrees)
Mean Wind Speed (m/s)
Dominant Wind Frequency (percent)
Extreme Maximum Temperature (C)
Extreme Minimum Temperature (C)

Table 6.2: The 14 climatological variables of the Long-Term Instrumental Climatic Data Base of the People's Republic of China

data were already used in several paper in the Symbolic data, for clustering interval data (see [Chavent et al., 2003], [Chavent et al., 2006] and [Diday and Noirhomme-Fraiture, 2008], chapter 11), usually with 5 clusters (we do not discuss this choice here).

We have chosen to work on the "Mean Minimum Temperature" variable, measured along the following periode: from 1978 to 1988. For each station we have computed the empirical cdf over these eleven years (i.e. 132 values for each stations). The resulting cdfs are shown in figure 6.5.

For using the Dynamical Clustering algorithm, with QAMML distributions we have used the following parameters:

- Surface of margins G: empirical distribution,
- Surface of densities G: kernel density distribution,
- QAMML and GQAMML generators: Clayton generator.

Then, the clustering algorithm was used with the classifying log-likelihood criterion (expression (6.11)). The algorithm was applied 20 times and we have retained the result which gives the best criterion: 22.083.

We show the results in the following figures:

- the figure 6.6 shows the overall clustered cdfs,
- the figure 6.7 shows a more detailed view: each cluster of cdf is plotted in its associated color and the remaining data are plotted in grey,

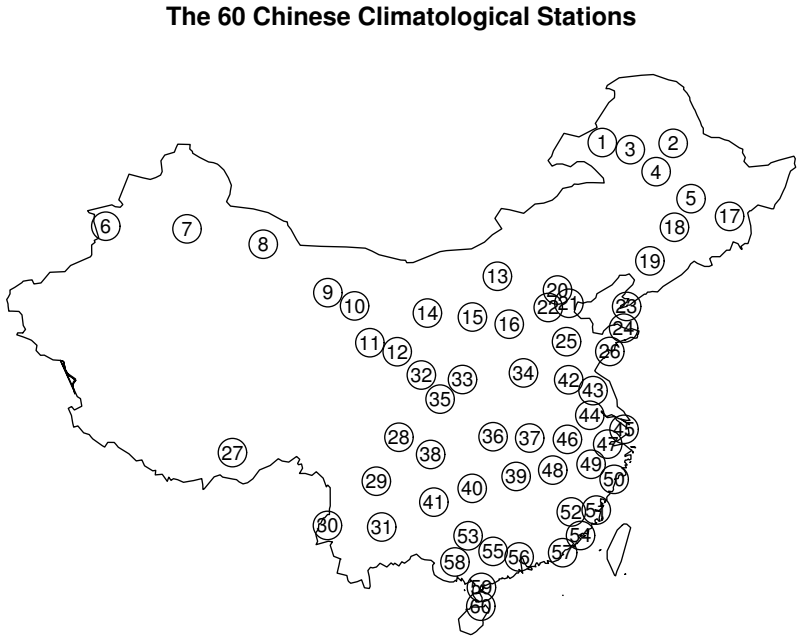


Figure 6.4: The 60 climatological stations of the Long-Term Instrumental Climatic Data Base of the People's Republic of China.

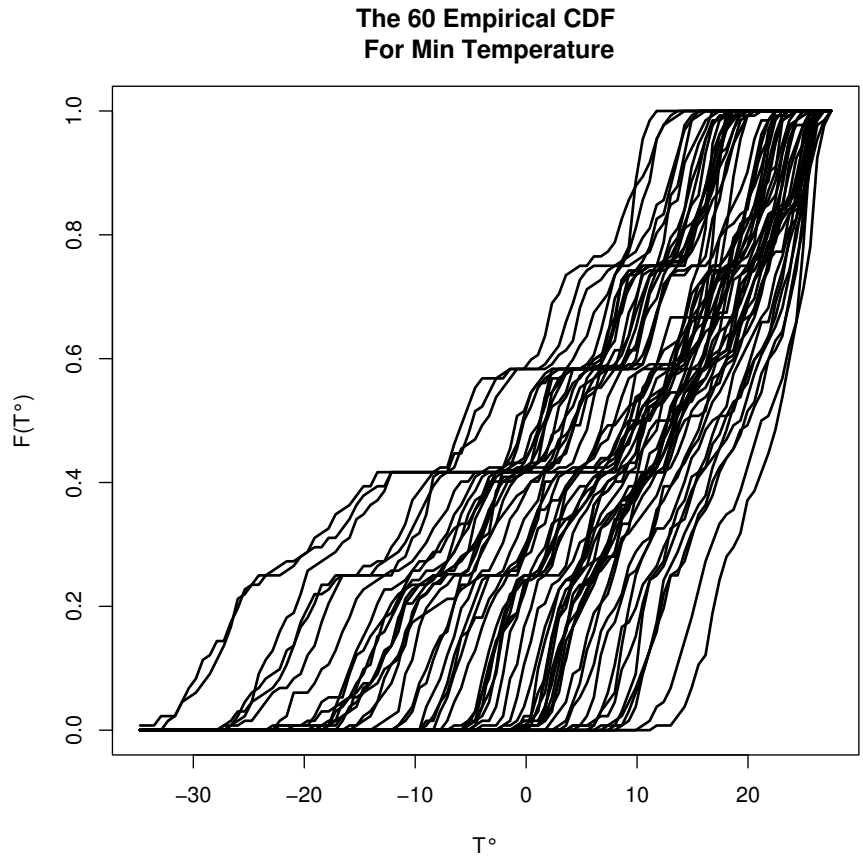


Figure 6.5: Empirical cdf for the 60 climatological stations of the Chinese Climatic Data Base.

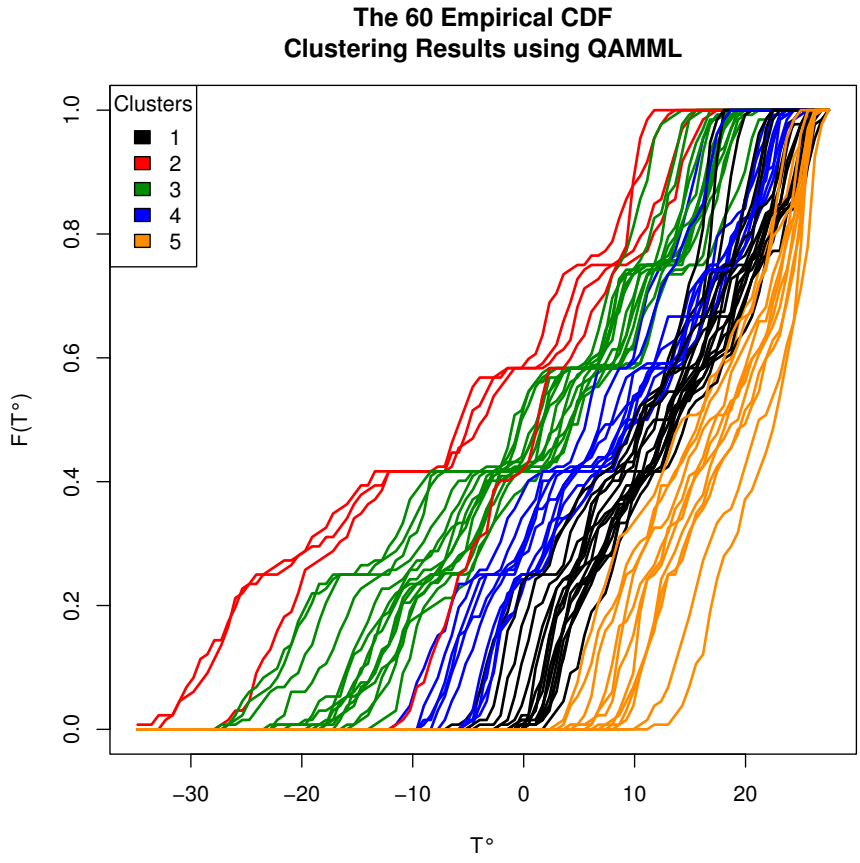


Figure 6.6: Clustering results on empirical cdfs of the climatological of the Chinese Climatic Data Base using QAMML.

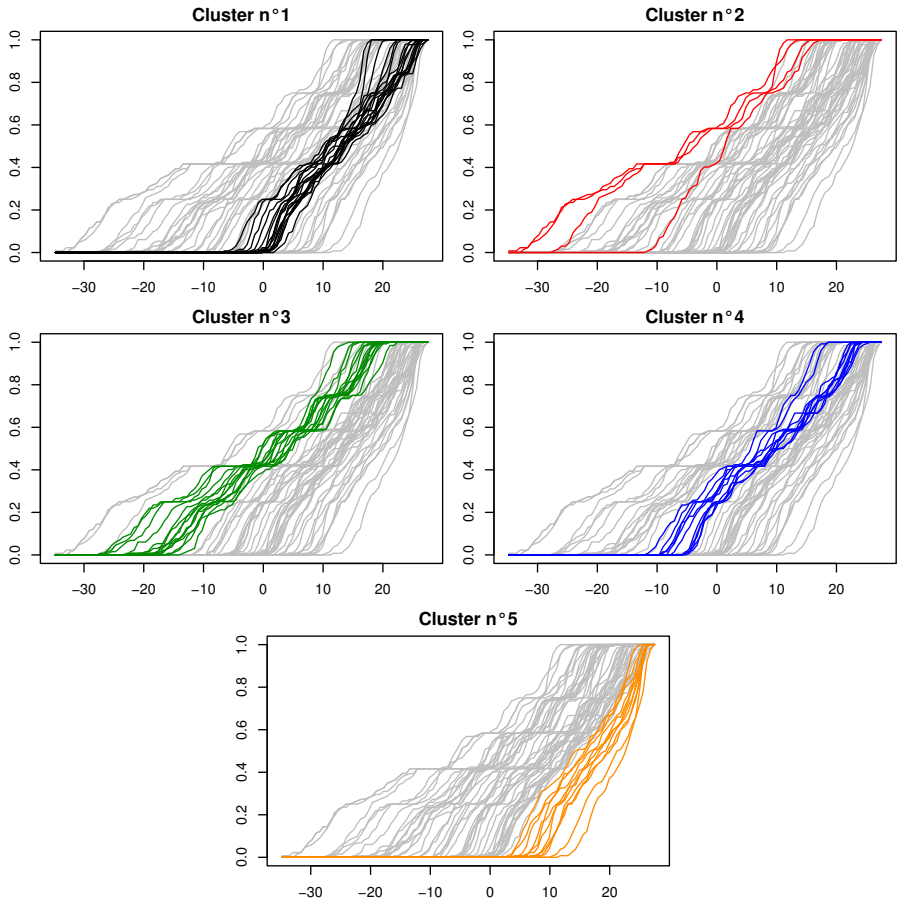


Figure 6.7: Detailed clustering results of empirical cdfs of the climatological of the Chinese Climatic Data Base using QAMML.

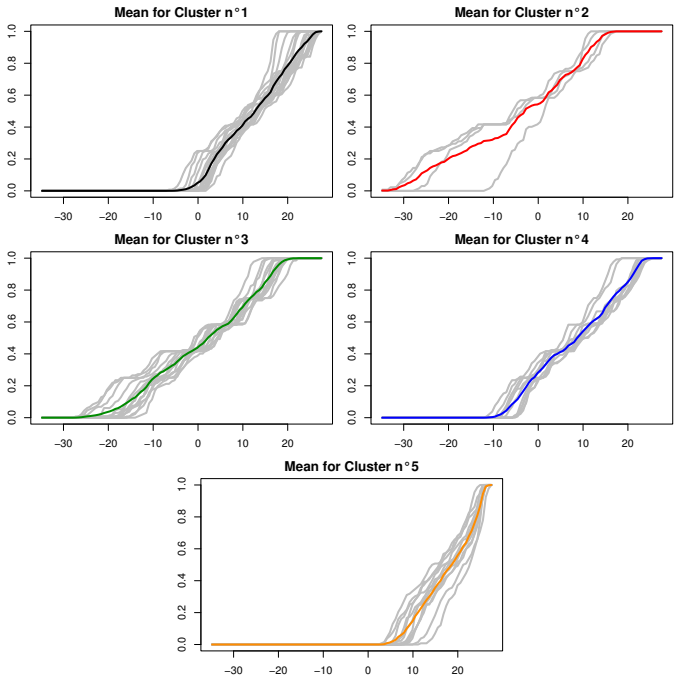


Figure 6.8: Means of the clusters of empirical cdfs of the climatological of the Chinese Climatic Data Base using QAMML.

- the figure 6.8 shows each cluster (plotted in grey) and its center (functional mean) in the associated color,
- the figure 6.9 shows the centers of the clusters in the same graph,
- the figure 6.10 shows each cluster (plotted in grey) and its functional standard deviations,
- the figure 6.11 represents the stations on the map of China with the number of the station in the color of its cluster,
- and the figure 6.12 represents the stations on the map of China with an altitude index (altitude divided by 100m) in the color of its cluster and we have added the 20th, 30th, 40th and 50th parallels.

