Universitat de Barcelona
Departament d'Estadística

Construction of Bivariate Distributions and Statistical Dependence Operations

M. Desamparados Casanova Gurrera

29 d'abril de 2005

Programa de Doctorat
Probabilitats i Estadística
Bienni 1997-99

Memòria presentada per a optar al grau de
Doctora en Medicina i Cirurgia

Certifico que la present memòria
ha estat realitzada per
M. Desamparados Casanova Gurrera,
en el Departament d'Estadística
de la Universitat de Barcelona
sota la meva direcció

Prof. Dr. Carles M. Cuadras Avellana
Catedràtic del Departament d'Estadística
Universitat de Barcelona

Barcelona, 25 de gener de 2005.

A la meua Família

Agraïments

Al llarg d'aquests anys he repetit moltes vegades la paraula Gràcies. Gràcies, en primer lloc, al director d'aquesta tesi, el Dr. Carles Cuadras, per la confiança que va dipositar en mi en acceptar aquesta tasca. En el cursos de doctorat em va introduïr en el mon de les Còpules, i també de l'Anàlisi Multivariant. I ho va fer confiant en que podria assolir el nivell necessari, encara que provenia d'una àrea de coneixement aliena. No s'impacientà si jo no aconseguia provar un teorema després de moltes hores de treball; i sabia conduïr-me perquè arribés per mi mateixa al resultat. Així, em dirigia a l'hora que m'ensenyava a avançar. Li agraeixo també les moltíssimes hores que ha dedicat a la revisió d'aquesta tesi.

Gràcies als meus pares i a la tieta Joaquina que em van ensenyar a no tenir por a les dificultats, i a buscar sempre la Veritat i el Perquè de les coses. Gràcies a la meva família que m'ha recolzat sempre. Han renunciat a molt perquè jo pogués acabar aquesta etapa de la meva formació professional, especialment els meus germans Josep, M. Teresa, Fernando, Ignasi, Pilar i Sindo. Gràcies, també, a les meves amigues, la Jose, l'Anna, l'Ana i les seves famílies, la Núria i la Glòria. Elles han mostrat la mateixa il.lusió que jo en aquesta tesi, i no s'han cansat de sentir-me dir que no puc compartir més temps amb elles. Gràcies a la Carmen i a la Rosa per tanta paciència. Gràcies als meus nebots i fillols que m'han fet passar tantes bones estones, fruïnt de les coses senzilles: la Mercè, el Fernando, l'Albert, la Beth, la Cristina, José Ramón, l'Albert, l'Eduard, la Marta, l'Ignasi, Joan i la Isaura.

Gràcies també als meus companys del Departament de Salut Pública a la Facultat de Medicina: de tots he aprés moltíssimes coses, no només professionals. Gràcies als meus professors del programa de doctorat del Departament d'Estadística; van posar tot de la seva part per tal que assolís els coneixements que m'han permés arribar fins aquí. Al Martín li agraeixo les seves permanents paraules d'ànim. Gràcies als meus companys de doctorat, especialment al Josep Anton, de qui vaig aprendre a programar, i tantes coses més; i a la Gloria, per la seva amistat.

Agraeixo també al Dr. Juan Quesada, al Dr. David Márquez i al Dr. Sindo Núñez llurs comentaris a aquesta memòria, i llur recolzament. Agraeixo també a la Montserrat les hores que ha passat al meu costat per a aconseguir millorar la presentació de la tesi, i a l'Alfonso les correccions al llenguatge i la seva paciència.

Finalment, dóno gràcies a Déu per haver-me'ls posat al costat.

Contents

Introduction

Interest in understanding the laws that govern universal phenomena can possibly be considered as intrinsically linked to the rational nature of humans. With the mere observation of the phenomenon as a starting point, human reason tries to know how, when, and why it occurs. Knowledge of the Laws that govern Nature allows Man to dominate it in some way. Man can fly, for instance, because he knows the laws of gravity, mechanics and hydrodynamics, among others. The complexity of these laws reflects the complexity of the physical world. The speed at which an object falls from some height is determined by the gravitational pull of the earth on this object. Its energy, produced by the movement of the force, can be computed by multiplying the mass of the object, the height from which it falls, and the gravitational acceleration (which is 9.8 $m\ s^{-2}$, as physicists have shown). In a similar and easy way, we can compute the speed of the impact when the object reaches the earth's surface. This is because we know that the energy of the object does not diminish as it falls, but rather is changed into another kind of energy: kinetic energy. If, additionally, we consider air friction, the problem grows in complexity but also becomes more realistic.

We can add complexity to the physical model that allows us to understand a phenomenon when our knowledge of this phenomenon improves. Laws governing the free fall of an object are mainly deterministic. We are absolutely sure that an apple will fall down if we toss it into the air. Moreover, the same thing will surely happen every time we toss an apple into the air. We can study this phenomenon numerically because our experience allows us to reject some complex factors as irrelevant in practice.

Other phenomena do not show such a deterministic behaviour. The apples on the same branch of a tree will each have a different weight. Some of the apples grow bigger than others, although all of them are exposed to nearly the same amount of sunlight. The variability in the weight of the apples alerts us to the fact that we are studying a phenomenon with diverse nature. This is not deterministic, but random. We cannot be sure about the weight of an apple before choosing it. But we are sure that it will fall every time that we drop it. We say that the weight is a *random variable*, in a statistical sense. Although we cannot be sure about the weight of an apple, if we weigh all the apples harvested by a farmer, and we know that the mean weight is 200 grams, it is reasonable to expect the weight of a given apple to be close to 200 grams. If we know that the farmer has harvested two kinds of apples, A and B, with odds of $4:1$, when choosing one hundred apples we expect to find 80 apples of the A kind, and 20 apples of the B kind, approximately. The frequency of observation of the values of weight, and the kind of apples, allow us to predict which event shows the highest "probability". Nevertheless, as in the deterministic instance, random phenomena are affected by many factors.

In this thesis, we study the dependence between random variables from a theoretical perspective, at various levels. As Jogdeo ([57]) notes, "Dependence relations between random variables is one of the most widely studied subjects in probability and statistics. The nature of the dependence can take a variety of forms and unless some specific assumptions are made about the dependence, no meaningful statistical model can be contemplated". First, it is convenient to study the *probability law* of only one variable, abstracting from other factors that may unnecessarily complicate the model. Once the law of a random variable has been characterized, we are ready to compute the probability of this variable taking a value in a particular range. The function that assigns to every real value, say, x, the probability of a variable to take values equal to or lower than x is called the *cumulative distribution function* of the variable. In mathematics, it is denoted as $F(x) = P(X \leq x)$. The distribution function characterizes the probabilistic behaviour of a random variable completely.

This is an example: let a variable X be defined by assigning 0 to the event "heads" when a coin is tossed and 1, to the event "tails". The distribution function $F(x)$ of X equals 0 if $x < 0$, equals $\frac{1}{2}$ if $0 \leq x < 1$, and 1 if $x \geq 1$. Let another variable be defined by assigning 0 to the event "even" when a die is thrown, and 1 to the event "odd". The distribution function is identical to that of the previous example. If we define another variable as the number of dots on the face of the die, its distribution function equals 0, if $x < 1$, it equals $\frac{k}{6}$ if $k \leq x < k+1$, for $k = 1, ..., 5$, and it equals 1 if $x \geq 6$.

Laws governing random phenomena possibly show more complexity than those of a deterministic nature, but a number of general models have been found that fit variables taken from experimental fields. The law (the probability function) of variables valued on either a countable or a finite set has been described, like those in the examples above that follow a discrete uniform law. A number of so-called continuous random variables that take values in some continuous (sub)set of the real numbers, $\mathbb{R}$, have also been introduced. In this case, the distribution function is continuous too. The most well-known is the Normal distribution, and it was the most useful one until about 1930. Later on, research in a number of fields expanded, and many data appeared whose probabilistic behaviour departed significantly from normality. Since then many other univariate distributions have been characterized (uniform, Cauchy, Gamma, Laplace, Pareto, Weibull, etc.). Variables with a probability function fitting one of these models (Normal, for instance) share the "form" of the function, and they can be distinguished from each other by other measures such as the mean value, or the dispersion from the mean, etc.

Once the study of univariate distributions began to cover a fairly wide range of experimental data, research focused shifted to relations between (two or more) random variables. Multivariate models constitute an area of increasing interest, in both Probability and Statistics. Statistical methods for multivariate data analysis and inference aim at identifying the underlying stochastic structure when given a specific sample of data, i.e., the goal is to specify the univariate marginals as well as the underlying dependence structure between each pair of variables. This is the main purpose of this thesis. We have applied the *diagonal expansions method* to construct bivariate distributions with given marginals and fixed *correlation* (see below).

The simplest multivariate distribution is that of two random variables, X, Y, coupled by a *random vector* (X, Y), with a *bivariate (joint) distribution function.* This function *models* (captures) the dependence between these two variables. Dependence can be total, in which case knowledge of one random variable determines the other. For instance, when $Y = X+3$, almost surely. The other extreme is that either random variable does not give any information about the other. We say that X and Y are *stochastically independent* in this case. Between these extrema, there are many possible dependence relations that have been addressed in literature. The most studied measure of dependence is the *covariance* of (X, Y), which is defined as the expectation of the product of the differences between each variable and its own expectation. We writte $Cov\,(X, Y) = E\,[(X - E\,(X))\,(Y - E\,(Y))]$. This measure of dependence is not unit-free. We can standardize it by dividing by the square root of the variances of each variable involved (i.e., the covariance of each variable with itself). This dimension-free coefficient is Pearson's *correlation coefficient*; it is bounded by -1 and 1, and reaches the bounds if there is a linear relation between the variables. Hence, it is considered as a measure of linear relation. In this thesis we have used the covariance to study the relations between two random variables. Previously, we have studied the relations between two distance matrices obtained from the same set of n objects; we have defined some operations between matrices and we have extended all the results to the continuous case, using kernels (i.e., real functions of two variables). We have also studied expansions of symmetric functions of two variables, in general. These operations involve the covariance between functions of random variables, which are joined by a bivariate distribution. The diagonal expansions method uses complete orthogonal sets of functions. In this particular construction, the correlation between two variables is determined by the covariance between two specific functions of the variables, the so-called *principal components*.

Some other common dependence relations are: dependence between variables with only two possible values, whose frequencies are shown in a 2×2 table (contingency table), stochastic independence (mentioned above), *positive* (or *negative*) *dependence* (i.e. large values of a variable tend to go together with large (or small) values of the other variable), etc. A large variety of dependence concepts have been studied by a number of authors, offering proper definitions and useful properties with applications.

As in the univariate case, the Bivariate Normal Distribution has been the main bivariate model for years. It is easy to manage, and most of the statistical techniques of data analysis assume this distribution (uni- or bivariate) for the data. However, such an assumption is not easy to verify and, moreover, to hold in a number of experimental situations, when data show evident non-normality. Nowadays, many other multivariate models have been described. They are models fitting properly experimental data, as shown by several authors. Construction of bivariate distributions with fixed marginals was first done by Hoeffding, 1940 ([52]) and Fréchet, 1951 ([42]); the latter finds the upper and lower bound of a class (a set) of all the bivariate distributions that share the marginals. This class is called the Fréchet class. The Fréchet class is too broad for practical interest. Too many dependence models exist in the same class. It is often more fruithful to build families of joint distributions with fixed marginals, following the same dependence model, as we have studied in this thesis. The dependence between the random variables involved is enclosed by one (or more) parameter different from the parameters of the univariate marginals. These parameters should be easy to understand. Kimeldorf and Sampson, 1975 ([65]) give five conditions that any parametric family of distributions must satisfy (see Chapter 1). There are many families, with applications in a wide range of fields, satisfying such conditions: the Bivariate Normal, Farlie-Gumbel-Morgenstern (FGM) family, Ali-Mikhail-Haq, Frank, Cuadras-Augé, Plackett, bivariate Pareto, Regression, Clayton-Oakes, etc. The range of dependence covered by any such family is a meaningful research subject. For instance, the FGM parameter θ takes values in $[-1, 1]$ and Pearson's correlation is $\frac{\theta}{3}$, hence, the maximum correlation is $\frac{1}{3}$. If a bivariate sample shows a correlation coefficient of, for instance, 0.7, the probability of the sample drawn from a population with bivariate distribution FGM is extremely low. On the other hand, we expect to observe stochastic independence when the parameter values 0, and maximum dependence when the parameter attains its maximum value.

Hence, the study of the joint distribution of two variables belongs to the broadest field of the study of multivariate relations. One of the main objectives of Statistics is to establish conclusions about population phenomena from data samples. Data allow us to estimate population parameters, and to test hypotheses regarding these parameters with specified *statistical confidence*. From this global perspective our purpose has been to study the dependence between two random variables by means of the covariance.