For the interpretation of the results, let us firstly look at the figures 6.6 and 6.7. We see that, except for cluster 2, all clusters are homogeneous. For the cluster 2, it should be noted that this cluster seems homogeneous except for one cdf which is related to clusters 2, 3 and 4, and "cross" completely the cluster 3. A look at the figure 6.11 shows that the number of this station is 27. The range of temperatures of this station is smaller that the ranges of clusters 2 and 3, and its largest quantiles are smaller than largest quantiles

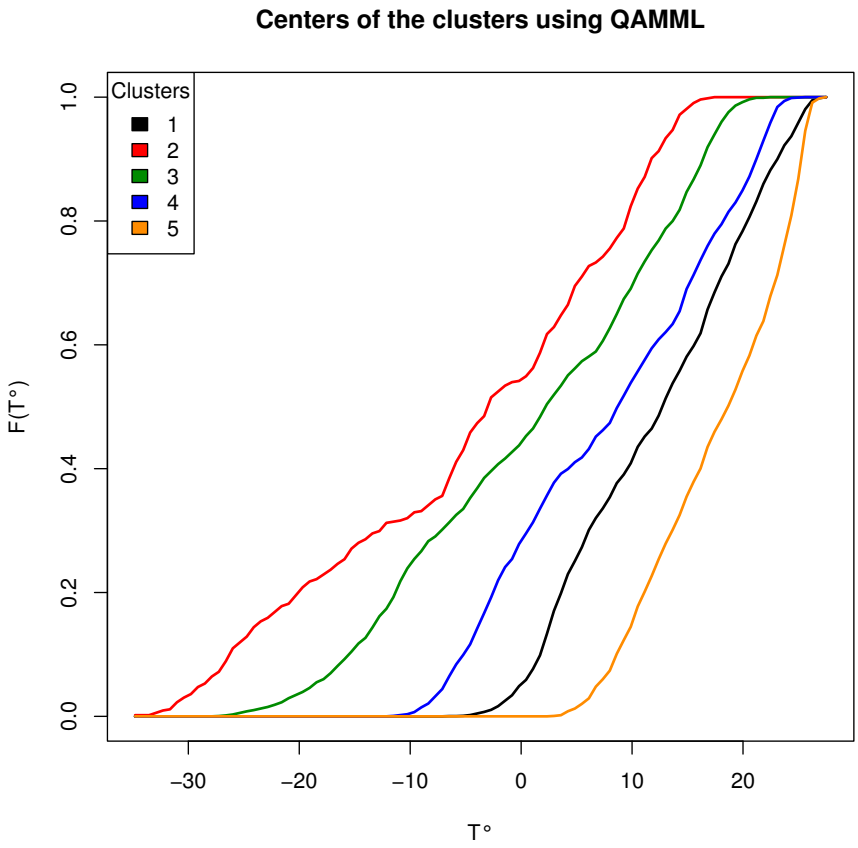


Figure 6.9: Centers of the clusters of empirical cdfs of the climatological of the Chinese Climatic Data Base using QAMML.

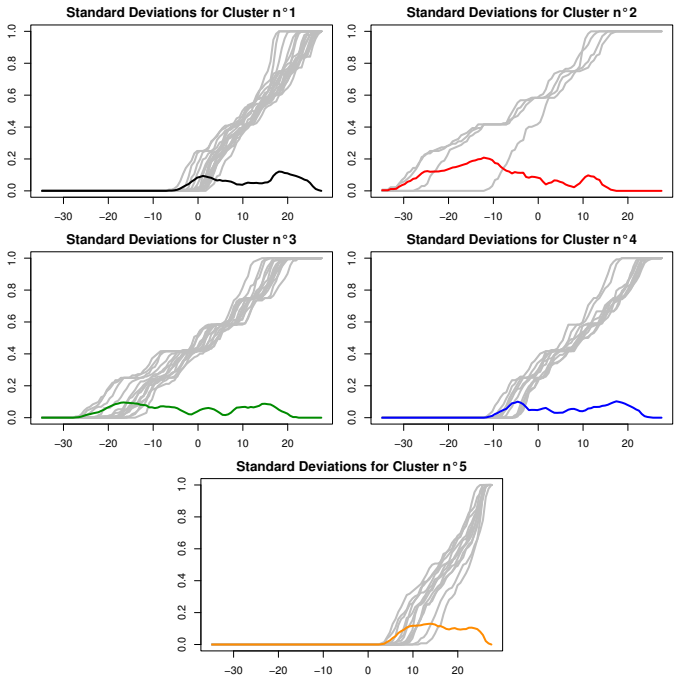


Figure 6.10: Standard deviations of the clusters of empirical cdfs of the climatological of the Chinese Climatic Data Base using QAMML.

of the cluster 3, ant its smaller quantiles are larger than the larger quantiles of the 3th cluster.

A look the figure 6.11 seems to show that the stations are grouped more or less following their latitudes. But the most interesting representation is in the figure 6.12. In this latter map we can see together clusters and three parameters that impact the climate of each station: the latitude, the altitude, the proximity of the sea. We can see that there is some sort of inverse relation between the altitude and the latitude of the stations inside of each cluster: the southerly stations of a clusters are stations with higher altitude. The more extreme case is the station 27 which belong to the cluster 2, formed by the northerly stations. Some stations of the cluster 5 seems to be exception to this remark, but there are all coastal stations, which involve a clear influence of the proximity of the sea in the climate.

Finally, the clearest interpretations concern cluster 2 and 5. The cluster 2 is formed by the northernly and/or coldest stations, and the cluster 5 contains the southerly and/or coastal stations. The cluster 3 seems formed by lowlands and highlands stations; clusters 1 and 4 are harder to interpret, their climate are maybe influenced by other parameters like precipitations or hydrology... More deep studies will be interesting in the future.

In introduction to this section, we have recalled the seminal works of

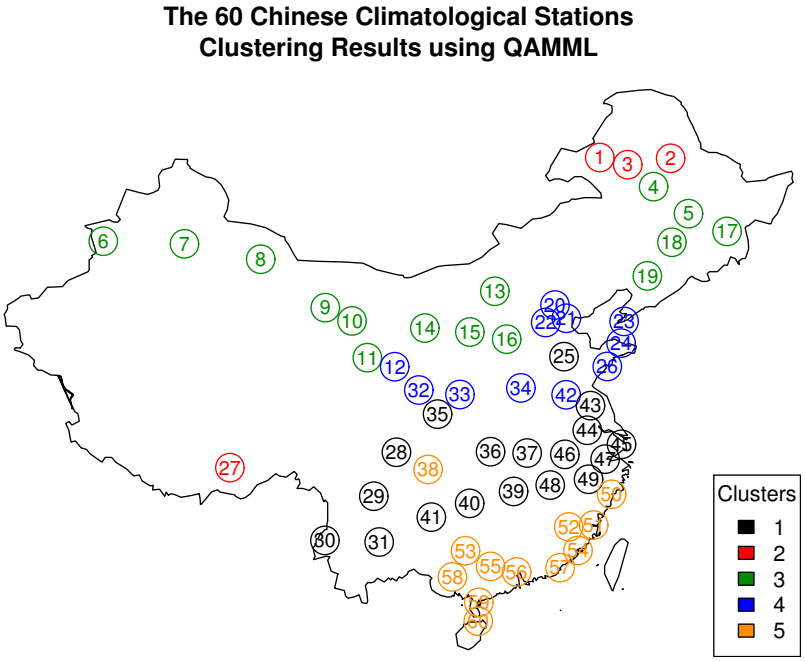


Figure 6.11: Map of the clustering results on empirical cdfs of the climatological of the Chinese Climatic Data Base using QAMML.

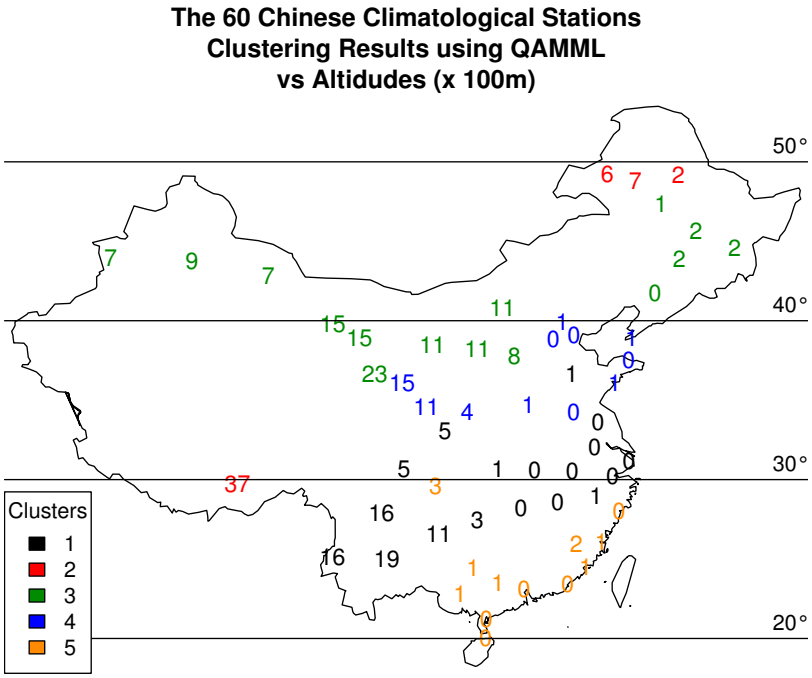


Figure 6.12: Map of the clustering results on empirical cdfs of the climatological of the Chinese Climatic Data Base using QAMML, with additional informations about altitudes and latitudes.

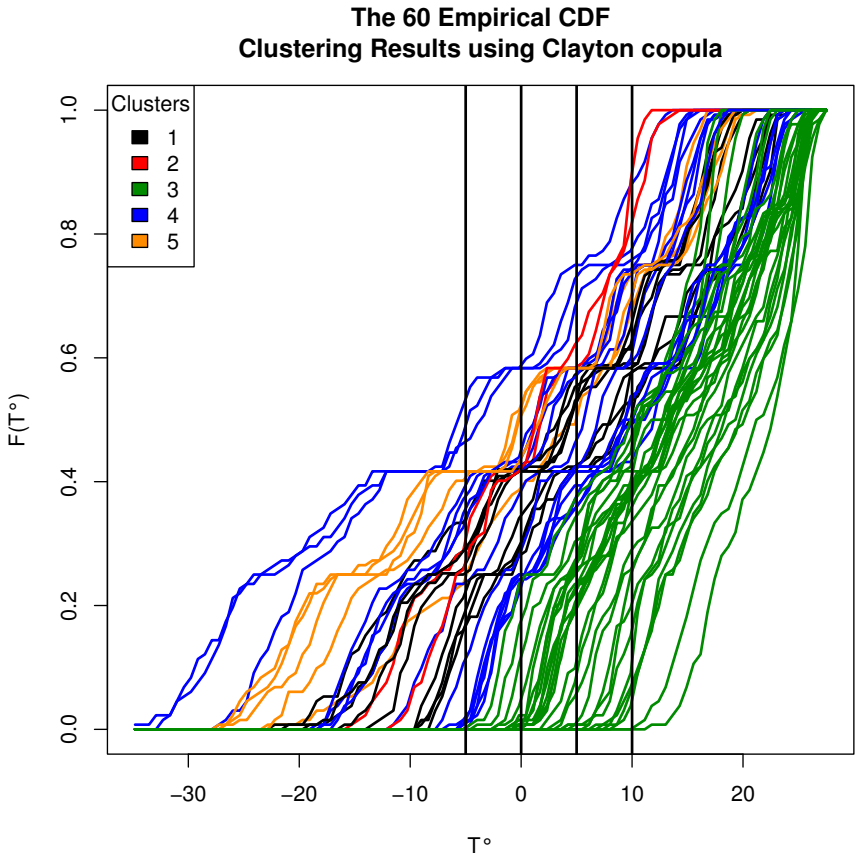


Figure 6.13: Clustering results on empirical cdfs of the climatological of the Chinese Climatic Data Base using the Clayton copula.

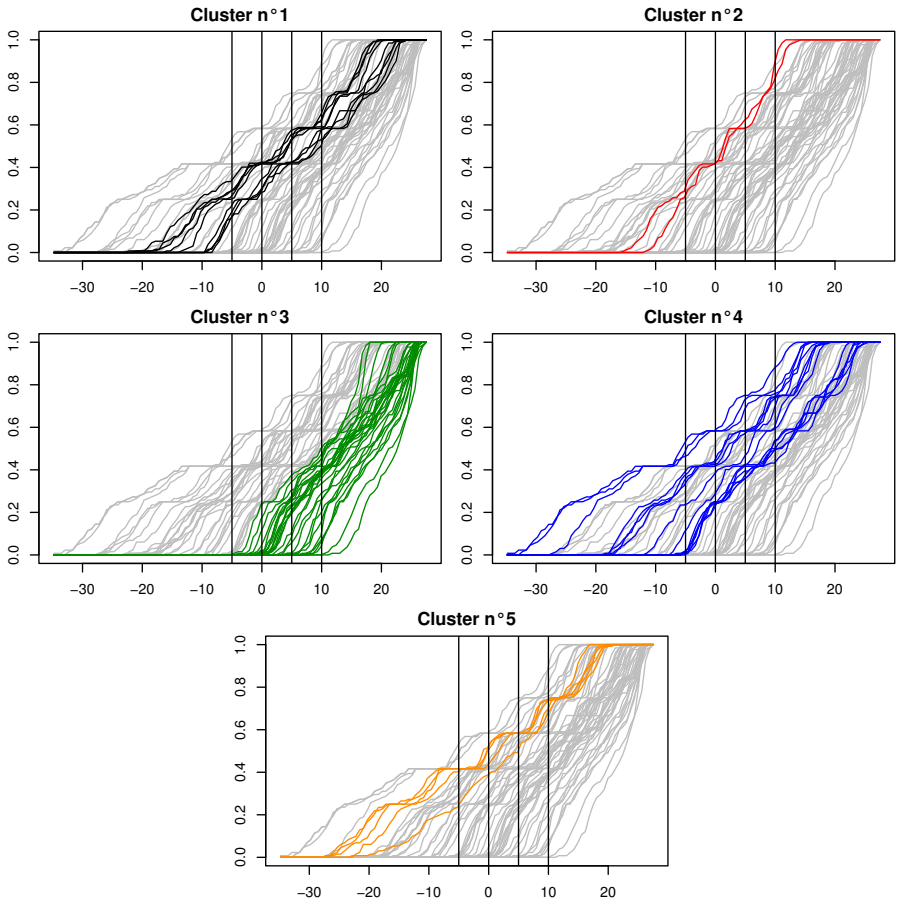


Figure 6.14: Detailed clustering results of empirical cdfs of the climatological of the Chinese Climatic Data Base using the Clayton copula.

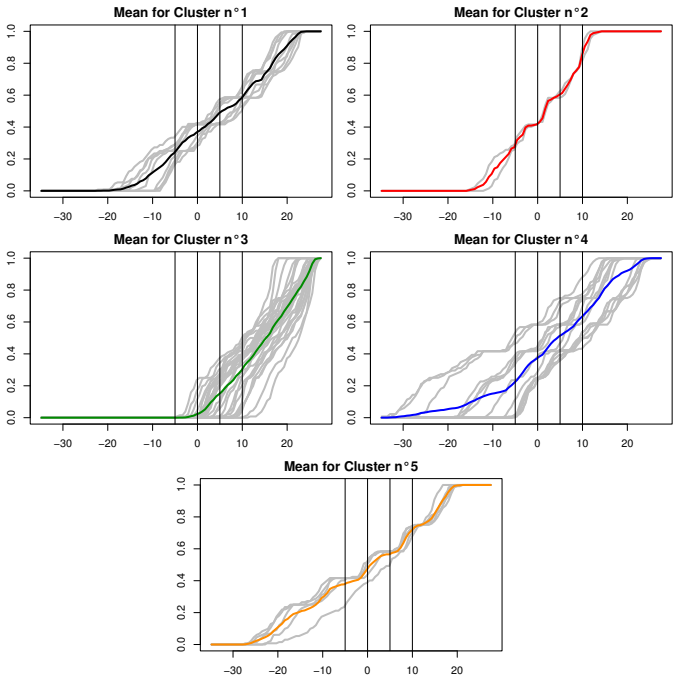


Figure 6.15: Means of the clusters of empirical cdfs of the climatological of the Chinese Climatic Data Base using the Clayton copula.

this thesis, then, for the sake of comparison, we have applied the mixture decomposition classification, using finite dimensional distributions build on Clayton copulas, as defined in chapter 4, section 4.8 (p. 139). In other words, we have used the same algorithm (dynamical clustering), but using the expression 4.125 (p. 140) to compute $f_{\mathcal{X}_\ell}(u_j|\theta_\ell)$,with the following parameters:

- Surface of margins G: empirical distribution,
- Surface of densities G: kernel density distribution,
- Archimedean generators: Clayton generator.

We have also chosen the classifying log-likelihood criterion (expression (6.11)), and again the algorithm was applied 20 times and we have retained the result which gives the best criterion.

But, first of all, let us remark that, to perform the classification, there is another mandatory choice to do: the choice of the set $\{t_1, \ldots, t_n\}$ for which finite dimensional distributions are computed. And this choice is not an easy choice, because it will be influential on the quality of the clustering results.

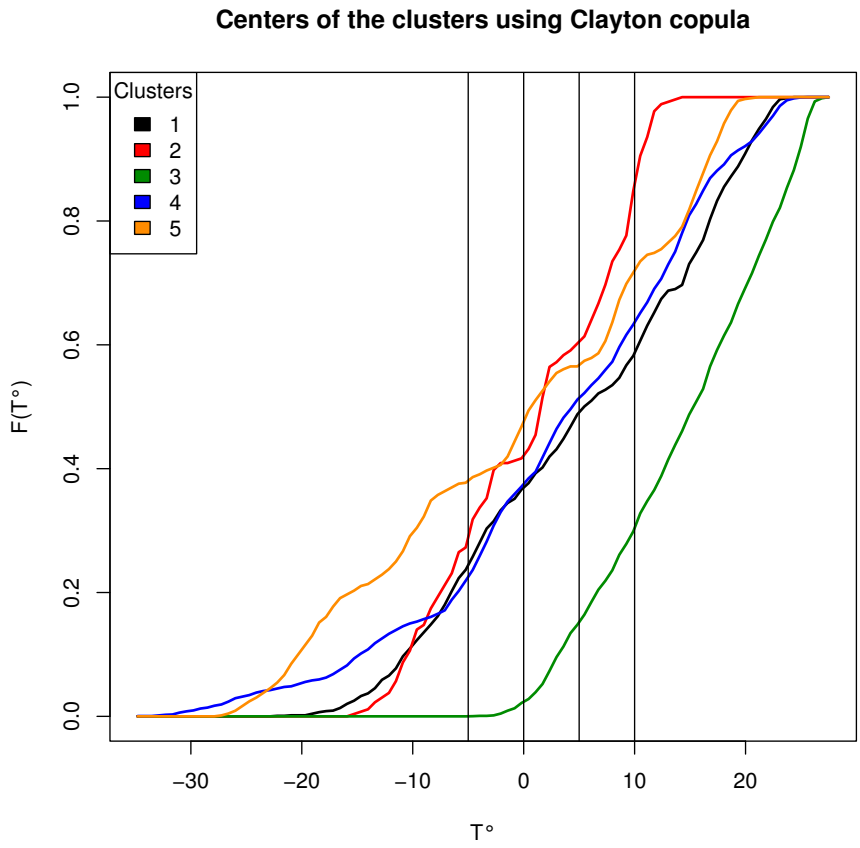


Figure 6.16: Centers of the clusters of empirical cdfs of the climatological of the Chinese Climatic Data Base using the Clayton copula.

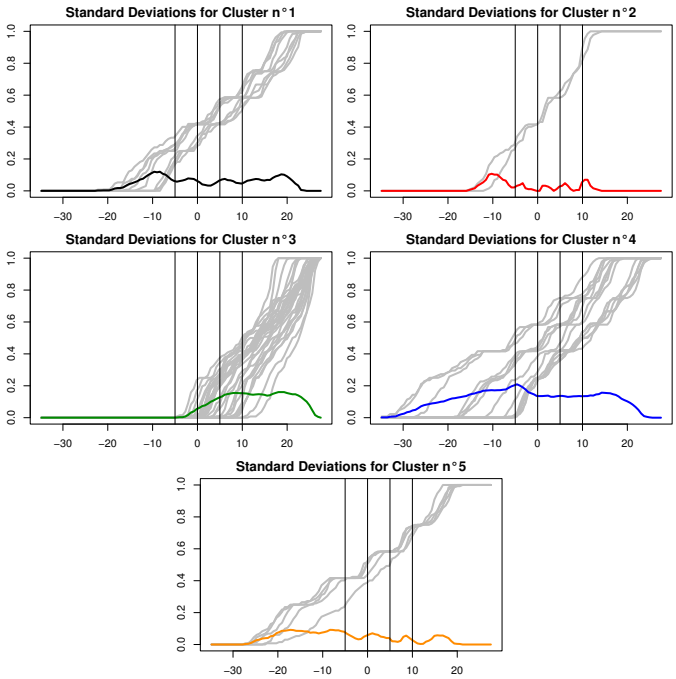


Figure 6.17: Standard deviations of the clusters of empirical cdfs of the climatological of the Chinese Climatic Data Base using the Clayton copula.

Nowadays, there is no any deterministic algorithm to proceed to this choice. But in [Cuvelier and Noirhomme-Fraiture, 2008a], we gave the results of an experimentation performed on synthetic datasets of cdfs, where the goal was to try to determine heuristics, or guidelines, for helping the choice of the set $\{t_1 \ldots, t_n\}$. In this experimentation, we have build several datasets of *cdfs* with very distinct clusters of *cdfs*, then we have applied the dynamical clustering, always with the same parameters and starting partition (necessary to compare the results of the classifications), but varying the number and the locations of the t_i[2]. And, even if it is very unusual in unsupervised classification, we have compared each resulting classification with the known a priori clusters. Then, using basic statistics, we issued the following heuristics based on the visualization of the data. Choose the set $\{t_1 \ldots, t_n\}$ such

1. there is a maximum of distinct data for each t_i,
2. there is, if it is possible, a maximum of distinct groups for each t_i,
3. the size of the set $\{t_1 \ldots, t_n\}$ is minimal.

[2]But we have chosen to use only equidistant t_i.

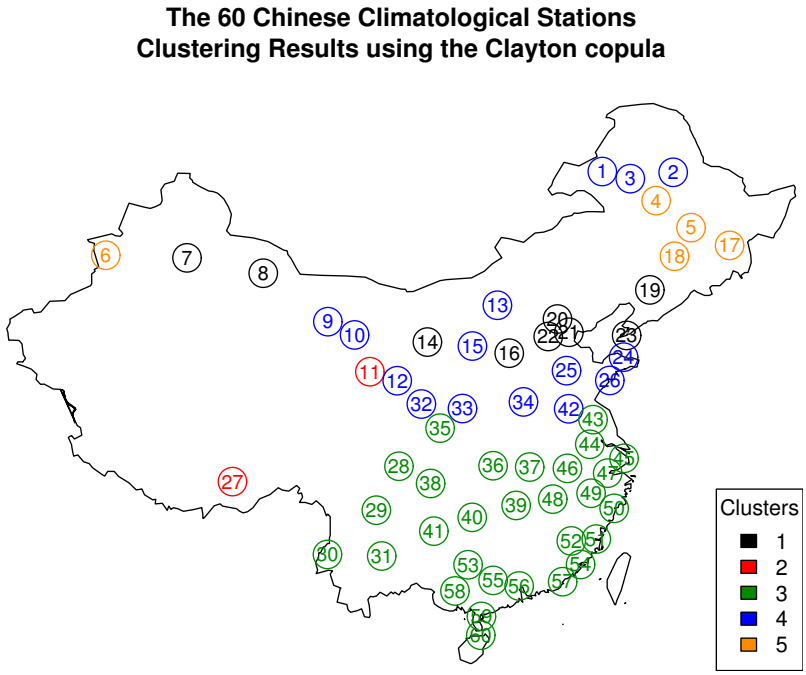


Figure 6.18: Map of the clustering results on empirical cdfs of the climatological of the Chinese Climatic Data Base using the Clayton copula.

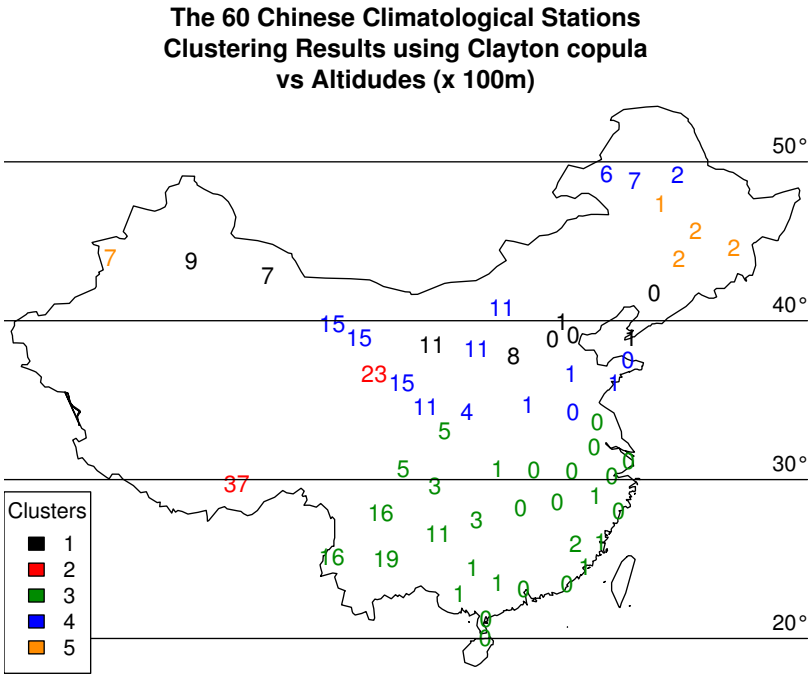


Figure 6.19: Map of the clustering results on empirical cdfs of the climatological of the Chinese Climatic Data Base using the Clayton copula, with additional informations about altitudes and latitudes.

Of course, there is an infinity of choices, but our experience leads us to remark that the first heuristic is the more important, because, firstly there is no interest to choose a t_i where there is only a few distinct data, and secondly if all data of a cluster are equals for a given t_i, then the algorithm fails (i.e. the algorithm stop without any result) when it try to estimate the standard deviation of the cluster for this t_i, mandatory to estimate the density of the cluster in this t_i.

In the case of the Chinese Climatic Data Base, if we look closely to the figure 6.5, we see that, it is difficult to follow the second heuristic, because we do not see clearly any group in the data. But the first heuristic leads us to remark that: greater will be a t_i (near 20 degrees), greater will be the probability of having a cluster with a null standard deviation for this t_i. In the same way, lesser will be a t_i (near -30 degrees), greater will be the probability of having a cluster with a null standard deviation for this t_i. That is why we have decided to choose our set of t_i in the following interval:$[-5\,°\mathrm{C}, 10\,°\mathrm{C}]$. Then we have chosen to use 4 equidistant values for the t_i: $\{-5\,°\mathrm{C}, 0\,°\mathrm{C}, 5\,°\mathrm{C}, 10\,°\mathrm{C}\}$.

As in the preceding case, we show the results in the following figures:

- the figure 6.13 shows the overall clustered cdfs,
- the figure 6.14 shows a more detailed view: each cluster of cdf is plotted in its associated color and the remaining data are plotted in grey,
- the figure 6.15 shows each cluster (plotted in grey) and its center (functional mean) in the associated color,
- the figure 6.16 shows the centers of the clusters in the same graph,
- the figure 6.17 shows each cluster (plotted in grey) and its functional standard deviations,
- the figure 6.18 represents the stations on the map of China with the number of the station in the color of its cluster,
- and the figure 6.19 represents the stations on the map of China with an altitude index (altitude divided by 100m) in the color of its cluster and we have added the 20th, 30th, 40th and 50th parallels.

The set $\{-5\,°\mathrm{C}, 0\,°\mathrm{C}, 5\,°\mathrm{C}, 10\,°\mathrm{C}\}$ is plotted in all figures, except for the last one.

The interpretation of the result is not so clear than in the QAMML case. Firstly, we see in figure 6.13 that, except for cluster number 3, clusters seem "interlaced". This fact is even more obvious in the figure 6.16: clusters 1, 2, 4 and 5 have close centers, but with different slopes, while using QAMML we see in the figure 6.9 that the clusters have distinct centers, with different slopes.

Geographical interpretations of the result in figures 6.18 and 6.19 are less clear than in the QAMML case. The link between clusters, altitude and latitude seems broken.

The figure 6.17 is interesting, because it shows that the algorithm, as expected, minimizes the intra-cluster variance for the set $\{t_1, \ldots, t_n\}$, but ignores what it happens in $\mathcal{D} \setminus \{t_1, \ldots, t_n\}$, and then intra-cluster variances can be higher than in $\{t_1, \ldots, t_n\}$.

Finally it would be artificial to proclaim, based on results obtained on one dataset, that the use of QAMML gives always better results in the unsupervised classification of *cdfs*. However, the competitive advantages of the use of QAMML, are that it is unnecessary to choose a restriction set $\{t_1, \ldots, t_n\}$ in the complete domain $\mathcal{D}$, and the fact that QAMML models take directly into account the problem of a null standard deviation on some subset of the initial domain (cf. remark 5.7.1).

Original contribution(s) 12. *In this section we have shown that a classical algorithm of clustering in multidimensional data can be used with success in the functional and symbolic data analysis, when the probability distribution is a QAMML distribution. QAMML distributions and Gâteaux density permit to build homogeneous clusters of functional data, and have competitive advantages in comparison with classical multidimensional models.*

6.3 Bayesian Supervised Classification

Another classical task of data analysis is the supervised classification: assign new individuals to their group, knowing the value(s) of the explicative variable(s).