Suppose we study several characteristics of individuals in a population. Each characteristic can be interpreted as a random variable. One sample of each random variable allows us to construct a data matrix with p columns, one for every variable, with all the data of one individual in a row. This matrix has a finite number of rows, equal to the number of individuals in the sample. From a theoretical perspective, continuous random variables require infinite matrices (bounded linear operators) to cover the whole range of their values. If we obtain two (finite) matrices from the same set of n individuals with different statistical distances, as well as defining two real-valued kernels over the same spaces, we can study the relationships between the two matrices (and the two kernels) by means of a suitable operation that involves the covariance between two random "objects". Subsequently the covariance between random variables and between any functions of random variables can be further studied. Clearly, the dependence between these pairs of random objects (matrices, kernels, functions of random variables) is perfectly defined by their joint distribution function. If we are able to construct a bivariate family of distributions with one (or more) multivariate parameters, we must study the theoretical aspects of the multivariate parameters: whether they are indeed dependence parameters, the range of dependence covered, possible values, etc.

In the following section, we describe the most relevant subjects that motivate this thesis.

Overview of the thesis and main results

We begin by studying dependence between random variables at various levels, and the last two chapters are devoted to the construction of bivariate distributions via principal components. Although we refer to classical bibliography on these topics for an overview, every chapter includes further explanations about less standard issues such as Related Metric Scaling, Principal Components of Random Variables, and finally, Diagonal Expansions of Bivariate Distributions. Readers familiar with any of these topics may skip these sections. Nevertheless, some specific concepts appear recurrently throughout this thesis, so Chapter 1 of Preliminaries is devoted to them. Nomenclature and notation are introduced for some specific dependence concepts. We briefly discuss Fréchet classes, copulas, and parametric families of distributions.

In Chapter 2, we generalize the operation introduced in [28] between two distances (the union and intersection operations of two distance matrices) to symmetric non-negative definite matrices. These operations are shown to be useful in the geometric interpretation of Related Metric Scaling (*RMS*), and possibly in other approaches of Multivariate Analysis. When two distance matrices associated with the same finite set of n objects are available, the underlying dependence between these matrices can be taken into account building an *intersection* distance matrix; on the other hand, the complete and possibly redundant information may be included in a joint distance matrix, considered as their *union*. These operations show relevant properties that are studied in this chapter. The behaviour of the operations is, in some way, analogous to that presented by the intersection and union between vector spaces (for instance, those spanned by the columns of the matrices); in particular, we prove that the intersection of orthogonal matrices is the null matrix, while the union is the direct sum of the matrices. Matrices that share their eigenvectors form an equivalence class, and a partial order relation is defined. This class is closed for the union and intersection operations.

A continuous extension of these operations is presented in Chapter 3. Infinite matrices are studied in the context of bounded integral operators and numerical kernels. Metric Scaling has already been studied from a continuous perspective by Cuadras and Fortiana ([24], [25], [26], [27]). We put the basis for extending *RMS* to continuous random variables and, hence, infinite matrices. The starting point is Mercer's Theorem, which ensures the existence of an orthogonal expansion of the covariance kernel $K(s,t) = \min\{F(s), F(t)\} - F(s)F(t)$, where F is the cumulative distribution function of each marginal variable. The sets of eigenvalues and eigenfunctions of K, whose existence is ensured by the cited theorem, allow us to define a product between symmetric and positive (semi)definite kernels, and, further, to define the *intersection* and the *union* between them. Results obtained in the discrete instance are extended in this chapter to continuous variables, with examples.

Such covariance kernels (symmetric and positive definite) are associated with symmetric and positive quadrant dependent (PQD) bivariate distributions. Covariance between functions of bounded variation defined on the range of some random variables, joined by distributions of this type, can be computed by means of their cumulative distribution functions (see [18]). In Chapter 4, further consequences are obtained, especially some relevant relations between the covariance and the Fréchet bounds, with a number of results that can be useful in the characterization of independence as well as in testing goodness-of-fit. The *intersection* of two kernels (defined in Chapter 3) is a particular instance of the covariance between functions. Covariance is a quasi-inner product defined through the joint distribution of the variables involved. A measure of affinity between functions with respect to H is defined, and also studied.

In Chapter 5, from the concept of affinity between functions via an extension of the covariance, we define the dimension of a distribution, we relate it to the diagonal expansion and find the dimension for some parametric families. Thus we find a finite, a countable and a continuous dimension for the generalized Farlie-Gumbel-Morgenstern, Ali-Mikhail-Haq and Cuadras-Augé families, respectively.

Diagonal expansions of bivariate distributions (Lancaster, [71]) allows us to construct bivariate distributions. It has proved to be adequate for constructing Markov processes (see [5], [94]), and has also been applied to engineering problems (see [103]), among other uses. This method has been generalized using the principal dimensions of each marginal variable that are, by construction, canonical variables. We introduce in Chapter 6 the theoretical foundations of this method. In Chapter 7 we study the bivariate, symmetric families obtained when the marginals are Uniform on $(0, 1)$, Exponential with mean 1, standard Logistic, and Pareto ($\alpha = 3$, $\theta = 1$). Conditions for the bivariate density, first canonical correlation and maximum correlation of each family of densities are given in some cases. The corresponding copulas are also obtained.

We conclude this thesis with a brief description of the derived ongoing and future research.

Chapter 1

Preliminaries

In the Introduction we intuitively discussed the concept of statistical dependence. The concept of marginal and joint distribution function is now properly introduced. Let $(\Omega, \mathcal{A}, P)$ be a probability space, i.e., Ω is a set, $\mathcal{A}$ is a $\sigma-$algebra, family of parts of Ω, and P is an application $P : \mathcal{A} \to [0,1]$ such that (i) $P(\Omega) = 1$, and (ii) if $\{A_n, \ n \geq 1\}$ is a sequence of sets of $\mathcal{A}$, with A_n, A_m disjoints, then $P(\cup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n)$. Let $\mathcal{B}(\mathbb{R})$ be the $\sigma-$algebra generated by the borelian sets in $\mathbb{R}$. A *random variable* (r.v.) X is an application $X : \Omega \to \mathbb{R}$ such that $\forall B \in \mathcal{B}(\mathbb{R}),\ X^{-1}(B) \in \mathcal{A}$. That is, a random variable is a measurable function on a probability space.

Definition 1.0.1 *Let X be a random variable. The cumulative distribution function of the random variable X is the function $F : \mathbb{R} \to [0,1]$ defined by $F(x) = (P \circ X^{-1})((-\infty, x]) = P(X \leq x)$, for $x \in \mathbb{R}$.*

The cumulative distribution function (or distribution function) is non-decreasing, right-continuous, and $\lim_{x \to -\infty} F(x) = 0$, $\lim_{x \to \infty} F(x) = 1$. This function assigns to $x \in \mathbb{R}$ the probability of $X \in (-\infty, x]$.

The joint distribution function is an extension. Let $(\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2, P_1 \times P_2)$ be the product space of two probability spaces.

Definition 1.0.2 *Let $\mathbf{X} = (X, Y)$ be a random vector. The joint distribution function of $\mathbf{X}$ is the function $H : \mathbb{R}^2 \to [0,1]$ defined by*

$$H(x, y) = (P \circ \mathbf{X}^{-1})((-\infty, x] \times (-\infty, y]) = P(X \leq x, Y \leq y)$$

for $(x, y) \in \mathbb{R}^2$.

The random variables X, Y are called the marginal variables of the random vector. The distribution functions of X and Y, say, F and G, are called the marginal distributions of (X, Y). The joint distribution function is right-continuous, and must also satisfy the following properties:

1. $\lim_{x\to\infty} H(x,y) = G(y)$, $\lim_{y\to\infty} H(x,y) = F(x)$, where F, G are the marginal distribution functions of X, Y, respectively,

2. $\lim_{(x,y)\to(\infty,\infty)} H(x,y) = 1$, where $(x,y) \to (\infty,\infty)$ stands for both variables x, y tending to infinite,

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

4. for all $(x_1, x_2), (y_1, y_1)$ with $x_1 < x_2$, $y_1 < y_2$,

$$H(x_2, y_2) - H(x_1, y_2) - H(x_2, y_1) + H(x_1, y_1) \geq 0.$$

As mentioned above, the joint distribution function encloses the dependence relation between the variables. Hence, we introduce some definitions.

1.1 Some dependence concepts

Some measures of dependence are specially relevant, with properties that make them useful to explain the meaning of a statistical model. The concept of total dependence is opposite to stochastic independence and appears in this context.

Let $(\Omega, \mathcal{A}, P)$ be a probability space, and let $(E_i, \mathcal{E}_i)$, $i \in I$ be a measurable space. One collection of events $\{A_i,\ i \in I\}$ is said to be independent if for any finite subset $J \subset I$, $P(\cap_{j\in J} A_j) = \Pi_{j\in J} P(A_j)$. We say that the family of measurable applications $\{X_i,\ i \in I\}$,

$$X_i : (\Omega, \mathcal{A}) \to (E_i, \mathcal{E}_i)$$

is independent if the collection of $\sigma-$algebra $\left\{X_i^{-1}(\mathcal{E}_i), i \in I\right\}$ is also independent. In particular, if $\{X_i,\ i \in I\}$ is a family of r.v.'s (i.e., $\mathcal{E}_i$ is the $\sigma-$algebra of the borelian sets in $\mathbb{R}$), the family (or the variables) are said to be independent if the collection of $\sigma-$algebra $\left\{X_i^{-1}(\mathcal{B}(\mathbb{R})),\ i \in I\right\}$ is also independent (see, for instance [40], [90]).

The following properties are useful in the characterization of stochastic independence between r.v.'s.

Properties

1. $X_1, X_2, ..., X_p$ are stochastically independent r.v.'s if and only if (iff) the law of the random vector $(X_1, X_2, ..., X_p)$ is equal to the product of the marginal laws. This is equivalent to

$$H(x_1, \dots, x_p) = F_1(x_1) \cdot \dots \cdot F_p(x_p),$$

where H and $F_1, \dots, F_p$ are the joint and marginal cumulative distribution functions (cdf's), respectively.

2. If $X_1, X_2, ..., X_p$ are independent and integrable, then the product $X_1 \cdot \dots \cdot X_p$ is integrable, and

$$E\left(X_1 \cdot \dots \cdot X_p\right) = E\left(X_1\right) \cdot \dots \cdot E\left(X_p\right)$$

where E is the expectation. Note that this condition is necessary but not sufficient for stochastic independence.

3. If $\{X_i,\ i \in I\}$ are independent r.v.'s and $g_i : \mathbb{R} \to \mathbb{R}$ are measurable functions, then the r.v.'s $\{g_i\left(X_i\right),\ i \in I\}$ are also independent.

Property 1 provides a very useful characterization of independence using the law of the random vector involved. Non-independence or dependence between random variables means some kind of relation between them. Two r.v.'s X, Y are said to be *implicitly dependent* if there exist two functions f, g such that $f(X) = g(Y)$, with $Var(f(X)) > 0$, where Var denotes the variance of a random variable (see below). The random variables X, Y are said to be *functionally dependent* if there exist some functions f, g such that either $Y = f(X)$, or $X = g(Y)$. The random variables X, Y are said to be *mutually completely dependent* if there exists a one-to-one function f such that, $P(Y = f(X)) = 1$. If f is linear we say that X, Y are *linearly dependent.*

In the next two sections we review some of the most useful measures of dependence. We will focus on the best known measure, its derivation and consequences.

1.1.1 Dependence Measures

Covariance and Pearson's Correlation coefficient

Let $X, Y \in L^2(\Omega, \mathcal{A}, P)$ be two square integrable random variables. The covariance between X and Y is defined by

$$Cov(X, Y) = E[(X - E(X))(Y - E(Y))] = E(XY) - E(X)E(Y).$$

From Property 2 it can be easily proved that if X, Y are square integrable and independent, their covariance is zero. We say that both variables are uncorrelated. It has been noted that zero correlation does not imply stochastic independence. Observe that the variance $Var(X)$ of a square integrable r.v. X is $Cov(X, X)$.

The Pearson correlation coefficient is defined as

$$\rho = Cor(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}}.$$

This coefficient is one of the most useful measures of dependence. If X, Y are centred r.v.'s on $L^2(\Omega, \mathcal{A}, P)$, their covariance coincides with the ordinary inner product of X and Y defined in the Hilbert space $L^2(\Omega, \mathcal{A}, P)$ and hence, Pearson's correlation coefficient ρ can be interpreted as the cosine of the angle between X and Y, which in turn can be viewed as vectors with norms $\sqrt{Var(X)}$ and $\sqrt{Var(Y)}$, respectively.