Let $(u_i, y_i)_{i=1,\ldots,n}$, be n independent observations of the pair of random variables $(\mathcal{X}, Y)$, valued in $\mathcal{I}^{\mathcal{D}} \times \{1, \ldots, K\}$, where $y_i = k$ means that the i^{th} observation belongs to the k^{th} group. The problem is the following: given x_i, what is the best value y_i associated to x_i ?

Several approaches exist, but the most popular are the *Support Vector Machines (SVM)* (see [Vapnik, 1995], [Scholkopf and Smola, 2002] or [Herbich, 2002]) and the *Bayes Classification Rule* (see [Hastie et al., 2001] or [Devroye et al., 1996]). In our probabilistic context, Bayes rules can be directly used.

With the Bayes rule, given a random variable u, we have firstly to esti-

mate the K posterior probabilities, defined as follow

$$\begin{aligned} p_k(u) &= P(Y = k|\mathcal{X} = u) && (6.13)\\ &= \frac{P(\mathcal{X} = u|Y = k) \cdot P(Y = k)}{P(\mathcal{X} = u)} && (6.14)\\ &= \frac{P(\mathcal{X} = u|Y = k) \cdot P(Y = k)}{\sum_{i=1}^{K} P(\mathcal{X} = u|Y = i) \cdot P(Y = i)} && (6.15)\\ &= \frac{f_k(u) \cdot \pi_k}{\sum_{i=1}^{K} f_i(u) \cdot \pi_i} && (6.16) \end{aligned}$$

where $\pi_k = P(Y = k)$ and $f_k(u) = P(\mathcal{X} = u|Y = k)$. Once these K probabilities are estimated by $(\widehat{p}_1(u), \ldots, \widehat{p}_K(u))$, u is assigned to the group with the highest posterior probability:

$$\widehat{y}(u) = \arg\max_k p_k(u). \tag{6.17}$$

Let us remark that the above expression is equivalent

$$\widehat{y}(u) = \arg\max_k f_k(u) \cdot \pi_k \tag{6.18}$$

since $P(\mathcal{X} = u)$ has the same value for all groups for a given u.

In [Ferraty and Vieu, 2003] and [Ferraty and Vieu, 2006] a nonparametric estimation of the posterior probabilities is proposed. This method, similar to the density estimation seen in section 4.2, is based on a kernel estimation:

$$\widehat{p}_k(u) = p_{k,h}(u) = \frac{\sum_{\{j:Y_j=k\}} k(h^{-1}d(u, \mathcal{X}_j))}{\sum_{i=1}^{n} k(h^{-1}d(u, \mathcal{X}_i))} \tag{6.19}$$

where k, called the kernel, is a decreasing positive function, and d is a semi-metric[3].

In the QAMML framework we propose naturaly to use the Gâteaux density to compute $f_k(u)$:

$$p_k(u) = \frac{f_{\mathcal{X}_k,\mathcal{D},s}(u) \cdot \pi_k}{\sum_{i=1}^{K} f_{\mathcal{X}_i,\mathcal{D},s}(u) \cdot \pi_i}. \tag{6.20}$$

If the surface of density $G_{\mathcal{X},\mathcal{D}}$ is given by the normal case (see expression (4.4) and (4.5), p.55) we are in a parametric case, and we will use the following direction $s = \sigma$ (cf. theorem 5.8.6).

And if the surface of density $G_{\mathcal{X},\mathcal{D}}$ is given by the estimation case (see expression (4.6) and (4.7), p.55) we are then in a semi parametric case, and in then we will use the following direction $s = h$ (cf. expression(5.148)).

[3] A semi-metric d is similar to a metric, but such $d(x, y) = 0 \nRightarrow x = y$.

We have performed this Bayesian classification on the spectrometric data from Tecator (see chapter 2, section 2.3). Let us recall that, these data consist in 100 channels of spectrum absorbance (wavelength from 850 nm to 1050 nm). The goal is to distinguish data with more than 20% of fat content, from data with less than 20% of fat content. We have performed a 10-fold cross validation on these data, the first derivatives, the second derivatives using the GQAMML distributions with the following parametrization:

- Surface of distributions $G_{\mathcal{X},\mathcal{D}}$: Normal distribution,
- GQAMML copula: Clayton copula.

For the sake of comparison, using the same 10 subsets, we have also performed the 10-cross validation using the nonparametric density estimation given by expression (6.19), denoted NPCD (non parametric curve discrimination in [Ferraty and Vieu, 2003]). The table 6.3 and the figure 6.20 show the results of these experimentations.

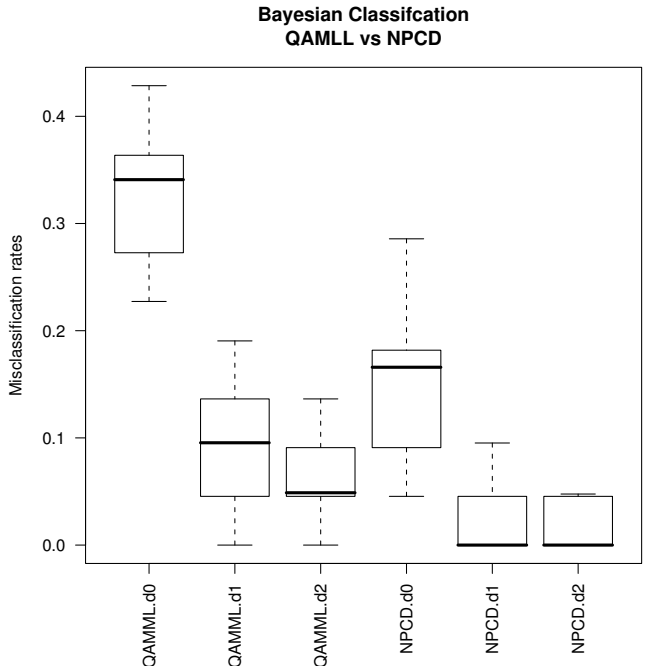


Figure 6.20: Misclassification rates of a 10-fold cross validation of the Bayesian supervised classification of the Tecator data (see p.21).

We see in table 6.3 that the best classification is given using the second derivatives and is performed by the NPCD. The best results for the QAMML distribution is almost 6% when performing, also on the second derivative, but it is well known that these derivatives of these data contain the more

Method	mean	median	sd
QAMML d0	33.51%	34.09%	06.26%
QAMML d1	09.26%	09.54%	06.21%
QAMML d2	05.97%	04.88%	04.29%
NPCD d0	15.31%	16.59%	07.59%
NPCD d1	01.86%	00.00%	03.29%
NPCD d2	01.83%	00.00%	02.37%

Table 6.3: Means, median and standard deviation of misclassification rates of the 10-fold cross validations using Normal QAMML distribution and non parametric functional density estimation.

interesting features [Ferraty and Vieu, 2003] and [Rossi and Villa., 2006]. Even if the results of QAMML distributions gives worse results then using NPCD, these results are promising, firstly because the use of QAMML gives, in the most interesting case, a mean misclassification rate of 6%, and secondly, because we must recall that we compare here a parametric method and a non parametric one. It seems that the choice of the normal distribution and the normal density for the surfaces of margins $G_{\mathcal{X},\mathcal{D}}$ and the surface of densities $g_{\mathcal{X},\mathcal{D}}$, is not optimal here, then it will be interesting to test the same procedure in the semi-parametric case, i.e. when the surface of margins $G_{\mathcal{X},\mathcal{D}}$ use the empirical cdf, and the surface of densities $g_{\mathcal{X},\mathcal{D}}$ use the kernel density estimation [4].

Original contribution(s) 13. *In this section we have shown that QAMML distributions and the Gâteaux density are directly usable in the classical Bayesian classification. The results on the classical Tecator dataset are promising using a complete parametric approach, and we can hope that, using the more flexible semi-parametric approach, these results could be improved.*

[4] In the preceding section, "Unsupervised Classification by Mixture Decomposition", we have used the empirical cdf and the kernel density estimation to compute respectively, the surface of margins $G_{\mathcal{X},\mathcal{D}}$ and the surface of densities $g_{\mathcal{X},\mathcal{D}}$, but the code was unsatisfactory, and was not planned to perform the supervised classification. Then, a complete code was rewritten from scratch, but the estimation case is not completely operational, mainly due to slow speed of execution. The problem is how compute quickly $\mathbf{G}_{\mathcal{X},\mathcal{D}}[t;u]$ for all $t \in \mathcal{D}$, or for a subset $t_1,\ldots,t_n$, especially, during the optimization of the (log)likelihood. A future solution is, given $u_1,\ldots,u_N$, to compute once $\mathbf{G}_{\mathcal{X},\mathcal{D}}[t_1;u_j],\ldots,\mathbf{G}_{\mathcal{X},\mathcal{D}}[t_n;u_j]$ for all j, and then search for the optimal parameter of the generator ϕ by maximum likelihood, using directly these values.

6.4 Functional Confidence Intervals

In the univariate case, a classical statistical use of the *cdf*'s is the construction of confidence intervals for estimations of parameters. In the functional case, even if [Ramsay and Silverman, 2005] build confidence intervals based on the basis functions, there is very few result on this field.

Let us remember the functional mean, variance and standard deviation of a given functional sample $u_1, \ldots, u_N$:

$$m(t) = \frac{1}{N} \sum_{i=1}^{N} u_i(t) \tag{6.21}$$

$$s^2(t) = \frac{1}{N-1} \sum_{i=1}^{N} [u_i(t) - m(t)]^2 \tag{6.22}$$

and the standard deviation function $s(t)$ is naturally the square root of the variance function.

Let us deal with the easiest case: the estimation of the functional mean $\mu(t)$ in the normal case[5], i.e. when the surface of margins $G_{\mathcal{X},\mathcal{D}}$ is given by the expression (4.4) with functional parameters $\mu, \sigma \in \mathcal{I}^{\mathcal{D}}$:

$$G_{\mathcal{X},\mathcal{D}}(t, y) = F_{\mathcal{N}(\mu(t),\sigma(t))}(y).$$

In other words $\mathcal{X}_t \sim \mathcal{N}(\mu(t), \sigma(t))$.

Now, like in the classical case, we search for two functional limits L and U, both belonging to $\mathcal{I}^{\mathcal{D}}$, and such:

$$P\left[L \leq_{\mathcal{D}} \mathcal{Y} \leq_{\mathcal{D}} U\right] = 1 - \alpha \tag{6.23}$$

for a chosen $\alpha \in]0, 1[$.

And , again like in the classical case, let us start considering $\mathcal{M} = \dfrac{m - \mu}{\sigma/\sqrt{n}}$, (where m, μ, σ and $\mathcal{M} \in \mathcal{I}^{\mathcal{D}}$).

Then for two functions $A, B \in \mathcal{I}^{\mathcal{D}}$ such $B \leq_{\mathcal{D}} A$, we have that

$$\begin{aligned} P\left[B \leq_{\mathcal{D}} \mathcal{M} \leq_{\mathcal{D}} A\right] &= P\left[B \leq_{\mathcal{D}} \frac{m - \mu}{\sigma/\sqrt{n}} \leq_{\mathcal{D}} A\right] && (6.24) \\ &= P\left[m - A\frac{\sigma}{\sqrt{n}} \leq_{\mathcal{D}} \mu \leq_{\mathcal{D}} m - B\frac{\sigma}{\sqrt{n}}\right]. && (6.25) \end{aligned}$$

Then, we can state the problem in the following manner: we search for $A, B \in \mathcal{I}^{\mathcal{D}}$, with $B \leq_{\mathcal{D}} A$, such

$$P\left[m - A\frac{\sigma}{\sqrt{n}} \leq_{\mathcal{D}} \mu \leq_{\mathcal{D}} m - B\frac{\sigma}{\sqrt{n}}\right] = 1 - \alpha. \tag{6.26}$$

[5]We will suppose that $\sigma(t)$ is known, and $\mu(t)$ unknown.

But

$$P\left[B \leq_{\mathcal{D}} \mathcal{M} \leq_{\mathcal{D}} A\right] = P\left[\mathcal{M} \leq_{\mathcal{D}} A\right] - P\left[\mathcal{M} \leq_{\mathcal{D}} B\right] \quad (6.27)$$
$$= \mathcal{F}_{\mathcal{M},\mathcal{D}}(A) - \mathcal{F}_{\mathcal{M},\mathcal{D}}(B) \quad (6.28)$$

then, a solution is to find A and B such

$$\mathcal{F}_{\mathcal{M},\mathcal{D}}(A) = 1 - \alpha^{'} \quad (6.29)$$
$$\mathcal{F}_{\mathcal{M},\mathcal{D}}(B) = \alpha^{''} \quad (6.30)$$

with $\alpha^{'} + \alpha^{''} = \alpha$.

Then if we set

$$A = Q_{\mathcal{M},\mathcal{D},1-\alpha'} \quad (6.31)$$
$$B = Q_{\mathcal{M},\mathcal{D},\alpha''} \quad (6.32)$$

the problem is solved, because we have

$$P\left[m - Q_{\mathcal{M},\mathcal{D},1-\alpha'}\frac{\sigma}{\sqrt{n}} \leq_{\mathcal{D}} \mu \leq_{\mathcal{D}} m - Q_{\mathcal{M},\mathcal{D},\alpha''}\frac{\sigma}{\sqrt{n}}\right] \quad (6.33)$$
$$= P\left[Q_{\mathcal{M},\mathcal{D},\alpha''} \leq_{\mathcal{D}} \mathcal{M} \leq_{\mathcal{D}} Q_{\mathcal{M},\mathcal{D},1-\alpha'}\right] \quad (6.34)$$
$$= \mathcal{F}_{\mathcal{M},\mathcal{D}}(Q_{\mathcal{M},\mathcal{D},1-\alpha'}) - \mathcal{F}_{\mathcal{M},\mathcal{D}}(Q_{\mathcal{M},\mathcal{D},\alpha''}) \quad (6.35)$$
$$= 1 - \alpha^{'} - \alpha^{''} \quad (6.36)$$
$$= 1 - \alpha. \quad (6.37)$$

It remains to determine $Q_{\mathcal{M},\mathcal{D},1-\alpha'}$ and $Q_{\mathcal{M},\mathcal{D},\alpha''}$. But let us recall that a functional quantile associated to a frv is only determined by the surface of margins, and then is only determined by what it happens in each $t \in \mathcal{D}$, and in our case we have the following implications

$$\mathcal{X}_t \sim \mathcal{N}(\mu(t), \sigma(t)) \Rightarrow m_t \sim \mathcal{N}(\mu(t), \frac{\sigma(t)}{\sqrt{(n)}}) \quad (6.38)$$
$$\Rightarrow \mathcal{M}_t \sim \mathcal{N}(0,1) \quad (6.39)$$

and then the surface of margins $G_{\mathcal{M},\mathcal{D}}$, of the *frv* $\mathcal{M}$ is given by

$$G_{\mathcal{M},\mathcal{D}}(t,y) = F_{\mathcal{N}(0,1)}(y). \quad (6.40)$$

Then with this relation (6.40), we have that $Q_{\mathcal{Y},\mathcal{D},p}(t) = Q_G(1-\alpha')$, for any $t \in \mathcal{D}$ and any $p \in [0,1]$, where Q_G is the quantile of the Gaussian distribution.