Canonical correlations

Suppose that $\mathbf{X} = (X_1, \ldots, X_p)$, $\mathbf{Y} = (Y_1, \ldots, Y_q)$ are two second-order random vectors. The first set of canonical correlations and variates is defined as the linear combinations $X^{(1)} = a_1 X_1 + \ldots + a_p X_p, \quad Y^{(1)} = b_1 Y_1 + \ldots + b_q Y_q$ such that $Cor(X^{(1)}, Y^{(1)})$ is maximum. The second set of canonical correlations and variates is defined as similar linear combinations $X^{(2)}, Y^{(2)}$ such that $Cor(X^{(2)}, Y^{(2)})$ is maximum given that $Cor(X^{(1)}, X^{(2)}) = Cor(Y^{(1)}, Y^{(2)}) = 0$. This procedure is continued until $\min\{p, q\}$ sets are obtained.

It is of interest to extend this type of analysis to more general classes of multivariate distributions. Lancaster [70] applied the methods of the theory of integral equations to find the canonical correlations and variables in the joint normal distribution and generalized the canonical correlation theory. The canonical variables $\{a_i\}, \{b_j\}$ are two sets of orthonormal functions defined on the marginal distributions in a recursive manner such that the correlation between corresponding members of the two sets is maximal, given the preceding canonical variables. The $\rho_i = Cor\left(a_i\left(X\right), b_i\left(Y\right)\right)$ are the canonical correlations and can be assumed positive. If a bivariate distribution is ϕ^2-bounded[1], it can be expanded in an eigenfunction expansion. The canonical variables on each marginal variable are the eigenfunctions (except for a factor) and form a subset of a complete orthonormal set. The canonical correlations are the eigenvalues of the expansion.

Maximal correlation

Gebelein, 1941 ([43]) introduced this relevant measure of dependence between two r.v.'s X and Y, defined by

$$\rho' = \sup Cor\left(\alpha\left(X\right), \beta\left(Y\right)\right),$$

where the supremum is taken over all Borel-measurable functions α and β, with $Var\left(\alpha\left(X\right)\right)$, $Var\left(\beta\left(Y\right)\right) > 0$. Rényi, 1959 ([93]) proposed a set of seven axioms that a measure of dependence must satisfy (see conditions 1 to 7, below), and showed that the maximal correlation satisfies all of them. One problem is that it often equals 1.

If the bivariate distribution is ϕ^2-bounded, then the maximal correlation equals ρ_1, the first canonical correlation.

If the supremum is taken over all monotone functions α and β, such that $Var\left(\alpha\left(X\right)\right)$, $Var\left(\beta\left(Y\right)\right) > 0$ we have the *monotone correlation*, a measure of the degree of monotone relation between two variables X, Y. The concept of monotone dependence was suggested by Kimeldorf and Sampson ([67]).

---

[1]Pearson's contingency coefficient ϕ^2 is defined by

$$\phi^2 + 1 = \int_a^b \int_c^d (dH(x,y))^2/(dF(x)dG(y)),$$

as a measure of association of the r.v.'s X and Y, with $[a,b], [c,d]$ the ranges of X, Y, respectively. We say that H is ϕ^2-bounded if $\phi^2 < \infty$.

Conditions on a Measure of Dependence

Hutchinson & Lai [54], p. 180, lists nine conditions proposed by a number of authors ([75], [93], [98]), with further comments. We reproduce this list and the comments. The quantity $\delta(X, Y)$ denotes an index of dependence between X, Y.

1. $\delta(X, Y)$ is defined for any pair of random variables, neither of them being constant with probability 1. This condition avoids trivialities.
2. $\delta(X, Y) = \delta(Y, X)$. Notice, however, that while independence is a symmetric property, total dependence is not, as one variable may be determined by the other, but not vice versa.
3. $0 \leq \delta(X, Y) \leq 1$. Lancaster ([75]) says that this is an obvious choice, but not everyone agrees.
4. $\delta(X, Y) = 0$ iff X and Y are mutually independent. Notice how strong this condition is made by the "only if".
5. If the functions f and g map the spaces of X and Y in a one-to-one manner, respectively, onto themselves, then $\delta(f(X), g(Y)) = \delta(X, Y)$. This condition means that the index remains invariant under one-to-one transformations of the marginal variables.
6. $\delta(X, Y) = 1$ iff X and Y are mutually completely dependent.
7. If X and Y are jointly normal, with correlation coefficient ρ, then $\delta(X, Y) = |\rho|$.
8. In any family of distributions defined by a vector parameter θ, $\delta(X, Y)$ must be a function of θ.
9. If (X, Y) and (X_n, Y_n), $n = 1, 2, ...$ are pairs of random variables with joint distributions H and H_n, respectively, and if $\{H_n\}$ converges in distribution to H, then $\lim_{n\to\infty} \delta(X_n, Y_n) = \delta(X, Y)$.

1.1.2 Concordance measures

Consonni and Scarsini, 1982 ([9]) define the concordance between two variables and a partial order relation between cdf's. Let H and H' be two continuous joint cdf's, with the same marginals, say F, G. If a point $(x_0, y_0) \in \mathbb{R}^2$ is fixed, we can determine four subsets of $\mathbb{R}^2$:

$$\begin{aligned} Q_1(x_0, y_0) &= \left\{(x, y) \in \mathbb{R}^2 : x \leq x_0, y \leq y_0\right\}, \\ Q_2(x_0, y_0) &= \left\{(x, y) \in \mathbb{R}^2 : x \leq x_0, y > y_0\right\}, \\ Q_3(x_0, y_0) &= \left\{(x, y) \in \mathbb{R}^2 : x > x_0, y > y_0\right\}, \\ Q_4(x_0, y_0) &= \left\{(x, y) \in \mathbb{R}^2 : x > x_0, y \leq y_0\right\}. \end{aligned}$$

We say that H is more concordant than H' if, for all $(x, y) \in \mathbb{R}^2$,

$$\begin{aligned} &P\{(X, Y) \in [Q_1(x, y) \cup Q_3(x, y)] \mid (X, Y) \sim H\} \\ \geq\ &P\{(X, Y) \in [Q_1(x, y) \cup Q_3(x, y)] \mid (X, Y) \sim H'\}, \end{aligned} \tag{1.1}$$

where $(X, Y) \sim H$ means that H is the cdf of (X, Y). Concordance gives the intuition of large values of one variable join with large values of the other variable.

Spearman's rho, and Kendall's tau

Let X, Y be two r.v.'s with cdf's F, G, respectively. Spearman's ρ, 1904 ([101]), denoted by ρ_S, is defined as

$$\rho_S = Cor\,(F(X), G(Y)). \tag{1.2}$$

This measure is invariant with respect to strictly increasing transformations of the variables. Notice that if the variables are uniformly distributed on the interval $(0, 1)$, $\rho_S = \rho$.

Let (X, Y), (X', Y') be independent random pairs of variables, identically distributed with common joint distribution H. Then Kendall's τ, 1938 ([64]) is defined as

$$\tau = P\{(X - X')(Y - Y') > 0\} - P\{(X - X')(Y - Y') < 0\}, \tag{1.3}$$

i.e., Kendall's τ is the difference of the probability of two random concordant pairs and the probability of two random discordant pairs.

This measure is also invariant under strictly increasing transformations (see [6]).

Conditions on a measure of concordance

Scarsini, 1984 ([95]) lists the axioms that a measure of concordance must satisfy. They are analogous to those conditions given for measures of dependence. Consider the space $\mathcal{H}$ of joint cdf's with continuous marginals. The partial order relation defined by (1.1) allows us to establish a total order relation on $\mathcal{H}$. Let I be a totally ordered set (usually, $I \subset \mathbb{R}$). Let (X, Y) be distributed according to the cdf H and define the map $J : \mathcal{H} \longrightarrow I$. We define the concordance between X and Y $(\theta(X, Y))$ as $J(H)$. This map J is called a measure of concordance if it satisfies the following axioms:

1. Domain: $\theta(X, Y)$ is defined for any (X, Y) with continuous cdf.
2. Symmetry: $\theta(X, Y) = \theta(Y, X)$.
3. Coherence: $\theta(X, Y)$ is monotone in the corresponding copula C_{XY} (see Section 1.2 below), i.e., if $C_{X_1Y_1} \geq C_{X_2Y_2}$, then $\theta(X_1, Y_1) \geq \theta(X_2, Y_2)$.
4. Range: $-1 \leq \theta(X, Y) \leq 1$.
5. Independence: $\theta(X, Y) = 0$ if X and Y are stochastically independent.
6. Change of sign: $\theta(-X, Y) = -\theta(X, Y)$.
7. Continuity: If $(X, Y) \sim H$ and $(X_n, Y_n) \sim H_n$, $n \in \mathbb{N}$, and if H_n converges pointwise to H (H and H_n continuous), then $\lim_{n\to\infty} \theta(X_n, Y_n) = \theta(X, Y)$.

1.2 Bivariate distributions

1.2.1 Fréchet classes

Two random variables X, Y with marginal cdf's F, G may show total complete dependence, and then we can predict one from the other; or they may show stochastic independence, or any other dependence relation. Dependence between the marginal variables lies on the joint cdf, say H; the joint cdf determines the marginals, but the opposite is false.

The set of joint cdf's with the same marginals forms an equivalence class. This class is the Fréchet class of distributions (see [42]) with marginals F, G. This is denoted $\mathcal{F}(F,G)$. The independence distribution $H^0(x,y) = F(x)G(y)$ belongs to $\mathcal{F}(F,G)$.

An order relation is defined on $\mathcal{F}(F,G)$. We say that H_1 is lower than H_2 ($H_1 \leq H_2$) iff for every pair $(x,y) \in \mathbb{R}^2$, $H_1(x,y) \leq H_2(x,y)$. It can be shown that every $H \in \mathcal{F}(F,G)$ satisfies

$$H^- \leq H \leq H^+,$$

where

$$\begin{aligned} H^-(x,y) &= \max\{F(x) + G(y) - 1, 0\}, \\ H^+(x,y) &= \min\{F(x), G(y)\}. \end{aligned}$$

These relevant cdf's are called the lower and upper Fréchet bounds, respectively, for obvious reasons (see [20] for a number of results involving the upper bound). The case of discrete random variables is studied in [86]. The best bounds for arbitrary sets of distributions are studied in [88].

The correlation coefficient of (X,Y) can be $|\rho| = 1$ if $F = G$, but in general, if F and G are different (e.g., normal and exponential) then $|\rho| < 1$. The minimum and maximum correlations between X and Y, say, ρ^-, ρ^+, are called the Hoeffding correlations, and they satisfy:

1. $\rho^- \leq \rho \leq \rho^+$,
2. $\rho = \rho^-$ iff $H = H^-$, and $\rho = \rho^+$ iff $H = H^+$.

The proof can be found in [42], [52], or [77].

1.2.2 Copulas

A 2−copula is a function $C : [0,1] \times [0,1] \rightarrow [0,1]$ which satisfies:

1. For every u, v in $[0,1]$,

   $$C(u,0) = C(0,v) = 0,$$

   and

   $$C(u,1) = u, \quad C(1,v) = v.$$

2. C is 2-increasing, i.e., for every u_1, u_2, v_1, v_2 in $[0,1]$ such that $u_1 \leq u_2$ and $v_1 \leq v_2$,
$$C(u_2, v_2) - C(u_1, v_2) - C(u_2, v_1) + C(u_1, v_1) \geq 0.$$

Sklar's theorem (see [100]) says that every joint cumulative distribution function H has an associated copula C such that:

$$H(x,y) = C(F(x), G(y)), \quad x, y \in \mathbb{R},$$

where F is the cdf of X, and G is the cdf of Y. If X, Y are absolutely continuous r.v.'s, C is unique. The r.v.'s $F(X)$, $G(Y)$ (i.e., the composition of the cdf and the corresponding random variable) follows a Uniform distribution on $(0,1)$. Hence, the change of variables $F(x) = u$, $G(y) = v$ gives the corresponding copula. In some way, copulas are "independent" of the univariate marginals, and represent the dependence structure, separately. Simplicity of copulas allows us to perform analysis which may present high difficulties using H. Let $\mathcal{F}(U,U)$ denotes the Fréchet class of the copulas, i.e., U denotes the cdf of a uniform r.v. on $[0,1]$. Every $C \in \mathcal{F}(U,U)$ satisfies

$$C^- \leq C \leq C^+,$$

where, for $u, v \in [0,1]$,

$$\begin{aligned} C^-(u,v) &= \max\{u+v-1, 0\}, \\ C^+(u,v) &= \min\{u,v\}, \end{aligned}$$

are the Fréchet bounds.

Schweizer and Wolf (see [98]) connect Spearman's ρ and Kendall's τ with copulas (see also [85]). See Nelsen, 1999 ([89]) for an excellent introduction to this topic. See also [32], [41], [44], [66], and [91].

1.2.3 Parametric families of distributions

Statistical modelling needs to characterize the dependence between X and Y (the bivariate distribution H), and the marginal distributions (the Fréchet class) as well. But the Fréchet class $\mathcal{F}(F,G)$ of bivariate distributions with the same marginals is very wide, as already mentioned. It is much more interesting to model bivariate distributions of $\mathcal{F}(F,G)$, with the same dependence structure given by a vector of parameters. For instance, if F is the univariate Normal distribution with parameters

$(\mu_1, \sigma_1) \in \mathbb{R} \times [0, \infty)$, G is the Normal distribution with parameters $(\mu_2, \sigma_2) \in \mathbb{R} \times [0, \infty)$, and $\rho \in [-1, 1]$ is the correlation coefficient of (X, Y), then $H_\rho \in \mathcal{F}(F, G)$ is the class of all bivariate distributions with these marginals.

Into this class we can find mixtures of Bivariate Normal Distributions (BND), densities defined piecewise, and a wide number of bivariate distributions whose shape is absolutely different from the Bivariate Normal (see, for instance, [69] for examples). The bivariate density function

$$h(x, y) = \omega_1 \phi_1(x, y) + \omega_2 \phi_2(x, y),$$

where ω_1, ω_2 are probability weights, and ϕ_i, $i = 1, 2$, is the density of a standard BND with correlation coefficient ρ_i, is non-normal, if $\rho_1 \neq \rho_2$, with Normal marginals. The correlation between X and Y is $\rho = \omega_1 \rho_1 + \omega_2 \rho_2$. Taking $\omega_1 = \omega_2 = \rho_1 = -\rho_2 = 0.5$, we have $\rho = 0$. The density is

$$h(x, y) = \frac{1}{2\pi\sqrt{3}} \left(\exp\left(-\frac{2}{3}\left(x^2 - xy + y^2\right) \right) + \exp\left(-\frac{2}{3}\left(x^2 + xy + y^2\right) \right) \right),$$

$(x, y) \in \mathbb{R}^2$. Notice that, in this example, X and Y are not stochastically independent. The parameter ρ encloses the dependence (the linear correlation) between X and Y. If H_ρ is the Bivariate Normal Distribution (BND), we have $\rho = 0$ iff X and Y are independent r.v.'s (i.e., $H_0(x, y) = F(x) G(y)$). If $|\rho| = 1$ we have the singular bivariate Normal.

It is evident that this bivariate distribution with Normal marginals shows a behaviour absolutely different from the BND, even though both joint distributions belong to the same Fréchet class. The density of the Bivariate Normal distribution with $\rho \neq 1$ is

$$h(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1 - \rho^2}} \exp\left(-\frac{1}{2} Q(x, y) \right),$$

for $(x, y) \in \mathbb{R}^2$, where

$$Q(x, y) = \frac{1}{1 - \rho^2} \left(\frac{(x - \mu_1)^2}{\sigma_1^2} - 2\rho \frac{(x - \mu_1)(y - \mu_2)}{\sigma_1 \sigma_2} + \frac{(y - \mu_2)^2}{\sigma_2^2} \right).$$

We say that all densities with this representation belong to the family of Bivariate Normal distributions. For instance, let X, Y be two r.v.'s following a standard Normal distribution $N(0, 1)$. If (X, Y) follow a BND with $\rho = 0$ their joint density is given by

$$h(x, y) = \frac{1}{2\pi} \exp\left(-\frac{1}{2}\left(x^2 + y^2\right) \right), \quad (x, y) \in \mathbb{R}^2,$$

i.e., X, Y are independent r.v.'s. Figure 1.1 shows this density.

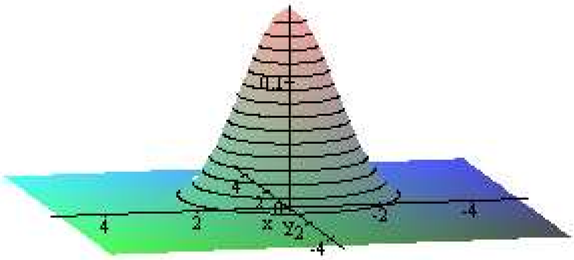

Figure 1.1: Bivariate Normal distribution with independent marginals

Parametric families of distributions have been constructed by a number of authors (Fréchet [42], Kimeldorf and Sampson [65], and Mardia [78], among others). Some of these families include the Fréchet bounds. Kimeldorf and Sampson, 1975 ([65]) proposes five conditions on one-parameter families of distributions. For any family of distributions $\{H_\theta, \ -1 \leq \theta \leq 1\}$, with absolutely continuous marginals, and parameter θ: (i) if $\theta = 1$, H_1 is the Fréchet upper bound (see Section 1.2); (ii) if $\theta = 0$, H_0 is the independence distribution; (iii) if $\theta = -1$, H_{-1} is the Fréchet lower bound; (iv) for fixed (x, y), $H_\theta(x, y)$ is continuous in $\theta \in [-1, 1]$; and (v) for fixed $\theta \in [-1, 1]$, H_θ is an absolutely continuous distribution function. We introduce some 1-parametric families which are shown to be useful in applications, and which appear through this work. See also [68].

Farlie-Gumbel-Morgenstern family (FGM)

Let (X, Y) be a random vector with absolutely continuous cdf $H_\theta \in \mathcal{F}(F, G)$. We write $(X, Y) \sim H_\theta$. The Farlie-Gumbel-Morgenstern of distributions (see [10]) is defined as

$$H_\theta(x, y) = F(x)\, G(y) \left[1 + \theta (1 - F(x)) (1 - G(y))\right],$$

for every pair (x, y) in the product of the ranges of X and Y, and with $\theta \in [-1, 1]$. The density is

$$h_\theta(x, y) = f(x)\, g(y) \left[1 + \theta (1 - 2F(x)) (1 - 2G(y))\right],$$

where $f = F'$, $g = G'$. The corresponding copula is

$$C_\theta(u,v) = uv\left[1 + \theta(1-u)(1-v)\right],$$

for $0 \leq u, v \leq 1$, with density

$$c_\theta(u,v) = 1 + \theta(1-2u)(1-2v).$$

The Fréchet bounds do not belong to this family (i.e., $H^- < H_{-1} < H_1 < H^+$). If $\theta = 0$ we have independence ($H_0 = FG$).

It can be proved that Pearson's correlation coefficient is given by $Cor(X,Y) = \frac{\theta}{3}$. If marginals are uniform, the maximum correlation is reached ($\frac{1}{3}$); for Normal marginals $Cor(X,Y) = \frac{1}{\pi}$; for exponential marginals $Cor(X,Y) = \frac{1}{4}$ (see [49], [97] for the proofs). For logistic marginals $Cor(X,Y) = \frac{3}{\pi^2}$ (see Chapter 7).

This family has been studied by Eyraud, 1936 ([38]), Morgenstern, 1956 ([84]), Farlie, 1960 ([39]) and Gumbel, 1960 ([48], [49], [50]). A number of applications have been published (for a extensive list, see [54]). FGM distributions have been used in statistical modelling, in studying test of association, and in other statistical applications such as generation of random variables ([33], [59], [60]).

Regression family (RG)

This family was proposed by Cuadras, 1996 ([13]). If the ranges of X, Y are the intervals $[a,b], [c,d]$, and $\varphi : [a,b] \to [c,d]$ is an increasing function, the family

$$H_\theta(x,y) = \theta F(\min\{x, \varphi^{-1}(y)\}) + (1-\theta)F(x)\,J_\theta(y), \quad 0 \leq \theta < \theta^+,$$

is a bivariate cdf with marginals F, G,

$$J_\theta(y) = (G(y) - \theta F(\varphi^{-1}(y)))/(1-\theta)$$

is a cdf. Here θ^+ is the maximum value for θ such that J_θ is a cdf. The regression curve is linear in φ, and $H_\theta(x,y)$ has a singular part with mass on the curve $y = \varphi(x)$. It can be proved that $P[Y = \varphi(X)] = \theta$. See [12], [13].

The simplest version of this family appears with $\varphi = G^{-1} \circ F$:

$$H_\theta(x,y) = \theta \min\{F(x), G(y)\} + (1-\theta)F(x)G(y), \quad 0 \leq \theta \leq 1.$$

The density with respect to the measure $\mu^2 + \mu_1$ is

$$h_\theta(x,y) = \theta f(x)\,\delta_{\{y=x\}} + f(x)\,(g(y) - \theta f(y))\,\delta_{\{y\neq x\}},$$

where μ^2 is the Lebesgue measure on $\mathbb{R}^2$, μ_1 is the Lebesgue measure concentrated in $y = x$, and δ is the indicator function. The corresponding copula is

$$C_\theta(u,v) = \theta \min\{u,v\} + (1-\theta)uv, \quad 0 \leq \theta \leq 1,$$

with density

$$c_\theta(u,v) = \theta\delta_{\{u=v\}} + (1-\theta)\,\delta_{\{u\neq v\}}.$$

The Fréchet upper bound is attained at $\theta = 1$, if $F = G$.

For this family, and assuming $F = G$ the correlation is $Cor(X,Y) = \theta$.

With this family we can generate bivariate data given either a linear or non-linear regression function.

Ali-Mikhail-Haq family (AMH)

Ali et al., 1978 ([1]) describes the bivariate cdf

$$H_\theta(x,y) = \frac{F(x)\,G(y)}{1-\theta\,(1-F(x))\,(1-G(y))},$$

with density that may be written as

$$\begin{aligned} h_\theta(x,y) &= f(x)\,g(y) \times [1 + \theta\,(1-2F(x))\,(1-2G(y)) \\ &\quad + \sum_{k=2}^{\infty} k\theta^k\,(1-F(x))^{k-1}\,(1-G(y))^{k-1}]. \end{aligned} \tag{1.4}$$

The corresponding copula is

$$C_\theta(u,v) = \frac{uv}{1-\theta\,(1-u)\,(1-v)},$$

with density

$$c_\theta(u,v) = \frac{1-\theta+2\theta\,[uv/\,(1-\theta\,(1-u)\,(1-v))]}{(1-\theta\,(1-u)\,(1-v))^2}.$$

The Fréchet bounds do not belong to this family. There is independence for $\theta = 0$.

The first term of this family includes the FGM (if the series in (1.4) is null).

Cuadras-Augé family (CA)

This distribution was introduced by Cuadras and Augé, 1981 ([22]), and is a weighted geometric mean of the independence distribution and the Fréchet upper bound:

$$H_\theta(x,y) = \min\{F(x), G(y)\}^\theta (F(x) G(y))^{1-\theta},$$

with density with respect to the measure $\mu^2 + \mu_1$

$$h_\theta(x,y) = (1-\theta) f(x) g(y) \min\{F(x), G(y)\}^{-\theta} + \theta f(x) (F(x))^{1-\theta} \delta_{\{(F(x)=G(y))\}},$$

where μ^2 is the Lebesgue measure on $\mathbb{R}^2$, μ_1 is the Lebesgue measure concentrated on the curve $F(x) = G(y)$, and δ is the indicator function. The copula is

$$C_\theta(u,v) = \min\{u,v\}^\theta (uv)^{1-\theta},$$

with density

$$c_\theta(u,v) = (1-\theta)\min\{u,v\}^{-\theta} + \theta u^{1-\theta}\delta_{\{u=v\}}.$$

The Fréchet upper bound is attained at $\theta = 1$. There is independence if $\theta = 0$.

Pearson's correlation for uniform marginals is $Cor(U,V) = \frac{3\theta}{4-\theta}$ and the maximum correlation is θ.

This distribution is the distribution of Marshall and Olkin if marginals are exponential (see [82], [83]), and provides a survival copula for the Marshall-Olkin family of distributions. Nelsen studies a generalization (see [87]).

Further results about these parametric families as well as many other families can be found in, for instance, [54].

Chapter 2

Dependence concepts and operations on matrices

The connection between matrix theory, geometry and statistics is natural in the context of several statistical methods such as multidimensional scaling (see [46], [92], [96], [105]), correspondence analysis, principal component analysis, as well as in many other distance-based approaches in multivariate analysis (see [14], [81]). For example, consider the well known relation between orthogonality and linear independence, where the geometric approach often makes the statistical interpretation clearer.

The main purpose of this chapter is to establish a theoretical framework which allows us to study the underlying dependence structure of a class of matrices. Intersection and union of vector spaces (spanned by the column vectors of two matrices, for example) are well known operations. Analogously, we define and study the *intersection* and *union* operations on symmetric, non-negative definite matrices. The choice of this terminology will be justified by the properties of these operations.

2.1 Symmetric non-negative definite matrices: basic definitions and properties

In this section we review some basic properties of symmetric non-negative definite matrices (see, for instance, [47], [51], [102]). Let $\mathcal{M}_n$ denote the set

$$\mathcal{M}_n = \{\mathbf{A} \mid \mathbf{A} \text{ is an } n \times n \text{ symmetric, positive (semi)definite matrix}\}.$$

That is, $\mathbf{A} = \mathbf{A}'$ and all its eigenvalues are strictly positive (positive definite, p.d.) or non-negative with at least one null eigenvalue (positive semidefinite, p.s.d.)[1].