And finally

$$P\left[m - Q_G(1-\alpha')\frac{\sigma}{\sqrt{N}} \leq_{\mathcal{D}} \mu \leq_{\mathcal{D}} m - Q_G(\alpha'')\frac{\sigma}{\sqrt{N}}\right] = 1 - \alpha \quad (6.41)$$

with m, μ and σ belonging to $\mathcal{I}^{\mathcal{D}}$.

The interesting, but disconcerting, conclusion is the fact that the confidence bounds, according to the QAMML distributions, are built pointwise, but the expression (6.41) gives the probabilistic justification for this construction, like in the classical univariate case.

For the sake of illustration, figures 6.21 to 6.23 show the confidence intervals of the mean for the Tecator dataset, using the expression (6.30), firstly for the original data, and then for the first and second derivatives. These figures draw the data in gray, and the functional confidence interval in shaded black. The top centered of the figures show the confidence interval for $\mu \in \mathcal{I}^{\mathcal{D}}$ for all data, and bottom figures show intervals for the two known groups: the low fat content and the high fat content. In all case, we can see clearly the (eventually slight) difference between the two groups.

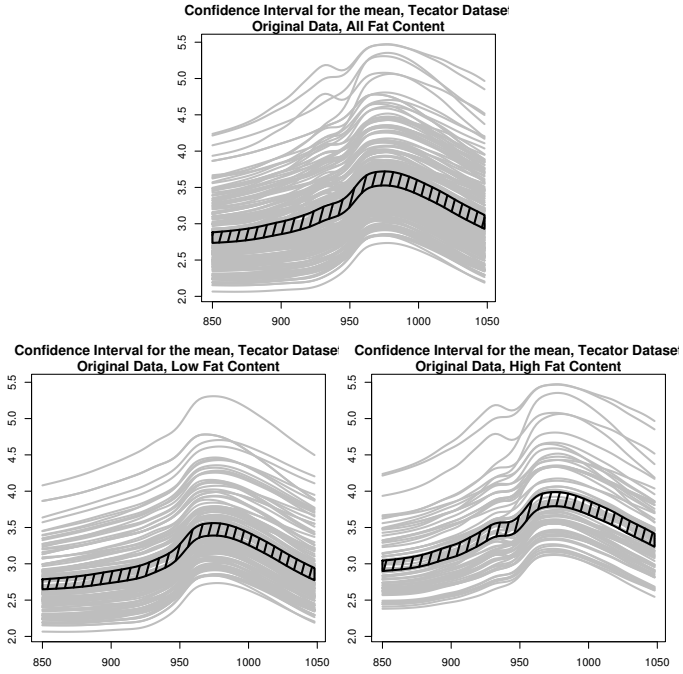


Figure 6.21: The Confidence Bounds for the mean of Tecator data (N=215).

In section 2.3, we have introduced the Phonemes Data (see figure 2.12). We have used expression (6.30) again to build the confidence interval for the mean of these data, but, due to the size of the sample ($N = 4509$) the results shown in figure 6.25 look like a single line, but are confidence intervals. For the sake of illustration we show in figure 6.24 the functional quantiles: $Q_{\mathcal{X},\mathcal{D},0.1}$, $Q_{\mathcal{X},\mathcal{D},0.2}$, $Q_{\mathcal{X},\mathcal{D},0.3}$, $Q_{\mathcal{X},\mathcal{D},0.4}$, $Q_{\mathcal{X},\mathcal{D},0.5}$, $Q_{\mathcal{X},\mathcal{D},0.6}$, $Q_{\mathcal{X},\mathcal{D},0.7}$, $Q_{\mathcal{X},\mathcal{D},0.8}$ and $Q_{\mathcal{X},\mathcal{D},0.9}$, using the empirical *cdf* (see expression (4.6)).

Original contribution(s) 14. *We have shown, that the use of QAMML*

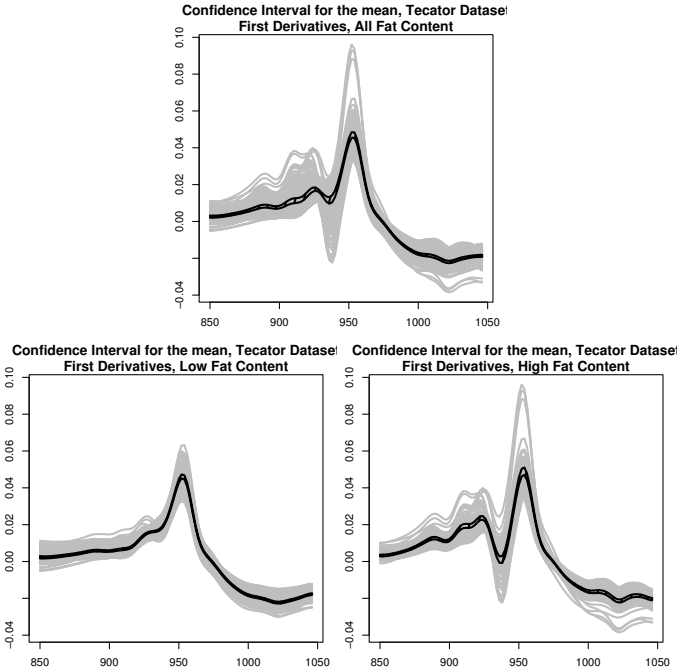


Figure 6.22: The Confidence Bounds for the mean of Tecator data, First Derivatives (N=215).

distributions and the functional quantiles $Q_{\mathcal{X},\mathcal{D}}$, give, in the functional case, a solution to the classical statistical problem of building confidence intervals for functional parameters. And, this solution give exactly a confidence interval which contains the desired functional parameter with a chosen a priori probability $1-\alpha$, like in the classical case.

6.5 Conclusions

Even if QAMML distributions and the Gâteaux density bring new theoretical tools for probability in the functional and symbolic data analysis frameworks, the starting point of these researches, was a practical problem: the unsupervised classification of probability distributions by mixture decomposition in Symbolic Data Analysis proposed in [Diday, 2002]. During all the developments of the theoretical model of QAMML, we have kept in mind this classification use. Moreover, we have decided to test QAMML distributions for the supervised classification. The primary goal of these classification tasks was not to prove that the use of QAMML is better than other methods, but to prove that existing methods in the classical multidimensional case are directly usable in the functional case. And this proof holds, and open new directions of research: what other classical existing

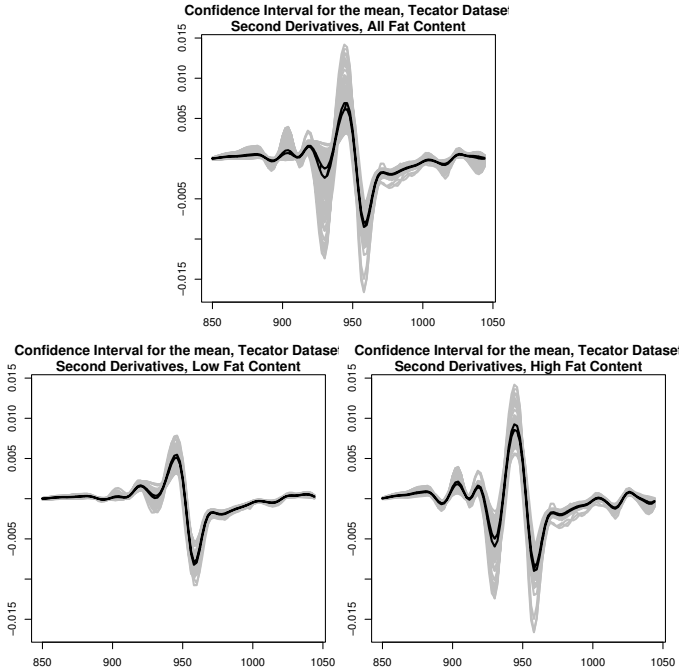


Figure 6.23: The Confidence Bounds for the mean of Tecator data, Second Derivatives (N=215).

procedures can be directly used with QAMML, how improve the results actually obtained,...

Moreover, in statistics, QAMML distributions bring a formal definition for the pointwise confidence interval build upon functional quantiles, and again we can hope that, it is a starting point to develop statistical tests in the functional case.

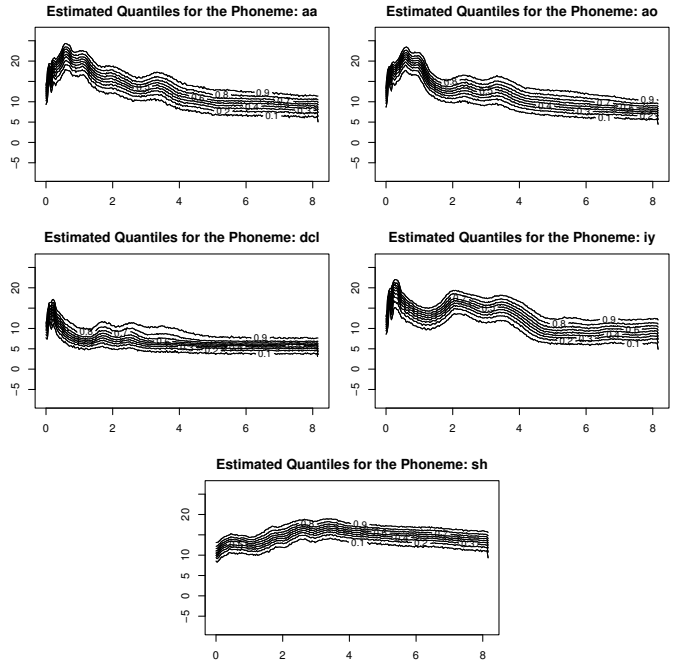


Figure 6.24: The estimated Quantiles $Q_{\mathcal{X},\mathcal{D}}(p;t)$ for Phonemes data (see fig. 2.12, p.22).

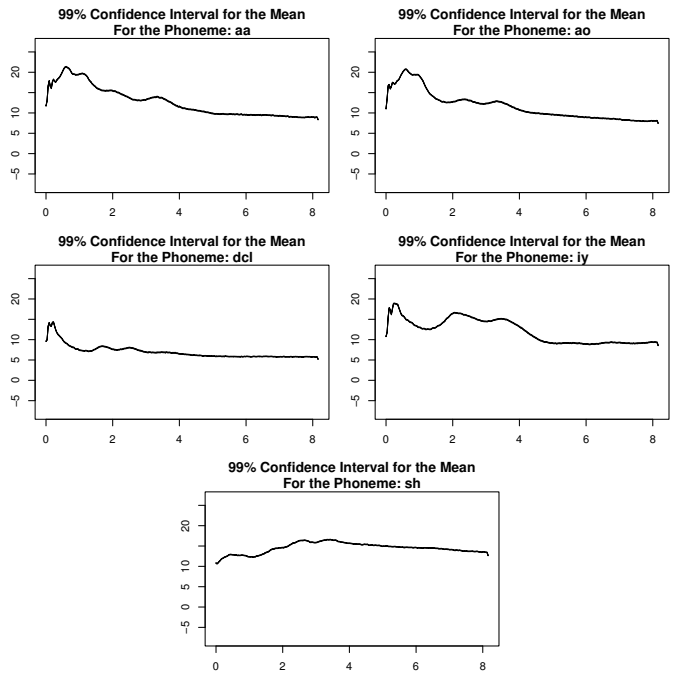


Figure 6.25: The Confidence Bounds for mean of Phonemes data.

Chapter 7

Conclusions and Perspectives

> Vous savez ce que c'est que la recherche: on part sur une question et on trouve en cours de route des faits qui vous en posent une autre.
>
> Philippe Meyer

In this work we propose an elegant solution to an arduous and complex problem: define and compute easily probability distributions in the infinite dimensional space of functional random variables (called *frv*).

For this, after having recalled the very basics of probability and real random variables in finite dimensional spaces, we have defined clearly the concept of probability distribution for functional random variable, using the non strict piecewise order on functions. We have called these distributions *fcdf* (functional cumulative distribution function). It should be noted that *fcdfs* are not calculable with the finite dimensional distributions used in the Daniell-Kolmogorov's extension theorem.

Then, in a first attempt to define the *fcdf* like the limit of finite dimensional distributions, we have, for a given *frv* $\mathcal{X}$, defined the surfaces of margins and densities, which give the distribution and the density of $\mathcal{X}_t$ for any t in the domain $\mathcal{D}$. The definition of the surface of margins leads us to define also the concept of functional quantile which is a function with a constant probability value for any margin. Like it is easy to obtain the margins for a given functional random variable, we have recalled how copulas decompose finite dimensional distributions, separating the margins from the dependence structures (the copula). We have focused specifically on the family of Archimedean copulas, and we have given formulas to calculate the joint density, valid for any finite dimensions, for the following Archimedean families: Clayton, Frank, Gumbel-Hougaard and Joe. We have achieved this attempt showing that for Archimedean copulas, and more broadly for almost all kinds of copulas, the limit of finite dimensional distributions of a *frv* is always equal to zero, and then can not lead to an *fcdf* directly defined

in the infinite dimensional space of functions. Then, for a given function u, it is impossible to calculate the *fcdf* of an *frv* $\mathcal{X}$ in u, as being the limit of finite dimensional distributions computed in the discretized values of u. In other words, it is impossible to reach the *fcdf* of an *frv* in the classical framework, where discretized values of the *frv* are considered as vectors of univariate random variables. That is why we have proposed an adapted framework for the *fcdf* of an *frv*.