Let $\mathbf{X}$ and $\mathbf{Y}$ be $n \times n$ matrices with $rank\,(X) = rank\,(Y) = p$ and $p \leq n$, and define

$$\mathbf{A} = \mathbf{X}\mathbf{X}', \qquad \mathbf{B} = \mathbf{Y}\mathbf{Y}'. \tag{2.1}$$

Then $\mathbf{A}, \mathbf{B} \in \mathcal{M}_n$. By the symmetry of $\mathbf{A}$ and $\mathbf{B}$, there exist (spectral decomposition theorem) orthogonal matrices $\mathbf{U}, \mathbf{V}$ and real diagonal matrices $\mathbf{\Lambda} = diag\,(\lambda_1, ..., \lambda_n)\,,$ $\mathbf{M} = diag\,(\mu_1, ..., \mu_n)$, such that

$$\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\,\mathbf{U}' = \sum_{i=1}^{n} \lambda_i \mathbf{u}_i \mathbf{u}_i', \qquad \mathbf{B} = \mathbf{V}\mathbf{M}\mathbf{V}' = \sum_{i=1}^{n} \mu_i \mathbf{v}_i \mathbf{v}_i',$$

where $\mathbf{u}_i$ and $\mathbf{v}_i$, $i = 1, ..., n$ are the orthonormal columns of $\mathbf{U}$ and $\mathbf{V}$, respectively. The symmetry and the positive semidefiniteness ensures the existence of the rational powers of a matrix, say $\mathbf{A}^{\frac{r}{s}}$, $r < s$, for any positive integers r, s. In particular, its square root can be expressed as

$$\mathbf{A}^{\frac{1}{2}} = \mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{U}' = \sum_{i=1}^{n} \lambda_i^{\frac{1}{2}} \mathbf{u}_i \mathbf{u}_i',$$

which are also symmetric and p.(s.)d. matrices. Note that $\mathbf{X}, \mathbf{Y}$ in (2.1) can be found from $\mathbf{A}, \mathbf{B}$ by taking $\mathbf{X} = \mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}$, $\mathbf{Y} = \mathbf{V}\mathbf{M}^{\frac{1}{2}}$. Also note that, if

$$\lambda_{(1)} \geq \lambda_{(2)} \geq ... \geq \lambda_{(p)} > \lambda_{(p+1)} = ... = \lambda_{(n)} = 0$$

(so $\lambda_{(j)}^{\frac{1}{2}} = 0$, $j = p+1, ..., n$), the spectral decomposition $\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\,\mathbf{U}'$ reduces to

$$\mathbf{A} = \mathbf{U}_p \mathbf{\Lambda}_p \mathbf{U}_p' = \sum_{i=1}^{p} \lambda_{(i)} \mathbf{u}_{(i)} \mathbf{u}_{(i)}',$$

where $\mathbf{\Lambda}_p = diag\left(\lambda_{(1)}, ..., \lambda_{(p)}\right)$ and $\mathbf{U}_p = \left[\mathbf{u}_{(1)}; ...; \mathbf{u}_{(p)}\right]$.

2.2 Orthogonality of matrices

The term orthogonality is extremely useful in algebra and geometry. Two vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ are said to be orthogonal if their inner product is zero:

$$\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u}'\mathbf{v} = \sum_{i=1}^{n} u_i v_i = 0.$$

[1]In this work, non-negative definite or p.(s.)d. matrices are equivalent to p.s.d. or p.d. matrices, indistinctly, when the cited properties hold for both types of matrices.

Two vector subspaces E, F of $\mathbb{R}^n$ are said to be orthogonal if every vector in E is orthogonal to every vector in F. In this case, the only vector belonging to $E \cap F$ (the intersection subspace) is the null vector. It is obvious that, if $\mathbf{u}$ is orthogonal with respect to $\mathbf{v}$, then $\mathbf{v}$ is also orthogonal with respect to $\mathbf{u}$ (i.e., the inner product of two vectors of $\mathbb{R}^n$ is commutative). The (usual) inner product of two matrices $\mathbf{A}, \mathbf{B}$ is defined as

$$\langle \mathbf{A}, \mathbf{B} \rangle = tr\left(\mathbf{AB}'\right),$$

provided the dimensions allow the operation. The trace of a squared $n \times n$ matrix $\mathbf{C}$ is the sum of the elements in its diagonal, i.e., $tr\left(\mathbf{C}\right) = \sum_{i=1}^{n} c_{ii}$. This is, obviously, commutative. Harville [51] defines two orthogonal matrices as two matrices whose inner product ($tr\left(\mathbf{AB}'\right)$) equals 0.

A natural definition of two matrices being orthogonal should express orthogonality between every row (or column) vector in the matrix on the left and every column (or row) vector in the matrix on the right. The column (row) space of a matrix is the subspace spanned by its columns (rows). Since the ordinary product of two matrices is not commutative we may have $\mathbf{AB} = \mathbf{0}$ while $\mathbf{BA} \neq \mathbf{0}$, being $\mathbf{0}$ the null matrix. Following the terminology used by Graybill [47], $\mathbf{A}$ and $\mathbf{B}$ are "disjoint", while $\mathbf{B}$ and $\mathbf{A}$ might not be "disjoint". Thus, the definition of orthogonality using the ordinary product is questionable. Nevertheless, if $\mathbf{A}$ and $\mathbf{B}$ are symmetric, $\mathbf{AB} = \mathbf{0}$ iff, $\mathbf{BA} = \mathbf{0}$ and also $tr\left(\mathbf{AB}\right) = tr\left(\mathbf{AB}'\right) = tr\left(\mathbf{BA}'\right) = 0$. Therefore we introduce the following definition of orthogonality between symmetric matrices.

Definition 2.2.1 *Two symmetric $n \times n$ matrices $\mathbf{A}$ and $\mathbf{B}$, are said to be orthogonal if their ordinary product equals the null matrix; i.e., $\mathbf{AB} = \mathbf{BA} = \mathbf{0}$.*

The following proposition characterizes p.(s.)d. orthogonal matrices:

Proposition 2.2.2 *Let $\mathbf{X}$ and $\mathbf{Y}$ be $n \times n$ matrices. Consider $\mathbf{A} = \mathbf{XX}'$ and $\mathbf{B} = \mathbf{YY}'$. The following statements are equivalent:*
(i) $\mathbf{A}$ and $\mathbf{B}$ are orthogonal,
(ii) $\mathbf{X}'\mathbf{Y} = \mathbf{0}$,
(iii) the column spaces of $\mathbf{X}$ and $\mathbf{Y}$ are orthogonal.

Proof. Equivalence between (*iii*) and (*ii*), as well as the implication of (*i*) by (*ii*), is evident. It suffices to prove that (*i*) implies (*iii*), although we will prove their equivalence. Let V_a, V_x, V_b, V_y be the column spaces of $\mathbf{A}, \mathbf{X}, \mathbf{B}, \mathbf{Y}$, respectively. The columns of $\mathbf{A}$ and $\mathbf{B}$ are linear combinations of the columns of $\mathbf{X}$ and $\mathbf{Y}$, respectively. Then

$$V_a \subseteq V_x, \qquad V_b \subseteq V_y. \tag{2.2}$$

It is known (see, for instance, [47]) that the number of linearly independent columns of these matrices (the rank) satisfies

$$\begin{aligned} rank\,(\mathbf{A}) &= rank\,(\mathbf{XX}') = rank\,(\mathbf{X})\,, \\ rank\,(\mathbf{B}) &= rank\,(\mathbf{YY}') = rank\,(\mathbf{Y})\,. \end{aligned}$$

Thus inclusions in (2.2) are equalities, and equivalence between (i) and (iii) is proved.

□

Remark 2.2.3 *If either* $\mathbf{A}$ *or* $\mathbf{B}$ *have at least one null eigenvalue,* $\mathbf{AB} = \mathbf{0}$ *does not necessarily imply* $\mathbf{U}'\mathbf{V} = \mathbf{0}$*, where* $\mathbf{U}, \mathbf{V}$ *are the orthogonal matrices with the eigenvectors of* $\mathbf{A}, \mathbf{B}$*, respectively.*

Corollary 2.2.4 *The square roots of two p.(s.)d. orthogonal matrices are also orthogonal.*

PROOF. Let $\mathbf{X} = \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}$ and $\mathbf{Y} = \mathbf{V}\mathbf{M}^{\frac{1}{2}}$ such that $\mathbf{A} = \mathbf{XX}'$ and $\mathbf{B} = \mathbf{YY}'$. Proposition 2.2.2 gives the equivalence between $\mathbf{AB} = \mathbf{0}$ and $\mathbf{X}'\mathbf{Y} = \mathbf{0}$. So, premultiplying $\mathbf{X}'\mathbf{Y}$ by $\mathbf{U}$ and postmultiplying by $\mathbf{V}'$ the desired result is obtained:

$$\begin{aligned} \mathbf{X}'\mathbf{Y} = \mathbf{0} &\Longleftrightarrow \boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{U}'\mathbf{V}\mathbf{M}^{\frac{1}{2}} = \mathbf{0} \\ &\Longleftrightarrow \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{U}'\mathbf{V}\mathbf{M}^{\frac{1}{2}}\mathbf{V}' = \mathbf{0} \\ &\Longleftrightarrow \mathbf{A}^{\frac{1}{2}}\mathbf{B}^{\frac{1}{2}} = \mathbf{0} \end{aligned}$$

□

2.3 Intersection and union of matrices

2.3.1 Definitions

Let $\mathbf{A}, \mathbf{B} \in \mathcal{M}_n$. Consider their square roots $\mathbf{A}^{\frac{1}{2}} = \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{U}'$, $\mathbf{B}^{\frac{1}{2}} = \mathbf{V}\mathbf{M}^{\frac{1}{2}}\mathbf{V}'$.

Definition 2.3.1 *The intersection* $\mathbf{A} \wedge \mathbf{B}$ *and the union* $\mathbf{A} \vee \mathbf{B}$ *of* $\mathbf{A}, \mathbf{B} \in \mathcal{M}_n$ *are defined by*

$$\begin{aligned} \mathbf{A} \wedge \mathbf{B} &= \frac{1}{2}\left(\mathbf{A}^{\frac{1}{2}}\mathbf{B}^{\frac{1}{2}} + \mathbf{B}^{\frac{1}{2}}\mathbf{A}^{\frac{1}{2}}\right), \\ \mathbf{A} \vee \mathbf{B} &= \mathbf{A} + \mathbf{B} - \mathbf{A} \wedge \mathbf{B}. \end{aligned}$$

Proposition 2.3.2 *Let* $\mathbf{A}, \mathbf{B} \in \mathcal{M}_n$*. Then* $\mathbf{A} \wedge \mathbf{B}$ *and* $\mathbf{A} \vee \mathbf{B}$ *are symmetric. Moreover* $\mathbf{A} \vee \mathbf{B} \in \mathcal{M}_n$*.*

PROOF. The symmetry of $\mathbf{A}$ implies the symmetry of $\mathbf{A}^{\frac{1}{2}}$. Thus:

$$\begin{aligned} (\mathbf{A} \wedge \mathbf{B})' &= \frac{1}{2}\left(\mathbf{A}^{\frac{1}{2}}\mathbf{B}^{\frac{1}{2}} + \mathbf{B}^{\frac{1}{2}}\mathbf{A}^{\frac{1}{2}}\right)' \\ &= \frac{1}{2}\left[\left(\mathbf{B}^{\frac{1}{2}}\right)'\left(\mathbf{A}^{\frac{1}{2}}\right)' + \left(\mathbf{A}^{\frac{1}{2}}\right)'\left(\mathbf{B}^{\frac{1}{2}}\right)'\right] \\ &= \mathbf{A} \wedge \mathbf{B}. \end{aligned}$$

The symmetry of the union is straightforward.
Let us write the union as:

$$\mathbf{A} \vee \mathbf{B} = \frac{1}{2}\mathbf{A} + \frac{1}{2}\mathbf{B} + \frac{1}{2}\left(\mathbf{A}^{\frac{1}{2}} - \mathbf{B}^{\frac{1}{2}}\right)^2, \tag{2.3}$$

where all terms are symmetric and p.(s.)d.. Since the linear combination with non-negative coefficients of p.(s.)d. matrices is a p.(s.)d. matrix (see, for instance, [47]), the positive (semi)definiteness of the union is proved. □

Remark 2.3.3 *The intersection is not a p.s.d. matrix, in general; it is the sum of two matrices with all its eigenvalues being non-negative (recall that all the eigenvalues of the product of two p.s.d. matrices are non-negative, [53]), but the linear combination with non-negative coefficients of such matrices, when they are not symmetric, is not necessarily a matrix with all its eigenvalues being non-negative. The matrices*

$$\mathbf{A} = \begin{pmatrix} 9 & 3 \\ 3 & 1 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} 4 & 2 \\ 2 & 1 \end{pmatrix},$$

provide a counterexample, as

$$\mathbf{A} \wedge \mathbf{B} = \begin{pmatrix} 5.9397 & 2.4749 \\ 2.4749 & 0.9899 \end{pmatrix}, \qquad \lambda_1 = 6.9648, \quad \lambda_2 = -0.0352.$$

2.3.2 Basic properties

In this section we obtain some basic properties of the intersection and union operations between matrices, which in some aspects are analogous to the intersection and union of sets.

Proposition 2.3.4 *Let* $\mathbf{A}, \mathbf{B} \in \mathcal{M}_n$*. The intersection and union of* $\mathbf{A}$ *and* $\mathbf{B}$ *have the following properties:*
(a) *Commutativity:*

$$\mathbf{A} \wedge \mathbf{B} = \mathbf{B} \wedge \mathbf{A}, \qquad \mathbf{A} \vee \mathbf{B} = \mathbf{B} \vee \mathbf{A}.$$

(b) *Orthogonality:* $\mathbf{AB} = \mathbf{0}$ *iff*

$$\mathbf{A} \wedge \mathbf{B} = \mathbf{0}, \qquad \mathbf{A} \vee \mathbf{B} = \mathbf{A} + \mathbf{B}.$$

(c) *Equality: if* $\mathbf{A} = \mathbf{B}$ *then*

$$\mathbf{A} \wedge \mathbf{A} = \mathbf{A}, \qquad \mathbf{A} \vee \mathbf{A} = \mathbf{A}.$$

(d) *Null element: the zero matrix is the null element for the intersection and the neutral element for the union:*

$$\mathbf{A} \wedge \mathbf{0} = \mathbf{0}, \qquad \mathbf{A} \vee \mathbf{0} = \mathbf{A}.$$

Moreover, $\mathbf{A} \vee \mathbf{B} = \mathbf{0}$ *iff* $\mathbf{A} = \mathbf{B} = \mathbf{0}$.

PROOF. Properties of the union follow from the properties of the intersection.
(a) follows from the definition of the operations and the commutativity of the sum of matrices.
The direct implication in (b) is evident. The proof of the reverse is based on the fact that $\mathbf{AB} = \mathbf{0}$ implies $\mathbf{A}^{\frac{1}{2}}\mathbf{B}^{\frac{1}{2}} = \mathbf{B}^{\frac{1}{2}}\mathbf{A}^{\frac{1}{2}} = \mathbf{0}$, as established in Corollary 2.2.4. Conversely, recall that the set of eigenvalues of a matrix $\mathbf{C}$ satisfies

$$\{\lambda_i(\mathbf{C})\} = \{\lambda_i(\mathbf{C}')\}. \tag{2.4}$$

From the symmetry of $\mathbf{A}$ and $\mathbf{B}$, note that $\left(\mathbf{A}^{\frac{1}{2}}\mathbf{B}^{\frac{1}{2}}\right)' = \mathbf{B}^{\frac{1}{2}}\mathbf{A}^{\frac{1}{2}}$. The hypothesis $\mathbf{A} \wedge \mathbf{B} = \mathbf{0}$ is equivalent to

$$\begin{aligned} \mathbf{0} &= \frac{1}{2}\left(\mathbf{A}^{\frac{1}{2}}\mathbf{B}^{\frac{1}{2}} + \mathbf{B}^{\frac{1}{2}}\mathbf{A}^{\frac{1}{2}}\right) \\ &\Rightarrow \mathbf{A}^{\frac{1}{2}}\mathbf{B}^{\frac{1}{2}} = -\mathbf{B}^{\frac{1}{2}}\mathbf{A}^{\frac{1}{2}}. \end{aligned}$$

This equality combined with property (2.4) implies

$$\left\{\lambda_i\left(\mathbf{A}^{\frac{1}{2}}\mathbf{B}^{\frac{1}{2}}\right)\right\} = \left\{\lambda_i\left(-\mathbf{A}^{\frac{1}{2}}\mathbf{B}^{\frac{1}{2}}\right)\right\}.$$

Thus all the eigenvalues are 0, and $\mathbf{A}^{\frac{1}{2}}\mathbf{B}^{\frac{1}{2}}$ is the zero matrix.
(c) is clear from the definitions.
To prove (d), observe that any orthogonal matrix $\mathbf{U}$ satisfies $\mathbf{U}\mathbf{0}\mathbf{U}' = \mathbf{0}$, and its square root is also $\mathbf{0}$. Then the proof is straightforward. Elementary algebra provides the uniqueness of the zero element (with the obvious restrictions on its dimension).
The proof of the last statement is based on the above mentioned result: the linear combination with non-negative coefficients of p.(s.)d. matrices is a p.(s.)d. matrix, and this is equal to the null matrix iff all the matrices being summed are the null matrix (see also [47]). Consider the decomposition (2.3) of the union and apply the cited property to the matrices

$$\mathbf{A} \vee \mathbf{B} = \frac{1}{2}\mathbf{A} + \frac{1}{2}\mathbf{B} + \frac{1}{2}\left(\mathbf{A}^{\frac{1}{2}} - \mathbf{B}^{\frac{1}{2}}\right)^2 = \mathbf{0} \Rightarrow \mathbf{A} = \mathbf{B} = \mathbf{0}.$$

□

Remark 2.3.5 *Unfortunately a neutral element for the intersection does not exist. Thus no matrix* $\mathbf{E}$ *exists such that, for all matrices* $\mathbf{A} \in \mathcal{M}_n$,

$$\mathbf{A} \wedge \mathbf{E} = \mathbf{A}. \tag{2.5}$$

If $\mathbf{E}$ *satisfies (2.5) for any* $\mathbf{A} \in \mathcal{M}_n$, *then for the identity matrix* $\mathbf{I}$,

$$\begin{aligned} \mathbf{I} = \mathbf{I} \wedge \mathbf{E} &= \frac{1}{2}\left(\mathbf{I}^{\frac{1}{2}}\mathbf{E}^{\frac{1}{2}} + \mathbf{E}^{\frac{1}{2}}\mathbf{I}^{\frac{1}{2}}\right) \\ &= \frac{1}{2}\left(\mathbf{E}^{\frac{1}{2}} + \mathbf{E}^{\frac{1}{2}}\right) \\ &= \mathbf{E}^{\frac{1}{2}}. \end{aligned}$$

Thus, $\mathbf{E} = \mathbf{I}^2 = \mathbf{I}$. *Then for any* $\mathbf{A} \in \mathcal{M}_n$

$$\begin{aligned} \mathbf{A} = \mathbf{A} \wedge \mathbf{I} &= \frac{1}{2}\left(\mathbf{A}^{\frac{1}{2}}\mathbf{I}^{\frac{1}{2}} + \mathbf{I}^{\frac{1}{2}}\mathbf{A}^{\frac{1}{2}}\right) \\ &= \mathbf{A}^{\frac{1}{2}}. \end{aligned}$$

This is not possible. For example, $\mathbf{A} = \begin{pmatrix} 4 & 0 \\ 0 & 1 \end{pmatrix} \neq \mathbf{A}^{\frac{1}{2}} = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}$.

Remark 2.3.6 *In general, the intersection and the union are not associative.*

2.3.3 Rank

Let $\mathbf{C}$ be an $n \times p$ matrix. The rank r of $\mathbf{C}$ is the number of linearly independent row or column vectors. Thus $r \leq \min\{p, n\}$. Here we study the rank of the matrices obtained by the operations intersection and union. We assume in this section that $\mathbf{X}$ and $\mathbf{Y}$ are of the same rank. It is not necessary, and similar results can be obtained if this condition does not hold.

Consider the matrices $\mathbf{A} = \mathbf{XX}' = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}'$ and $\mathbf{B} = \mathbf{YY}' = \mathbf{VMV}'$. Obviously, $\mathbf{A}, \mathbf{B}, \boldsymbol{\Lambda}, \mathbf{M} \in \mathcal{M}_n$ and $\mathbf{X}$, $\mathbf{Y}$, $\mathbf{U}$, $\mathbf{V}$ are also $n \times n$. Recall that such matrices satisfy, for $p \leq n$,

$$rank\,(\mathbf{A}) = rank\,(\boldsymbol{\Lambda}) = rank\left(\mathbf{A}^{\frac{1}{2}}\right) = rank\,(\mathbf{X}) = p.$$

The following proposition gives bounds for the ranks of the intersection and union of matrices $\mathbf{A}$ and $\mathbf{B}$. Of course, these bounds are meaningful only if either $2p < n$ or $4p < n$; otherwise, the result of the proposition is trivial.

Proposition 2.3.7 *Let* $\mathbf{X}$ *and* $\mathbf{Y}$ *be* $n \times n$ *matrices, with* $rank\,(\mathbf{X}) = rank\,(\mathbf{Y}) = p$, $p \leq n$, $\mathbf{A} = \mathbf{XX}'$ *and* $\mathbf{B} = \mathbf{YY}'$. *Then:*
(a) $0 \leq rank\,(\mathbf{A} \wedge \mathbf{B}) \leq \min\{2p, n\}$ *and* $rank\,(\mathbf{A} \wedge \mathbf{B}) = 0$ *iff* $\mathbf{AB} = \mathbf{0}$,
(b) $0 \leq rank\,(\mathbf{A} \vee \mathbf{B}) \leq \min\{4p, n\}$ *and* $rank\,(\mathbf{A} \vee \mathbf{B}) = 0$ *iff* $\mathbf{A} = \mathbf{B} = \mathbf{0}$.

PROOF. The non-negativeness of the rank of any matrix can directly be derived from the definition. It is also evident from the definition that the rank of any $n \times p$ matrix cannot exceed $\min\{p, n\} = p$. To prove the second inequality in the first statement of (a), two additional properties are useful (assume that the dimensions of the following matrices allow the operations). Let $\mathbf{C}, \mathbf{D}$ be two matrices, then

$$rank\,(\mathbf{CD}) \leq \min\left(rank\,(\mathbf{C}), rank\,(\mathbf{D})\right), \tag{2.6}$$

$$rank\,(\mathbf{C} + \mathbf{D}) \leq rank\,(\mathbf{C}) + rank\,(\mathbf{D}). \tag{2.7}$$

From the symmetry of $\mathbf{A}$ and $\mathbf{B}$ we have that $\mathbf{B}^{\frac{1}{2}}\mathbf{A}^{\frac{1}{2}}$ and $\mathbf{A}^{\frac{1}{2}}\mathbf{B}^{\frac{1}{2}}$ have the same rank. From the definition of $\mathbf{A} \wedge \mathbf{B}$ and (2.6) applied to $\mathbf{A}^{\frac{1}{2}}\mathbf{B}^{\frac{1}{2}}$ and $\mathbf{B}^{\frac{1}{2}}\mathbf{A}^{\frac{1}{2}}$ we see that the (common) rank satisfies

$$rank\left(\mathbf{A}^{\frac{1}{2}}\mathbf{B}^{\frac{1}{2}}\right) \leq \min\{rank\,(\mathbf{A}), rank\,(\mathbf{B})\} \leq p.$$

As $\mathbf{A}^{\frac{1}{2}}\mathbf{B}^{\frac{1}{2}} + \mathbf{B}^{\frac{1}{2}}\mathbf{A}^{\frac{1}{2}}$ is $n \times n$, from property (2.7) we have

$$rank\left(\mathbf{A}^{\frac{1}{2}}\mathbf{B}^{\frac{1}{2}} + \mathbf{B}^{\frac{1}{2}}\mathbf{A}^{\frac{1}{2}}\right) \leq \min\{2p, n\}.$$

The proof of the second inequality in the first statement of (b) can be obtained by applying (2.7) and (a) to $\mathbf{A} \vee \mathbf{B} = \mathbf{A} + \mathbf{B} - \mathbf{A} \wedge \mathbf{B}$. Thus

$$rank\,(\mathbf{A} \vee \mathbf{B}) \leq p + p + \min\{2p, n\}\,.$$

The second statement in (a) follows from the fact that $rank\,(\mathbf{A} \wedge \mathbf{B}) = 0$ iff $\mathbf{A} \wedge \mathbf{B} = \mathbf{0}$. And $\mathbf{A} \wedge \mathbf{B} = \mathbf{0}$ iff $\mathbf{AB} = \mathbf{0}$ (see Proposition 2.3.4, (b)).
To prove the second statement in (b), analogously to the proof given for (a), note that $rank\,(\mathbf{A} \vee \mathbf{B}) = 0$ iff $\mathbf{A} \vee \mathbf{B} = \mathbf{0}.$ This is equivalent to $\mathbf{A} = \mathbf{B} = \mathbf{0}$, as shown in Proposition 2.3.4. □

2.3.4 Trace

The following results on traces of matrices involved in the definitions of the intersection and union can easily be obtained from well-known general results on traces (see, for instance, [47]). In the following, $\mathbf{A}, \mathbf{B} \in \mathcal{M}_n$ are such that $\mathbf{A} = \mathbf{XX}' = \mathbf{U\Lambda U}'$, $\mathbf{B} = \mathbf{YY}' = \mathbf{VMV}'$, where $\mathbf{\Lambda}$ and $\mathbf{M}$ are diagonal matrices with the eigenvalues. Matrix $\mathbf{A}$ satisfies:

(Pr. 1) $tr\,(\mathbf{A}) = tr\,(\mathbf{A}') = \sum_{i=1}^{n} \lambda_i = \sum_{i=1}^{n} \sum_{j=1}^{p} x_{ij}^2 \geq 0,$
(Pr . 2) $tr\,(\mathbf{AB}) \geq 0$ and $tr\,(\mathbf{AB}) = \mathbf{0}$ iff $\mathbf{AB} = \mathbf{0}$,
(Pr . 3) $tr\,(a\mathbf{A}+b\mathbf{B}) = a\; tr\,(\mathbf{A}) + b\; tr\,(\mathbf{B})\,.$

Results on the traces of the intersection and union of the $n \times n$ symmetric matrices $\mathbf{A}$ and $\mathbf{B}$, are given next.