This new framework is based on the following principles. Firstly we always consider the function as a whole, even if the function is only known through discretized values. In this last case we have defined the quantilized functions which are piecewise functional quantiles. We have also shown that, when the surface of margins is strictly increasing for each t, then any function is the limit of at most a sequence of quantilized functions. The quantilized function of such a sequence are quantilized versions of the limit function. Functional quantiles determine the second principle: the *fcdf* of a functional quantile of order p must be equal to p, and when we work on discretized versions of functions, this value cannot depend of the number of discretized values. The third principle, states that the *fcdf* of a quantilized function must be calculated as the aggregation of the probability values of the quantiles which constitute the quantilized function. And finally, the *fcdf* of a function must be the limit of the *fcdfs* of any sequence of quantilized functions converging toward the considered function. We have called *fractal and quantile based distribution* any *fcdf* which fulfill these principles.

Then, we have shown that the quasi-arithmetic mean of margins of a discretized version of a *frv* gives a fractal and quantile based distribution, if the generator of the quasi-arithmetic mean is also a generator of an Archimedean copula of dimension higher than two. In the discretized case we have called these new kinds of distributions *Quasi-Arithmetic Means of Margins (QAMM)*. By the last principle of this new framework, the limit of QAMM distributions of a converging sequence of quantilized functions leads us directly to define the *Quasi-Arithmetic Means of Margins Limit (QAMML)* distributions. QAMML distributions are fractal and quantile based distributions which are directly defined in the infinite dimensional space of functions. Due to the properties of Archimedean generators, QAMML distributions are not able to capture all kind of dependence structures, but only the comonotonicity and not the anticomonotonicity. That is why we were led to define a generalized version of the QAMML distributions, the GQAMML distributions which are able to capture the comonotonicity and the anticomonotonicity.

Once we have reached our first goal, we had to deal with a new problem, because a lot of data analysis applications use in fact the density of the probability distributions. And, in an infinite dimensional space, the classical definition of the density do not hold anymore. To solve this problem we have proposed a definition of the density function for the QAMML dis-

tributions which leads us to use a directional derivative called the Gâteaux derivative: it is the derivative of functional[1] computed in a function, in the direction of another function. Using this type of derivative we have defined the Gâteaux density of the QAMML distributions and we have shown that when the distribution used for the surface of margins has a scale parameter, this latter one is the direction to choose in the Gâteaux derivative. And, under some conditions, we can choose the same direction when we have to estimate the surface of margins in an empirical and kernel way. We also gave the Gâteaux density for the GQAMML distributions. Then, for QAMML and GQAMML distributions, we have also defined a multi-steps parameter estimation procedure inspired by the IFM method used for copulas.

We have ended this work with three successful applications of (G)QAMML distributions and the Gâteau density: an unsupervised classification by mixture decomposition in the Symbolic Data Analysis framework, a Bayesian supervised classification of a classical functional dataset, and a functional confidence intervals calculation.

We have chosen these applications because they use directly probability distributions and density functions. And it was exactly our goal: to define a probability distribution for functional data directly usable in algorithm and methods, previously designed for the classical multidimensional case. Then, QAMML distributions open the door to the experimentation of other classical methods in the functional data analysis framework, and for probability distributions in the symbolic data analysis framework.

In addition to these experiments, QAMML distributions open many questions and outlooks of development.

Firstly, we use for QAMM and QAMML distributions, generators of Archimedean copulas, and, in the copulas framework, for the main families of Archimedean copulas there is a direct interpretation of the value of the generator parameter (see examples 4.4.1 to 4.4.4, pp 108-109). Can we find such an interpretation for the parameter value in the QAMM(L) case? And what becomes the QAMML distributions in bounds values of the parameters?

Secondly, we have shown how the use of generators of Archimedean copulas is essential to assure that QAMM distributions are joint distributions, but once we have defined QAMML distributions, and if we deal uniquely with QAMML distributions, can we use other kinds of generators?

Thirdly, apart from standard deviation, is there other possible directions usable for the Gâteaux derivative in the Gâteaux density?

Fourthly, let us recall how, for real random variables, parametric densities are used as kernel in nonparametric density estimation. Is it possible to use the Gâteaux density as kernel in a non parametric density estimation for functional data?

[1]Functional is the name given to functions of functions.

Fifthly, when we define the QAMML distribution in expression 5.63, we can define it equivalently by the following expression

$$\mathcal{F}_{\mathcal{X},\mathcal{D},\phi}(u) = \psi\left[\int_{\mathcal{D}} \phi\left(\mathbf{G}_{\mathcal{X},\mathcal{D}}\left[t;u\right]\right) dU_{\mathcal{D}}(t)\right] \tag{7.1}$$

where $U_{\mathcal{D}}$ is the uniform distribution over $\mathcal{D}$. But other distributions over $\mathcal{D}$ can be envisaged depending of the situation where the QAMML distribution is used.

Sixthly, in the section 5.7, we have introduced the GQAMML distributions as a matter of flexibility to capture the dependence structures, but this generalization of QAMML distributions can be used in other situations. Let us suppose that we know the distribution of $\mathcal{X}$ on the interval $A = [a, b]$: $\mathcal{F}_{\mathcal{X},A,\phi_A}$. And let us suppose that, after a while, we know also the distribution of $\mathcal{X}$ on $B = [b, c]$, then we can update our knowledge for $\mathcal{D} = A \cup B$, by computing $\mathcal{F}_{\mathcal{X},\mathcal{D},\phi_{\mathcal{D}}}$ using the GQAMML expression:

$$\mathcal{F}_{\mathcal{X},\mathcal{D},\phi_{\mathcal{D}}}(u) = \psi_{\mathcal{D}}\left(p_A \cdot \phi_{\mathcal{D}}\left[\mathcal{F}_{\mathcal{X},A,\phi_A}(u)\right] + p_B \cdot \phi_{\mathcal{D}}\left[\mathcal{F}_{\mathcal{X},B,\phi_B}(u)\right]\right) \tag{7.2}$$

with $p_A, p_B \in]0, 1[$ and $p_A + p_B = 1$.

A direct utility could be found in the study of streaming data, where the flow of data evolves continuously. If, in the above formula, t represent the time, A the known past interval and B the new known interval, then it is possible to update the distribution of $\mathcal{X}$ over $A \cup B = [a, c]$ without recomputing what it happened in A.

And let us finish with a last interesting path of development: the functional multidimensional extension.

In functional data analysis it is well known that all usable informations are not always directly available when working on the "raw data" (i.e. the given functions). If the data are sufficiently "smooth", it, or its derivatives can provide interesting work. In our classical example of Tecator data (see figure 2.11), if we want to proceed to a supervised classification, then first and second derivatives (see figures 2.20 and 2.21) are more suited to the achievement of this task (see [Ferraty and Vieu, 2006]). Thus we can work with $\mathcal{F}_{\mathcal{X},\mathcal{D},\phi}(u)$, but also with $\mathcal{F}_{\mathcal{X}^{(1)},\mathcal{D}}(u^{(1)})$ and/or $\mathcal{F}_{\mathcal{X}^{(2)},\mathcal{D}}(u^{(2)})$. But, it could be interesting to gather all this information in one function:

$$P[\mathcal{X} \leq_{\mathcal{D}} u, \mathcal{X}^{(1)} \leq_{\mathcal{D}} u^{(1)}, \ldots, \mathcal{X}^{(n)} \leq_{\mathcal{D}} u^{(n)}].$$

In the same way if we consider more than one *frv*: $(\mathcal{X}_1, \ldots, \mathcal{X}_n)$, then it could be valuable to work on all these data together:

$$P[\mathcal{X}_1 \leq_{\mathcal{D}} u_1, \ldots, \mathcal{X}_n \leq_{\mathcal{D}} u_n]. \tag{7.3}$$

In both cases, knowing $\mathcal{F}_{\mathcal{X},\mathcal{D}}(u), \mathcal{F}_{\mathcal{X}^{(1)},\mathcal{D}}(u^{(1)}), \ldots, \mathcal{F}_{\mathcal{X}^{(n)},\mathcal{D}}(u^{(n)})$, or $\mathcal{F}_{\mathcal{X}_1,\mathcal{D}_1}(u_1), \ldots, \mathcal{F}_{\mathcal{X}_n,\mathcal{D}_n}(u_n)$, it would be natural to define the multidimen-

sional version of QAMML distributions using the functional univariate distributions and a copulas:

$$\begin{aligned} &P[\mathcal{X} \leq_{\mathcal{D}} u, \mathcal{X}^{(1)} \leq_{\mathcal{D}} u^{(1)}, \ldots, \mathcal{X}^{(n)} \leq_{\mathcal{D}} u^{(n)}] \\ &= C\left(\mathcal{F}_{\mathcal{X},\mathcal{D}}(u), \mathcal{F}_{\mathcal{X}^{(1)},\mathcal{D}}(u^{(1)}), \ldots, \mathcal{F}_{\mathcal{X}^{(n)},\mathcal{D}}(u^{(n)})\right) \end{aligned} \tag{7.4}$$

or

$$P[\mathcal{X}_1 \leq_{\mathcal{D}} u_1, \ldots, \mathcal{X}_n \leq_{\mathcal{D}} u_n] = C\left(\mathcal{F}_{\mathcal{X}_1,\mathcal{D}_1}(u_1), \ldots, \mathcal{F}_{\mathcal{X}_n,\mathcal{D}_n}(u_n)\right). \tag{7.5}$$

In the same idea it could be interesting to evaluate jointly the probability distribution of an *rrv* Y and an *frv* $\mathcal{X}$:

$$P[Y \leq r, \mathcal{X} \leq_{\mathcal{D}} u] = C\left(F_Y(r), \mathcal{F}_{\mathcal{X},\mathcal{D}}(u)\right) \tag{7.6}$$

where $r \in \mathbb{R}$ and $u \in \mathcal{D}$. The above joint distribution could lead directly to the following conditional probability

$$P[Y \leq r | \mathcal{X} \leq_{\mathcal{D}} u] = \frac{C\left(F_Y(r), \mathcal{F}_{\mathcal{X},\mathcal{D}}(u)\right)}{\mathcal{F}_{\mathcal{X},\mathcal{D}}(u)} \tag{7.7}$$

where $r \in \mathbb{R}$ and $u \in \mathcal{D}$. Then it could be possible to envisage the use of this above probability to compute the linear regression of the *rrv* Y with a given *frv* $\mathcal{X}$.

Let us recall how, in section 4.3, we explained that copulas are powerful tools to build distributions with given marginals. This construction was built upon the lemma 4.3.3, which states that, for a given random real variable X of distribution F, the distribution of the real random variable $F(X)$ is an uniform distribution on $[0, 1]$. And the proof of this lemma was based on the quantile Q of the continuous distribution F.

In the functional case we no longer have anymore an inverse for the QAMML distributions. However, if we prove that

$$P[\mathcal{F}_{\mathcal{X},\mathcal{D},\phi}(\mathcal{X}) \leq p] = p \tag{7.8}$$

for any $p \in [0, 1]$, then the same reasoning could be used. For this it "suffices" to prove that

$$\{\omega \in \Omega : \mathcal{F}_{\mathcal{X},\mathcal{D},\phi}(\mathcal{X}(\omega)) \leq p\} = \{\omega \in \Omega : \mathcal{X}(\omega) \leq_{\mathcal{D}} Q_{\mathcal{X},\mathcal{D},p}\}. \tag{7.9}$$

The following inclusion

$$\{\omega \in \Omega : \mathcal{F}_{\mathcal{X},\mathcal{D},\phi}(\mathcal{X}(\omega)) \leq p\} \supset \{\omega \in \Omega : \mathcal{X}(\omega) \leq_{\mathcal{D}} Q_{\mathcal{X},\mathcal{D},p}\} \tag{7.10}$$

is easy to prove:

$$\mathcal{X}(\omega) \leq_{\mathcal{D}} Q_{\mathcal{X},\mathcal{D},p} \Rightarrow \mathcal{F}_{\mathcal{X},\mathcal{D},\phi}(\mathcal{X}(\omega)) \leq p.$$

But the converse inclusion is not so obvious, because it is possible to have $\mathcal{F}_{\mathcal{X},\mathcal{D},\phi}(\mathcal{X}(\omega)) \leq p$ and $\mathcal{X}(\omega) \not\leq_{\mathcal{D}} Q_{\mathcal{X},\mathcal{D},p}$, i.e. $\mathcal{X}(\omega)$ and $Q_{\mathcal{X},\mathcal{D},p}$ are not comparables.

A possible path to try to prove this second inclusion it to decompose the second set of (7.9) in the following manner

$$\begin{aligned}
&\{\omega \in \Omega : \mathcal{F}_{\mathcal{X},\mathcal{D},\phi}(\mathcal{X}(\omega)) \leq p\} \\
&= \{\omega \in \Omega : \mathcal{X}(\omega) \leq_{\mathcal{D}} Q_{\mathcal{X},\mathcal{D},p}\} \\
&\cup \{\omega \in \Omega : \mathcal{X}(\omega) \not\leq_{\mathcal{D}} Q_{\mathcal{X},\mathcal{D},p}; \mathcal{F}_{\mathcal{X},\mathcal{D},\phi}(\mathcal{X}(\omega)) \leq p\}
\end{aligned} \tag{7.11}$$

and then prove that the measure of the last subset is zero:

$$P\{\omega \in \Omega : \mathcal{X}(\omega) \not\leq_{\mathcal{D}} Q_{\mathcal{X},\mathcal{D},p}; \mathcal{F}_{\mathcal{X},\mathcal{D},\phi}(\mathcal{X}(\omega)) \leq p\} = 0. \tag{7.12}$$

This functional multidimensional extension is our main direction for subsequent research because, being able to deal with many *frv* would be very interesting, and could leads us to consider jointly *frv*'s and their derivatives[2]. This could then leads us to more precision in the probabilistic modeling of functional data and very interesting applications

[2]When they exist.

Bibliography

[Abbott, 1884] Abbott, E. A. (1884). *Flatland, A Romance of Many Dimensions*.

[Aczel, 1966a] Aczel, J. (1966a). *Functional Equations in Several Variables*. Encyclopedia of Mathematics and its Applications. Cambridge University Press, Cambridge.

[Aczel, 1966b] Aczel, J. (1966b). *Lectures on Functional Equations and Their Applications*. Mathematics in Science and Engineering. Academic Press, New York and London.

[Atkinson and Han, 2001] Atkinson, K. and Han, W. (2001). *Theoretical Numerical Analysis*. texts in Applied Mathematics. Springer, New-York.

[Bartlett, 1978] Bartlett, M. S. (1978). *An introduction to stochastic processes*. Cambridge University Press, Cambridge.

[Beliakov et al., 2007] Beliakov, G., Pradera, A., and Calvo, T. (2007). *Aggregation Functions: A Guide for Practitioners*, volume 221 of *Studies in Fuzziness and Soft Computing*. Springer.