Proposition 2.3.8 *The traces of the intersection* $\mathbf{A} \wedge \mathbf{B}$ *and union* $\mathbf{A} \vee \mathbf{B}$ *satisfy:*
(a) $tr\,(\mathbf{A} \wedge \mathbf{B}) \geq 0,$ *and* $tr\,(\mathbf{A} \wedge \mathbf{B}) = 0$ *iff* $\mathbf{AB} = \mathbf{0}$,
(b) $tr\,(\mathbf{A} \vee \mathbf{B}) \geq 0,$ *and* $tr\,(\mathbf{A} \vee \mathbf{B}) = 0$ *iff* $\mathbf{A} = \mathbf{B} = \mathbf{0}$.

Proof. Property (Pr .2) applied to $\mathbf{A}^{\frac{1}{2}}\mathbf{B}^{\frac{1}{2}}$ gives the non-negativeness of the trace of this $n \times n$ matrix. The same result is obtained for $\mathbf{B}^{\frac{1}{2}}\mathbf{A}^{\frac{1}{2}}$, from (Pr .1) . From (Pr .3) ,

$$\begin{aligned} tr\,(\mathbf{A} \wedge \mathbf{B}) &= \frac{1}{2} tr\left(\mathbf{A}^{\frac{1}{2}}\mathbf{B}^{\frac{1}{2}}\right) + \frac{1}{2} tr\left(\mathbf{B}^{\frac{1}{2}}\mathbf{A}^{\frac{1}{2}}\right) \\ &= tr\left(\mathbf{A}^{\frac{1}{2}}\mathbf{B}^{\frac{1}{2}}\right) \geq 0 \end{aligned}$$

Then the second statement in (a) follows from (Pr .2) and Corollary 2.2.4.
Now, since $\mathbf{A} \vee \mathbf{B} \in \mathcal{M}_n$, the non-negativity of the trace is direct. Obviously, if $\mathbf{C} \in \mathcal{M}_n$ then $tr(\mathbf{C}) = 0$ iff $\mathbf{C} = \mathbf{0}$. Thus, from Proposition 2.3.4

$$\mathbf{A} \vee \mathbf{B} = \mathbf{0} \text{ iff } \mathbf{A} = \mathbf{B} = \mathbf{0},$$

and Part (b) is proved. □

The properties found in the previous sections are analogous to those that appear in the context of the theory of vector spaces. Now we must define the concept of *inclusion*. We may expect that if a matrix (subspace) $\mathbf{A}$ is "included" in another $\mathbf{B}$, then the intersection should be the "smallest" matrix and the union, the "largest" in a sense.

2.4 Binary relations between matrices

In this section, we define an order relation among symmetric p.(s.)d. matrices under certain restrictions. The order of the (real) eigenvalues is used.

2.4.1 Equivalence relation

Let $\mathbf{A}, \mathbf{B} \in \mathcal{M}_n$. Suppose that there exists a vector $\mathbf{u} \in \mathbb{R}^n$ such that

$$\mathbf{A}\mathbf{u} = \lambda \mathbf{u}, \qquad \mathbf{B}\mathbf{u} = \mu \mathbf{u},$$

for some $\lambda, \mu \in \mathbb{R}$. We will say that $\mathbf{u}$ is a common eigenvector of $\mathbf{A}$ and $\mathbf{B}$. In other words, $\mathbf{A}, \mathbf{B}$ share the same eigenvector $\mathbf{u}$.

An interesting result follows from this definition.

Lemma 2.4.1 *If* $\mathbf{A}$ *and* $\mathbf{B}$ *share an eigenvector* $\mathbf{u}$*, with eigenvalues* λ, μ*, respectively, then:*
1. $\mathbf{A} \wedge \mathbf{B}$ *and* $\mathbf{A} \vee \mathbf{B}$ *share the same eigenvector* $\mathbf{u}$.
2. $\lambda^{\frac{1}{2}}\mu^{\frac{1}{2}}$ *and* $\lambda + \mu - \lambda^{\frac{1}{2}}\mu^{\frac{1}{2}}$ *are the eigenvalues of* $\mathbf{A} \wedge \mathbf{B}$ *and* $\mathbf{A} \vee \mathbf{B}$ *corresponding to this eigenvector* $\mathbf{u}$.

PROOF. $\lambda^{\frac{1}{2}}$ and $\mu^{\frac{1}{2}}$ are eigenvalues of $\mathbf{A}^{\frac{1}{2}}$ and $\mathbf{B}^{\frac{1}{2}}$, respectively, with common eigenvector $\mathbf{u}$ (see, [47]). Then

$$\begin{aligned}(\mathbf{A}\wedge\mathbf{B})\,\mathbf{u} &= \frac{1}{2}\left(\mathbf{A}^{\frac{1}{2}}\mathbf{B}^{\frac{1}{2}}+\mathbf{B}^{\frac{1}{2}}\mathbf{A}^{\frac{1}{2}}\right)\mathbf{u}\\ &= \frac{1}{2}\left(\mathbf{A}^{\frac{1}{2}}\mathbf{B}^{\frac{1}{2}}\mathbf{u}+\mathbf{B}^{\frac{1}{2}}\mathbf{A}^{\frac{1}{2}}\mathbf{u}\right)\\ &= \frac{1}{2}\left(\mathbf{A}^{\frac{1}{2}}\mu^{\frac{1}{2}}\mathbf{u}+\mathbf{B}^{\frac{1}{2}}\lambda^{\frac{1}{2}}\mathbf{u}\right)\\ &= \frac{1}{2}\left(\lambda^{\frac{1}{2}}\mu^{\frac{1}{2}}\mathbf{u}+\mu^{\frac{1}{2}}\lambda^{\frac{1}{2}}\mathbf{u}\right)\\ &= \lambda^{\frac{1}{2}}\mu^{\frac{1}{2}}\mathbf{u}.\end{aligned}$$

Similarly

$$\begin{aligned}(\mathbf{A}\vee\mathbf{B})\,\mathbf{u} &= (\mathbf{A}+\mathbf{B}-\mathbf{A}\wedge\mathbf{B})\,\mathbf{u}\\ &= \mathbf{A}\mathbf{u}+\mathbf{B}\mathbf{u}-(\mathbf{A}\wedge\mathbf{B})\,\mathbf{u}\\ &= \lambda\mathbf{u}+\mu\mathbf{u}-\lambda^{\frac{1}{2}}\mu^{\frac{1}{2}}\mathbf{u}\\ &= \left(\lambda+\mu-\lambda^{\frac{1}{2}}\mu^{\frac{1}{2}}\right)\mathbf{u}.\end{aligned}$$

□

Definition 2.4.2 *Let $\mathbf{A},\mathbf{B}\in\mathcal{M}_n$, with normalized eigenvectors $\{\mathbf{u}_i\}_{i=1}^n$ and $\{\mathbf{v}_i\}_{i=1}^n$, respectively. We say that $\mathbf{A}$ is equivalent to $\mathbf{B}$ if $\{\mathbf{u}_i\}_{i=1}^n=\{\mathbf{v}_i\}_{i=1}^n$. We write $\mathbf{A}\sim\mathbf{B}$.*

Thus $\mathbf{A},\mathbf{B}$ are related by the equivalence relation $\sim$ iff they share all their eigenvectors. This implies that we can rearrange the eigenvectors

$$\mathbf{V}^{(*)}=\left[\mathbf{v}^{(1)},...,\mathbf{v}^{(n)}\right]=\mathbf{U},$$

where $\mathbf{v}^{(i)}$ is the column vector after rearranging.

Given an $n\times n$ orthogonal matrix $\mathbf{U}$, we can define the equivalence classes by

$$[\mathbf{U}]=\{\mathbf{C}\in\mathcal{M}_n\mid\mathbf{C}=\mathbf{U}\boldsymbol{\Lambda}\mathbf{U}',\ \text{for some diagonal matrix }\boldsymbol{\Lambda}\geq 0\}.$$

2.4.2 Partial order relation

Definition 2.4.3 *Let $\mathbf{A},\mathbf{B}\in\mathcal{M}_n$, $\mathbf{A}\sim\mathbf{B}$ with eigenvalues $\{\lambda_i\}_{i=1}^n$ and $\{\mu_i\}_{i=1}^n$, respectively and common matrix of eigenvectors $\mathbf{U}$. We say that $\mathbf{A}$ is smaller than*

$\mathbf{B}$, *and we denote this partial order by* $\mathbf{A} \lesssim \mathbf{B}$, *if* $\mathbf{A} \sim \mathbf{B}$ *and*

$$\lambda_i \leq \mu_i, \qquad i = 1, ..., n,$$

where the index i *refers to the* $i-th$ *eigenvector.*

Example 2.4.4 *The following matrices are ordered*

$$\mathbf{A} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \lesssim \mathbf{B} = \begin{pmatrix} 1 & 0 \\ 0 & 0.5 \end{pmatrix} \lesssim \mathbf{I}_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \lesssim \mathbf{C} = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}.$$

2.4.3 Properties of equivalent matrices

Intersection and union

We can obtain the union and intersection of matrices belonging to the same equivalence class $[\mathbf{U}]$.

Lemma 2.4.5 *Let* $\mathbf{A} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}'$ *and* $\mathbf{B} = \mathbf{U}\mathbf{M}\mathbf{U}'$ *matrices of* $\mathcal{M}_n$ *such that* $\mathbf{A} \sim \mathbf{B}$, *with* $\mathbf{U}$, *the (common) orthogonal matrix containing the eigenvectors,* $\boldsymbol{\Lambda}$ *and* $\mathbf{M}$ *diagonal. Then* $\mathbf{A} \wedge \mathbf{B}$, $\mathbf{A} \vee \mathbf{B} \in \mathcal{M}_n$ *and*

$$\mathbf{A} \wedge \mathbf{B} = \mathbf{U}\boldsymbol{\Gamma}\mathbf{U}', \quad \mathbf{A} \vee \mathbf{B} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{U}',$$

where $\boldsymbol{\Gamma} = \boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{M}^{\frac{1}{2}}$ *and* $\boldsymbol{\Sigma} = \boldsymbol{\Lambda} + \mathbf{M} - \boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{M}^{\frac{1}{2}}$ *are diagonal.*

PROOF. Recall that $\mathbf{A} \vee \mathbf{B} \in \mathcal{M}_n$ and the symmetry of $\mathbf{A} \wedge \mathbf{B}$ was proved for every $\mathbf{A}, \mathbf{B} \in \mathcal{M}_n$ in Proposition 2.3.2. Then it suffices to obtain the spectral decomposition of these matrices. $\mathbf{U}$ is orthogonal and $\boldsymbol{\Lambda}^{\frac{1}{2}}$, $\mathbf{M}^{\frac{1}{2}}$ are diagonal, so $\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{M}^{\frac{1}{2}} = \mathbf{M}^{\frac{1}{2}}\boldsymbol{\Lambda}^{\frac{1}{2}}$ and

$$\begin{aligned} \mathbf{A} \wedge \mathbf{B} &= \frac{1}{2}\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{U}'\mathbf{U}\mathbf{M}^{\frac{1}{2}}\mathbf{U}' + \mathbf{U}\mathbf{M}^{\frac{1}{2}}\mathbf{U}'\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{U}'\right) \\ &= \frac{1}{2}\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{M}^{\frac{1}{2}}\mathbf{U}' + \mathbf{U}\mathbf{M}^{\frac{1}{2}}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{U}'\right) \\ &= \frac{1}{2}\left(\mathbf{U}\left(\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{M}^{\frac{1}{2}} + \mathbf{M}^{\frac{1}{2}}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)\mathbf{U}'\right) \\ &= \mathbf{U}\boldsymbol{\Gamma}\mathbf{U}', \end{aligned} \tag{2.8}$$

with $\boldsymbol{\Gamma}$ the diagonal matrix with elements $\gamma_i = \lambda_i^{\frac{1}{2}}\mu_i^{\frac{1}{2}}$, $i = 1, ..., n$.