[Bernstein, 1928] Bernstein, S. N. (1928). Sur les fonctions absolument monotones. *Acta mathematica*, 51:1–66.

[Billard and Diday, 2006] Billard, L. and Diday, E. (2006). *Symbolic Data Analysis: Conceptual Statisitcs and Data Mining*. Wiley, Chichester.

[Billingsley, 1968] Billingsley, P. (1968). *Convergence of probability measures*. John Wiley & Sons.

[Bock and Diday, 2000] Bock, H. and Diday, E. (2000). *Analysis of Symbolic Data. Exploratory methods for extracting statistical information from complex data*. Springer Verlag.

[Bonferroni, 1927] Bonferroni, C. E. (1927). Sulle medie dedotte da funzioni concave. *Giornale di Matematica Finanziaria*, 9:13–24.

[Bourbaki, 1967] Bourbaki, N. (1967). *Intégration. I.* Hermann.

[Bowman et al., 1998] Bowman, A., Hall, P., and T., P. (1998). Bandwidth selection for the smoothing of distribution functions. *Biometrika*, 85:799–808.

[Bullen et al., 1988] Bullen, P., Mitrinovic, D., and Vasic, P. (1988). *Means and their inequalities*. D.Reidel Publishing Company.

[Burril, 1972] Burril, C. W. (1972). *Measure, integration and probability*. McGraw-Hill, New-York.

[Celeux and Govaert, 1992] Celeux, G. and Govaert, G. (1992). A classification em algorithm for clustering and two stochastic versions. *Computational Statistics and Data Analysis*, 14:315–332.

[Chavent et al., 2003] Chavent, M., de A. T. de Carvalho, F., Lechevallier, Y., and Verde, R. (2003). Trois nouvelles méthodes de classification automatique de données symboliques de type intervalle. *Revue de Statistique Appliquée*, 51(4):5–29.

[Chavent et al., 2006] Chavent, M., de A. T. de Carvalho, F., Lechevallier, Y., and Verde, R. (2006). New clustering methods for interval data. *Computational Statistics*, 21(2):211–229.

[Conway, 1979] Conway, D. (1979). Multivariate distributions with specified marginals. Technical Report 145, Stanford university.

[Cox and Miller, 1965] Cox, D. R. and Miller, H. (1965). *The theory of stochastic processes*. Methuen, London.

[Cramér, 1971] Cramér, H. (1971). *Mathematical Methods of Statistics*. Princeton University Press.

[Cuvelier and Noirhomme-Fraiture, 2003] Cuvelier, E. and Noirhomme-Fraiture, M. (2003). Mélange de distributions de distributions, décomposition de mélange avec la copule de clayton. In *XXXV èmes Journées de Statistiques*, pages 377–380.

[Cuvelier and Noirhomme-Fraiture, 2005] Cuvelier, E. and Noirhomme-Fraiture, M. (2005). Clayton copula and mixture decomposition. In *ASMDA 2005*, pages 699–708.

[Cuvelier and Noirhomme-Fraiture, 2006] Cuvelier, E. and Noirhomme-Fraiture, M. (2006). A probability distribution of functional random variable with a functionnal data analysis application. In *ICDM 06 - MCD 06*.

[Cuvelier and Noirhomme-Fraiture, 2007] Cuvelier, E. and Noirhomme-Fraiture, M. (2007). Classification de fonctions continues á l'aide d'une

distribution et d'une définies dans un espace de dimension infinie. In *EGC07*.

[Cuvelier and Noirhomme-Fraiture, 2008a] Cuvelier, E. and Noirhomme-Fraiture, M. (2008a). Classification de données fonctionnelles par décomposition de mélange: Apports de la visualisation dans le cas des distributions de probabilité. *Revue d'Intelligence Artificielle*, 22(3-4):421–442.

[Cuvelier and Noirhomme-Fraiture, 2008b] Cuvelier, E. and Noirhomme-Fraiture, M. (2008b). Parametric families of probability distributions for functional data using quasi-arithmetic means with archimedean generators. In *Functional and Operatorial Statistics*, collection Contributions to Statistics, pages 127–133.

[Dabo-Niang, 2002] Dabo-Niang, S. (2002). Estimation de la densité dans un espace de dimension infinie : Application aux diffusions. *C. R. Acad. Sci. Paris Ser.*, 334:213–216.

[Dabo-Niang, 2004] Dabo-Niang, S. (2004). Kernel density estimator in an infinite-dimensional space with a rate of convergence in the case of diffusion process. *Applied Mathemathics Letters*, 17:381–386.

[Daniell, 1919a] Daniell, P. J. (1919a). Functions of limited variation in an infinite number of dimensions. *Annals of Mathematics*, 21(2):30–38.

[Daniell, 1919b] Daniell, P. J. (1919b). Integrals in an infinite number of dimensions. *Annals of Mathematics*, 20(2):281–288.

[de Boor, 2002] de Boor, C. (2002). *A Practical Guide to Splines. Second Edition.* Springer, New York.

[De Finetti, 1931] De Finetti, B. (1931). Sul concetto di media. *Giornale dell' Instituto Itialiano degli Attuari*, 2:369–396.

[Dempster et al., 1977] Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society (Series B)*, 39(1):1–38.

[Devroye et al., 1996] Devroye, L., Gy$\acute{}$orfi, L., and Lugosi, G. (1996). *A Probabilistic Theory of Pattern Recognition*. Springer-Verlag.

[Diday, 2001] Diday, E. (2001). A generalization of the mixture decomposition problem in the symbolic data analysis framework. *Cahiers du CEREMADE*, 2001(12):1–14.

[Diday, 2002] Diday, E. (2002). Mixture decomposition of distributions by copulas. In *Classification, Clustering and Data Analysis*, pages 297–310. Springer, Verlag.

[Diday and Noirhomme-Fraiture, 2008] Diday, E. and Noirhomme-Fraiture, M. (2008). *Symbolic Data Analysis And The Sodas Software*. Wiley.

[Diday et al., 1974] Diday, E., Schroeder, A., and Ok, Y. (1974). The dynamic clusters method in pattern recognition. In *IFIP Congress*, pages 691–697.

[Diday and Vrac, 2005] Diday, E. and Vrac, M. (2005). Mixture decomposition of distributions by copulas in the symbolic data analysis framework. *Discrete Applied Mathematics*, 147(1):27–41.

[Doob, 1953] Doob, J. L. (1953). *Stochastic Processes*. John Wiley & Sons.

[Ferraty and Vieu, 2003] Ferraty, F. and Vieu, P. (2003). Curves discrimintation: a non parametric functional approach. *Computational Statistics & Data Analysis*, 44:161–173.

[Ferraty and Vieu, 2006] Ferraty, F. and Vieu, P. (2006). *Nonparametric functional data analysis : theory and practice*. Springer.

[Fodor and Roubens, 1994] Fodor, J. and Roubens, M. (1994). *Fuzzy Preference Modelling and Multicriteria Decision Support*. Kluwer Academic Publishers.

[Fréchet, 1951] Fréchet, M. (1951). Sur les tableaux de corrélation dont les marges sont données. *Ann. Univ. Lyon, Sat. A*, 9:53–77.

[Galambos, 1978] Galambos, J. (1978). *The Asymptotic Theory of Extreme Order Statistics*. John Wiley & Sons, New-York.

[Gâteaux, 1913] Gâteaux, R. (1913). Sur les fonctionnelles continues et les fonctionnelles analytiques. *Comptes rendus de l'academie des sciences (Paris)*, 157:325–327.

[Gâteaux, 1919a] Gâteaux, R. (1919a). Fonctions d'une infinité de variables indépendantes. *Bulletin de la Société Mathmatique de France*, 47:70–96.

[Gâteaux, 1919b] Gâteaux, R. (1919b). Sur la notion d'intégrale dans le domaine fonctionnel et sur la théorie du potentiel. *Bulletin de la Société Mathmatique de France*, 47:47–70.

[Gâteaux, 1922] Gâteaux, R. (1922). Sur diverses questions de calculs fonctionnel. *Bulletin de la Société Mathmatique de France*, 50:1–37.

[Genest, 1987] Genest, C. (1987). Frank's family of bivariate distributions. *Biometrika*, 74:549–555.

[Genest et al., 1995] Genest, C., Ghoudi, K., and Rivest, L.-P. (1995). A semiparametric estimation procedure for dependence parameters in multivariate families of distributions. *Biometrika*, 83(3):543–552.

[Genest and MacKay, 1986a] Genest, C. and MacKay, J. (1986a). Copules archimédiennes et familles de lois bidimensionnelles dont les marges sont données. *Canad. J. Statist.*, 14:145–159.

[Genest and MacKay, 1986b] Genest, C. and MacKay, J. (1986b). The joy of copulas : Bivariate distributions with uniform marginals. *Amer. Statist.*, 40:280–285.

[Genest and Rivest, 1993] Genest, C. and Rivest, L. (1993). Statistical inference procedures for bivariate archimedean copulas. *Journal of the American Statistical Association*, 88:1034–1043.

[Genest and Werker, 2002] Genest, C. and Werker, B. (2002). Conditions for the asymptotic semiparametric efficiency of an omnibus estimator of dependence paramters in copula models. In Cuadras, C. and Lallena, J. R., editors, *Proceedings of the Conference on Distributions with Given Marginals and Statistical Modelling (Barcelona, Espagne)*. Kluwer Academic Publishers, Dordrecht.

[Gihman and Skorohod, 1974] Gihman, I. I. and Skorohod, A. V. (1974). *The theory of stochastic process I.* Die grundleheren der mathematischen wissenschaften ineinzeldarstellungen. Springer, Berlin.

[Han and Kamber, 2006] Han, J. and Kamber, M. (2006). *Data mining: concepts and techniques*. Morgan Kaufmann.

[Hand et al., 2001] Hand, D. J., Mannila, H., and Smyth, P. (2001). *Principles of data mining*. MIT Press.

[Hardy et al., 1934] Hardy, G. H., Littlewood, J. E., and Polya, G. (1934). *Inequalities*. Cambridge University Press, Cambridge.

[Hastie et al., 1995] Hastie, T., Buja, A., and Tibshirani, R. (1995). Penalized discriminant analysis. *Ann. Statist.*, 23:73–102.

[Hastie et al., 2001] Hastie, T., Tibshirani, R., and Friedman, J. (2001). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer-Verlag.

[Herbich, 2002] Herbich, R. (2002). *Learning Kernel Classifiers: Theory and Algorithms*. MIT Press.

[Hjort and Walker, 2001] Hjort, N. L. and Walker, S. G. (2001). A note on kernel density estimators withoptimal bandwidths. *Statistics & Probability Letters*, 54:153–159.

[Hoeffding, 1940] Hoeffding, W. (1940). Masstabinvariante korrelationstheorie. *Schriften des Matematischen Instituts und des Instituts fr Angewandte Mathematik des Universitt*, 5(3):179–233.

[Hoeffding, 1941] Hoeffding, W. (1941). Masstabinvariante korrelationsmasse. *Arkiv fr mathematischen Wirtschaften und Sozialforschung*, 7:49–70.

[Hollander and Wolfe, 1973] Hollander, M. and Wolfe, D. (1973). *Nonparametric Statistical Methods*. John Wiley & Sons.

[Jecklin, 1953] Jecklin, H. (1953). Trigonometrische mittelwerte. *Elemente der Mathematik eine Zeitschrift der Schweizerischen Mathematischen Gesellschaft*, 8:54–60.

[Joe, 1997] Joe, H. (1997). *Multivariate models and dependence concepts*. Chapman and Hall, London.

[Joe and Xu, 1996] Joe, H. and Xu, J. J. (1996). The estimation method of inference functions for margins formultivariate models. Technical report, Department of Statistics, University of British Columbia.

[Kimberling, 1974] Kimberling, C. (1974). A probabilistic interpretation of complete monotonicity. *Aequationes Math.*, 10:152–164.

[Kolmogorov, 1930] Kolmogorov, A. N. (1930). Sur la notion de moyenne. *Rendiconti Accademia dei Lincei*, 12(6):388–391.

[Kolmogorov, 1933] Kolmogorov, A. N. (1933). *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Springer.

[Kolmogorov, 1950] Kolmogorov, A. N. (1950). *Foundations of probability*. Chelsea Publishing Company.

[Kotz et al., 2000] Kotz, S., Balakrishnan, N., and Johnson, N. L. (2000). *Continuous Multivariate Distributions-Volume 1: Models and Applications*. John Wiley & Sons, New York.

[Kruskal, 1958] Kruskal, W. H. (1958). Ordinal measures of association. *J. Amer. Statist. Assoc.*, 53:814–861.

[Lamperti, 1966] Lamperti, J. (1966). *Probability : a survey of the mathematical theory*. Mathematics Monograph Series. W.A. Benjamin, Inc.

[Lamperti, 1977] Lamperti, J. (1977). *Stochastic Processes : A Survey of the Mathematical Theory*. Applied Mathematical Sciences. Springer-Verlag.

[Lehmann, 1975] Lehmann, E. (1975). *Nonparametrics: Statistical Methods Based on Ranks*. Holden-Day Inc.

[Ling, 1965] Ling, C.-H. (1965). Representation of associative functions. *Publ. Math. Debrecen*, 12:189–212.

[Lusin, 1912] Lusin, N. (1912). Sur les propriétés des fonctions mesurables. *Comptes Rendus de l'Académie des Sciences de Paris*, 154:1688–1690.

[Lusternik and Sobolev, 1974] Lusternik, L. A. and Sobolev, V. J. (1974). *Elements of Functional Analysis*. Hindustan Publishing Corpn., Delhi.

[Mazliak, 2007] Mazliak, L. (2007). Les fanômes de l'ecole normale. Technical report, Université Paris VI, France.

[Meyn and L, 1993] Meyn, S. P. and L, T. R. (1993). *Markov chains and stochastic stability*. Communications and Control. Springer-Verlag, New York.

[Nagumo, 1930] Nagumo, M. (1930). Uber eine klasse der mittelwerte. *Japan Journal of Mathematics*, 7:71–79.

[Nelsen, 1999] Nelsen, R. (1999). *An introduction to copulas*. Springer, London.

[Neveu, 1964] Neveu, J. (1964). *Bases mathmatiques du calcul des probabilits*. Masson.

[NIST/SEMATECH, 2008] NIST/SEMATECH (2008). e-handbook of statistical methods. "`http://www.itl.nist.gov/div898/handbook/`".

[Parzen, 1962] Parzen, E. (1962). On the estimation of a probability density function and the mode. *Ann. Math. Statist.*, 33:1065–1075.

[Pratelli, 1940] Pratelli, A. (1940). Sulle medie trigonometriche. *Atti. Prima Riern. Soc. Ital.*, pages 99–104.

[R Development Core Team, 2005] R Development Core Team (2005). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

[Ramsay and Silverman, 2002] Ramsay, J. O. and Silverman, B. W. (2002). *Applied Functional Data Analysis : Methods and Case Studies*. Springer Series in Statistics. Springer, New-York.

[Ramsay and Silverman, 2005] Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis*. Springer Series in Statistics. Springer, New-York.

[Roncalli, 2002] Roncalli, T. (2002). *Gestion des risques multiples*. Groupe de recherche Opérationnelle - Crédit Lyonnais.

[Rossi and Villa., 2006] Rossi, F. and Villa., N. (2006). Support vector machine for functional data classification. *Neurocomputing*, 69:730–742.

[Rudin, 1987] Rudin, W. (1987). *Real and complex analysis*. McGraw-Hill.

[Schiaparelli, 1868] Schiaparelli, G. (1868). Sul principio dell'uso della media aitmetica nel calcolo dei risultati delle osservazioni. *Rendiconti del Regio Istituto Lombardo di Scienze e Lettere*, 1:771–778.

[Schiaparelli, 1875] Schiaparelli, G. (1875). Sur le principe de la moyenne arithmetique. *Astronom. Nachr.*, 87:55–58.

[Scholkopf and Smola, 2002] Scholkopf, B. and Smola, A. (2002). *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press.

[Schumaker, 1981] Schumaker, L. (1981). *Spline Functions: Basic Theory*. John Wiley and Sons, New York.

[Schweizer and Sklar, 1974] Schweizer, B. and Sklar, A. (1974). Operations on distribution functions not derivable from operations on random variables. *Studia Mathematica*, LII:43–52.

[Schweizer and Sklar, 1983] Schweizer, B. and Sklar, A. (1983). *Probabilistic Metric Spaces*. North-Holland, New-York.

[Schweizer and Wolff, 1981] Schweizer, B. and Wolff, E. (1981). On nonparametric measures of dependence for random variables. *Annals of Statistics*, 9:879–885.

[Shih and Louis, 1995] Shih, J. H. and Louis, T. (1995). Inferences on the asociation parameter in copula models for bivariate survival data. *Biometrics*, 51:1384–1399.

[Silverman, 1986] Silverman, B. W. (1986). *Density estimation for statistics and data analysis*. Chapman and Hall, London.

[Sklar, 1959] Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publications Statistiques Université de Paris*, 8:229–231.

[Sklar, 1973] Sklar, A. (1973). Random variables, joint distribution functions and copulas. *Kybernetika*, 9:449–460.

[Stirzaker, 2005] Stirzaker, D. (2005). *Stochastic processes and models*. Oxford University Press, Oxford.

[Tuddenham and Snyder, 1954] Tuddenham, R. D. and Snyder, M. M. (1954). Physical growth of california boys and girls from birth to eighteen years. *University of California Publications in Child Development*, 1:183364.

[Vapnik, 1995] Vapnik, V. (1995). *The Nature of Statistical Learning Theory*. Springer Verlag.

[Vrac, 2002] Vrac, M. (2002). *Analyse et modélisation de données probabilistes par décompostion de mélange de copule et application á une base de données climatologiques*. PhD thesis, Université Paris IX - Dauphine.

[Vrac et al., 2004] Vrac, M., Diday, E., and Chédin, A. (2004). Décomposition de mélange de distributions et application á des données climatiques. *Revue de Statistique Appliquée*, LII:67–96.

[Vrac et al., 2001] Vrac, M., Diday, E., Chédin, A., and Naveau, P. (2001). Mélange de distributions de distributions, décomposition de mélange de copules et application à la climatologie. In *Actes du VIIIème congrés de la Société Francophone de Classification*, pages 348–355.

[Wand and Jones, 1995] Wand, M. P. and Jones, M. C. (1995). *Kernel Smoothing*. Chapman and Hall.

[Wasserman, 2004] Wasserman, L. (2004). *All of Statistics, A Concise Course in Statistical Inference*. Springer.

[Widder, 1941] Widder, D. V. (1941). *The Laplace Transform*. Princeton University Press.

List of Personal Publications

Written and oral communications

Journal papers:

1. Etienne Cuvelier, Monique Noirhomme-Fraiture, *Classification de données fonctionnelles par décomposition de mélange: Apports de la visualisation dans le cas des distributions de probabilité,* ***Revue d'Intelligence Artificielle,*** volume 22, issue 3-4, pp. 421-442, 2008,

2. Etienne Cuvelier, Monique Noirhomme-Fraiture, *Classification de fonctions continues à l'aide d'une distribution et d'une densité définies dans un espace de dimension infinie,* ***Revue des Nouvelles Technologies de l'Information (RNTI),*** volume RNTI-E-9, pp. 679-690, Cépaduès-Éditions, 2007.

Chapters in collective books:

1. Etienne Cuvelier, Monique Noirhomme-Fraiture, *An approach to stochastic process using quasi-arithmetic means,* ***Recent Advances in Stochastic Modeling and Data Analysis,*** pp. 2-9, World Scientific, 2008,

2. Etienne Cuvelier, Monique Noirhomme-Fraiture, *Parametric families of probability distributions for functional data using quasi-arithmetic means with archimedean generators,* ***Functional and Operatorial Statistics, collection Contributions to Statistics,*** pp. 127-133, Physica-Verlag HD, 2008,

3. Monique Noirhomme-Fraiture, Etienne Cuvelier, *Symbolic Markov Chains,* ***Selected Contributions in Data Analysis and Classification, collection Studies in Classification, Data Analysis and Knowledge Organization,*** pp. 103-111, Springer, 2008.

International conferences proceedings:

1. Etienne Cuvelier, Monique Noirhomme-Fraiture, *A Probability Distribution of Functional Random Variable with a Functional Data Analysis Application,* ***The Second International Workshop on Mining Complex Data - MCD'06 - In Conjunction with IEEE ICDM'06, Hong-Kong,*** pp 247-252, IEEE Computer Society, 2006,

2. Etienne Cuvelier, Monique Noirhomme-Fraiture, *Clayton copula and mixture decomposition,* ***Applied Stochastic Models and Data Analysis (ASMDA 2005),*** Brest, 17-20 May, pp 699-708, 2005,

3. Etienne Cuvelier, Monique Noirhomme-Fraiture, *Mélange de distributions de distributions: Décomposition de mélange avec la copule de Clayton,* ***XXXV èmes Journées de Statistiques,*** pp. 377-380, 2003.

International conferences oral communications:

1. Etienne Cuvelier, Monique Noirhomme-Fraiture, *From quasi-arithmetic means to parametric families of probability distributions for functional data,* ***COMPSTAT08 (International Conference on Computational Statistics),*** Porto, August 24th-29th, 2008,

2. Etienne Cuvelier, Monique Noirhomme-Fraiture, *The QAMML families of parametric probability distributions for functional data,* ***IWAP 2008 (International Workshop on Applied Probability),*** Compiègne, July 7-10, 2008,

3. Etienne Cuvelier, Monique Noirhomme-Fraiture, *Classification de distributions par décomposition de mélange de copules archimédiennes: choix de la dimension des copules par visualisation,* ***4e Atelier Visualisation et extraction de connaissances, EGC'2006,*** Lille, 2006,

4. Etienne Cuvelier, Monique Noirhomme-Fraiture, *Classification of probability distributions by mixture decomposition using Archimedean copulas,* ***3rd world conference on Computational Statistics and Data Analysis,*** Limassol, 2005.

National conferences oral communications:

1. Etienne Cuvelier, Monique Noirhomme-Fraiture,, *The Generalized QAMML Families Of Probability Distributions For Functional Data,* ***16th Annual meeting of the Belgian Statistical Society,*** Wépion-Namur, 2008,

2. Etienne Cuvelier, Monique Noirhomme-Fraiture,, *Modeling Stochastic Process Using Quasi-Arithmetic Means, A Functional Data Analysis Approach,* ***15th Annual meeting of the Belgian Statistical Society,*** Antwerpen, 2007,

3. Etienne Cuvelier, Monique Noirhomme-Fraiture, *Probability distribution and density for functional random variable, A new tool for functional data analysis,* ***14th Annual meeting of the Belgian Statistical Society,*** Houffalize, 2006.

Local working groups and seminars:

1. Etienne Cuvelier, Monique Noirhomme-Fraiture, *Distributions QAMML, un outil pour la classification probabiliste de données fonctionnelles,* ***Séminaires scientifiques de la Facultés d'Informatique(Staffsem),*** FUNDP, Namur, 2008,

2. Etienne Cuvelier, Monique Noirhomme-Fraiture, *Classification de données fonctionnelles: le cas des distributions de probabilités dans le cadre de l'analyse symbolique,* ***Séminaires scientifiques de la Facultés d'Informatique(Staffsem),*** FUNDP, Namur, 2006,

3. Etienne Cuvelier, Monique Noirhomme-Fraiture, *A probability distribution and density for functional random variables,* ***1st Research Contact Day Computational Intelligence and Learning,*** Doctoral School, Bruxelles, 2006.

Forthcoming communication:

1. Etienne Cuvelier, Monique Noirhomme-Fraiture, *QAMML distributions: a spell to break the curse of dimensionality for stochastic process,* ***Applied Stochastic Models and Data Analysis (ASMDA 2009),*** Vilnius, Lithuania, 2009.

Symbols and Acronyms

$2^{\aleph_0}$ cardinality of $\mathbb{IR}$, page 48

$\aleph_0$ cardinality of $\mathbb{N}$, page 48

$\mathcal{A}_{\mathcal{X},\mathcal{D}}(u)$ the set $\{\omega \in \Omega : \mathcal{X} \leq_{\mathcal{D}} u\}$, page 44

$\mathcal{A}_{\mathcal{X},\mathcal{D},n}(u)$ the set $\bigcap_{i=1}^{n}\{\omega \in \Omega : \mathcal{X}(t_i,\omega) \leq u(t_i)\}$, page 51

$\mathcal{D}$ real interval, domain of an *frv* , page 42

$\mathcal{F}_{\mathcal{X},\mathcal{D}}$ *fcdf* of the *frv* $\mathcal{X}$ defined on $\mathcal{D}$, page 44

$\mathbb{F}_{\mathcal{X},\mathcal{D}}(u)$ GQAMML *fcdf* of the *frv* $\mathcal{X}$ defined on $\mathcal{D}$, page 205

$\mathcal{F}_{\mathcal{X},\mathcal{D},\phi}$ *Quasi-Arithmetic Mean of Margins Limit fcdf* of the *frv* $\mathcal{X}$ defined on $\mathcal{D}$, using the generator ϕ, page 190

$f_{\mathcal{X},\mathcal{D},\phi}$ Gâteaux density of $\mathcal{F}_{\mathcal{X},\mathcal{D},\phi}$, page 214

$\mathcal{X}$ functional random variable, page 42

$\mathbf{G}_{\mathcal{X},\mathcal{D}}$ functional surface of margins of the *frv* $\mathcal{X}$ defined on $\mathcal{D}$, page 76

$\mathbf{g}_{\mathcal{X},\mathcal{D}}$ functional surface of densitiesof the *frv* $\mathcal{X}$ defined on $\mathcal{D}$, page 76

$\mathcal{I}$ closed real interval, image of an *frv* , page 48

$\mathcal{I}^{\mathcal{D}}$ set of all functions from $\mathcal{D}$ into $\mathcal{I}$, page 48

$\leq_{\mathcal{D}}$ pointwise order between two functions defined on the interval $\mathcal{D}$, page 42

$\mathcal{M}$ aggregation operator, page 173

$\mathbb{P}_{\mathbb{N},\{i_1,\ldots,i_n\}}$ natural projection from $\mathbb{IR}^{\infty}$ into $\mathbb{IR}^n$, page 49

$\mathbf{X}$ vector of real random variables X_i, page 39

$\mathbf{x}$ vector of real numbers x_i, page 39

$\mathcal{BF}$ the smallest $\sigma-$algebra containing $\mathcal{F}$, page 49

$\mathcal{C}_{i_1,\ldots,i_n}(A^{(n)})$ cylindrical set of $A^{(n)} \subset \mathbb{R}^n$, page 49

$\mathcal{F}$ Boolean algebra of Borel cylindrical set, page 49

$\mathcal{P}_{\mathcal{I}^{\mathcal{D}},\mathbb{R}^\infty}$ projection operator from $\mathcal{I}^{\mathcal{D}}$ into $\mathbb{R}^\infty$, page 48

$\mathcal{R}^n$ class of n-dimensional Borel sets, page 37

Ω sample space, page 36

ω elementary event, page 36

ϕ generator of an archimedean copula, page 107

Π product copula, page 87

$\pi_{n,k}$ natural projection from $\mathbb{R}^n$ into $\mathbb{R}^k$, page 40

ψ inverse of the generator ϕ of an archimedean copula, page 107

$\mathbb{R}^\infty$ set of sequences of real numbers, page 46

$\mathbb{R}^n$ n-dimensional Euclidean space, page 37

τ Kendall's tau, page 97

C copula, page 85

$d_1(u,v)$ Euclidean distance between u and v, page 207

$DF(u;s)$ Gâteaux differential of F in u in direction of s, page 209

F_X distribution function of the real random variable X, page 38

f_X density function of the real random variable X, page 38

$G_{\mathcal{X},\mathcal{D}}$ surface of margins of the *frv* $\mathcal{X}$ defined on $\mathcal{D}$, page 55

$g_{\mathcal{X},\mathcal{D}}$ surface of densities of the *frv* $\mathcal{X}$ defined on $\mathcal{D}$, page 55

$H_{\mathbf{X}}$ distribution function of the random vector $\mathbf{X}$, page 39

$h_{\mathbf{X}}$ density function of the random vector $\mathbf{X}$, page 39

$H_{QAMM,\mathbf{X}}$ QAMM joint distribution of the random vector $\mathbf{X}$, page 186

$L^1(\mathcal{D})$ set of square integrable functions belonging to $\mathcal{I}^{\mathcal{D}}$, page 207

M Fréchet-Hoeffding upper bound copula, page 87

N_0 strictly positive natural numbers, page 39

P probability function, page 37

Index