Similarly,

$$\begin{aligned} \mathbf{A} \vee \mathbf{B} &= \mathbf{U} \mathbf{\Lambda} \mathbf{U}' + \mathbf{U} \mathbf{M} \mathbf{U}' - \mathbf{U} \mathbf{\Gamma} \mathbf{U}' \\ &= \mathbf{U} \left(\mathbf{\Lambda} + \mathbf{M} - \mathbf{\Gamma} \right) \mathbf{U}' \\ &= \mathbf{U} \mathbf{\Sigma} \mathbf{U}', \end{aligned} \tag{2.9}$$

with $\mathbf{\Sigma}$, the diagonal matrix with elements $\sigma_i = \lambda_i + \mu_i - \lambda_i^{\frac{1}{2}} \mu_i^{\frac{1}{2}}$, $i = 1, ..., n$. Alternatively, we may prove this result by applying 2) of Lemma 2.4.1 to the eigenvectors of $\mathbf{U}$.

□

The properties of the ordered matrices are summarized by the following result. However, note that $\mathbf{A}$ is "smaller" than $\mathbf{A} \wedge \mathbf{B}$, whereas $\mathbf{A} \vee \mathbf{B}$ is not "larger" than $\mathbf{B}$.

Proposition 2.4.6 *Let* $\mathbf{A}, \mathbf{B} \in \mathcal{M}_n$, *with* $\mathbf{A} \lesssim \mathbf{B}$. *Then*
(i) $\mathbf{A} \wedge \mathbf{B} \sim \mathbf{A}$ $(\sim \mathbf{B})$ *and also* $\mathbf{A} \vee \mathbf{B} \sim \mathbf{A}$ $(\sim \mathbf{B})$,
(ii) $\mathbf{A} \lesssim \mathbf{A} \wedge \mathbf{B} \lesssim \mathbf{A} \vee \mathbf{B} \lesssim \mathbf{B}$.

Proof. Let $\mathbf{A} = \mathbf{U} \mathbf{\Lambda} \mathbf{U}'$ and $\mathbf{B} = \mathbf{U} \mathbf{M} \mathbf{U}'$. Part *(i)* is a direct corollary of Lemma 2.4.5. To prove *(ii)*, we may use the expression for the eigenvalues of the intersection and the union of ordered matrices obtained in Lemma 2.4.5, (2.8), (2.9), and the inequalities

$$\lambda_i \leq \lambda_i^{\frac{1}{2}} \mu_i^{\frac{1}{2}} \leq \lambda_i + \mu_i - \lambda_i^{\frac{1}{2}} \mu_i^{\frac{1}{2}} \leq \mu_i, \qquad i = 1, ..., n.$$

To prove these inequalities it suffices to consider that $\mathbf{A} \lesssim \mathbf{B}$ iff $\lambda_i \leq \mu_i$ and the fact that $\mathbf{A}, \mathbf{B}$ are p.(s.)d.. Thus, for $i = 1, ..., n$:

$$\begin{aligned} \lambda_i &= \lambda_i^{\frac{1}{2}} \lambda_i^{\frac{1}{2}} \leq \lambda_i^{\frac{1}{2}} \mu_i^{\frac{1}{2}}, \\ 0 &\leq \left(\lambda_i^{\frac{1}{2}} - \mu_i^{\frac{1}{2}} \right)^2 = \lambda_i + \mu_i - 2 \lambda_i^{\frac{1}{2}} \mu_i^{\frac{1}{2}}, \end{aligned}$$

and this is equivalent to

$$\lambda_i^{\frac{1}{2}} \mu_i^{\frac{1}{2}} \leq \lambda_i + \mu_i - \lambda_i^{\frac{1}{2}} \mu_i^{\frac{1}{2}}.$$

Finally,

$$\lambda_i + \mu_i - \lambda_i^{\frac{1}{2}} \mu_i^{\frac{1}{2}} \leq \lambda_i + \mu_i - \lambda_i^{\frac{1}{2}} \lambda_i^{\frac{1}{2}} = \mu_i.$$

□

These computations have some implications, as illustrated below.

Example 2.4.7 *Let* $\mathbf{u}_1, \mathbf{u}_2,$ *and* $\mathbf{u}_3$ *be orthogonal vectors in* $\mathbb{R}^3$ *and* $\lambda_1 > \lambda_2 > \lambda_3 > 0$. *Consider the matrices belonging to* $\mathcal{M}_3$

$$\begin{aligned}
\mathbf{M} &= \lambda_1 \mathbf{u}_1 \mathbf{u}_1' + \lambda_2 \mathbf{u}_2 \mathbf{u}_2' + \lambda_3 \mathbf{u}_3 \mathbf{u}_3', \\
\mathbf{A} &= \lambda_1 \mathbf{u}_1 \mathbf{u}_1' + 0 \mathbf{u}_2 \mathbf{u}_2' + 0 \mathbf{u}_3 \mathbf{u}_3', \\
\mathbf{B} &= \lambda_2 \mathbf{u}_2 \mathbf{u}_2' + 0 \mathbf{u}_1 \mathbf{u}_1' + 0 \mathbf{u}_3 \mathbf{u}_3', \\
\mathbf{C} &= \lambda_1 \mathbf{u}_1 \mathbf{u}_1' + \lambda_3 \mathbf{u}_3 \mathbf{u}_3' + 0 \mathbf{u}_2 \mathbf{u}_2'.
\end{aligned}$$

The "reference" matrix is $\mathbf{M} = \mathbf{U} \boldsymbol{\Lambda} \mathbf{U}'$, *with* $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3]$ *and* $\boldsymbol{\Lambda} = diag(\lambda_1, \lambda_2, \lambda_3)$. *Observe that*

$$\mathbf{A} = \mathbf{U} \boldsymbol{\Lambda}_1 \mathbf{U}', \qquad \mathbf{B} = \mathbf{U} \boldsymbol{\Lambda}_2 \mathbf{U}', \qquad \mathbf{C} = \mathbf{U} \boldsymbol{\Lambda}_{13} \mathbf{U}',$$

with $\boldsymbol{\Lambda}_1 = diag(\lambda_1, 0, 0)$, $\boldsymbol{\Lambda}_2 = diag(0, \lambda_2, 0)$, *and* $\boldsymbol{\Lambda}_{13} = diag(\lambda_1, 0, \lambda_3)$ *and these matrices may be related. Thus we cannot compare* $\mathbf{B}$ *to* $\mathbf{A}$ *and* $\mathbf{B}$ *to* $\mathbf{C}$, *whereas* $\mathbf{A}, \mathbf{B}$, *and* $\mathbf{C}$ *are smaller than* $\mathbf{M}$ *and* $\mathbf{A} \lesssim \mathbf{C} \lesssim \mathbf{M}$. *Next, their intersections and unions are found, applying the previous lemma:*

$$\begin{aligned}
\mathbf{A} \wedge \mathbf{B} &= \mathbf{U} \boldsymbol{\Lambda}_1 \boldsymbol{\Lambda}_2 \mathbf{U}' = \mathbf{0}, \\
\mathbf{A} \wedge \mathbf{C} &= \mathbf{U} \boldsymbol{\Lambda}_1 \boldsymbol{\Lambda}_{13} \mathbf{U}' = \mathbf{A}, \\
\mathbf{A} \wedge \mathbf{M} &= \mathbf{U} \boldsymbol{\Lambda}_1 \boldsymbol{\Lambda} \mathbf{U}' = \mathbf{A}, \\
\mathbf{B} \wedge \mathbf{C} &= \mathbf{U} \boldsymbol{\Lambda}_2 \boldsymbol{\Lambda}_{13} \mathbf{U}' = \mathbf{0}, \\
\mathbf{B} \wedge \mathbf{M} &= \mathbf{U} \boldsymbol{\Lambda}_2 \boldsymbol{\Lambda} \mathbf{U}' = \mathbf{B}, \\
\mathbf{C} \wedge \mathbf{M} &= \mathbf{U} \boldsymbol{\Lambda}_{13} \boldsymbol{\Lambda} \mathbf{U}' = \mathbf{C}.
\end{aligned}$$

Hence,

$$\begin{aligned}
\mathbf{A} \vee \mathbf{B} &= \mathbf{A} + \mathbf{B} - \mathbf{0} &&= \mathbf{A} + \mathbf{B}, \\
\mathbf{A} \vee \mathbf{C} &= \mathbf{A} + \mathbf{C} - \mathbf{A} &&= \mathbf{C}, \\
\mathbf{A} \vee \mathbf{M} &= \mathbf{A} + \mathbf{M} - \mathbf{A} &&= \mathbf{M}, \\
\mathbf{B} \vee \mathbf{C} &= \mathbf{B} + \mathbf{C} - \mathbf{0} &&= \mathbf{B} + \mathbf{C}, \\
\mathbf{B} \vee \mathbf{M} &= \mathbf{B} + \mathbf{M} - \mathbf{B} &&= \mathbf{M}, \\
\mathbf{C} \vee \mathbf{M} &= \mathbf{C} + \mathbf{M} - \mathbf{C} &&= \mathbf{M}.
\end{aligned}$$

Thus, the non-ordered matrices turn out to be orthogonal (the intersection is $\mathbf{0}$ *and the union is the direct sum). On the other hand, the ordered matrices present the expected properties, for instance,* $\mathbf{A} \lesssim \mathbf{C}$, *being* $\mathbf{A} \wedge \mathbf{C} = \mathbf{A}$ $(\lesssim \mathbf{C})$ *and* $\mathbf{A} \vee \mathbf{C} = \mathbf{C}$ $(\lesssim \mathbf{C})$. *Finally, observe that the intersection and union matrices (even* $\mathbf{0}$, $\mathbf{A} + \mathbf{B}$, *and* $\mathbf{B} + \mathbf{C}$*) belong to the same equivalence class* $[\mathbf{U}]$, *all of them being smaller than* $\mathbf{M}$.

Bounds for ordered matrices

We investigate the existence of bounds (minimum and maximum elements) within a subclass of ordered matrices included in a class of matrices related by the above equivalence relation. However, the largest eigenvalue might be unbounded, so we focus on the case that the largest eigenvalue is finite and bounded. We are interested in fixing a finite (arbitrary) upper bound for the largest eigenvalue. Without loss of generality we can normalize this bound to 1.

Proposition 2.4.8 *Let* $[\mathbf{U}^1]$ *be the class of all the p.(s.)d. symmetric* $n \times n$ *matrices whose eigenvectors are the columns of the orthogonal matrix* $\mathbf{U}$ *and the corresponding eigenvalues are bounded by* 1*. Then if* $\mathbf{0}_n \equiv \mathbf{0}$*,* $\mathbf{I}_n \equiv \mathbf{I}$,
(i) $\mathbf{0}, \mathbf{I} \in [\mathbf{U}^1]$;
(ii) $\mathbf{0} \lesssim \mathbf{A} \lesssim \mathbf{I}$ *for all* $\mathbf{A} \in [\mathbf{U}^1]$.

PROOF. (i) follows from $\mathbf{0} = \mathbf{U0U}'$, $\mathbf{I} = \mathbf{UIU}'$. To avoid confusions, let us denote $\lambda_k(\mathbf{A})$ the $k-$th eigenvalue of the matrix $\mathbf{A}$. To prove (ii) note that, for every $\mathbf{A} \in [\mathbf{U}^1]$, $\lambda_1(\mathbf{A}) \leq 1$ and $\lambda_n(\mathbf{A}) \geq 0$, so:

$$\begin{aligned} \lambda_n(\mathbf{0}) &= ... = \lambda_1(\mathbf{0}) = 0 \leq \lambda_n(\mathbf{A}) \leq ... \leq \lambda_1(\mathbf{A}) \leq 1 = \lambda_n(\mathbf{I}) = ... = \lambda_1(\mathbf{I}), \\ &\Rightarrow \lambda_i(\mathbf{0}) \leq \lambda_i(\mathbf{A}) \leq \lambda_i(\mathbf{I}), \qquad \forall i = 1, ..., n. \end{aligned}$$

□

Rank

The rank of $\mathbf{A} \in \mathcal{M}_n$ is the number of non-null eigenvalues (see, for instance, [51]). The proposition below gives the rank of $\mathbf{A}$ from the rank of the diagonal matrix of its eigenvalues.

Proposition 2.4.9 *Let* $\mathbf{A}, \mathbf{B} \in \mathcal{M}_n$*,* $\mathbf{A} \lesssim \mathbf{B}$*. Then*

$$rank(\mathbf{A}) \leq rank(\mathbf{B}) \leq n,$$

with equality in the left if both matrices have the same number of non-zero eigenvalues. The second equality holds if $\mathbf{B}$ *is p.d..*

PROOF. If k, l are the number of positive eigenvalues of $\mathbf{A}, \mathbf{B}$, respectively, then $k \leq l$ if $\lambda_i(\mathbf{A}) \leq \lambda_i(\mathbf{B})$, $i = 1, ..., n$. □

Trace

Proposition 2.4.10 *Let* $\mathbf{A},\mathbf{B} \in \mathcal{M}_n$*, with* $\mathbf{A}\sim\mathbf{B}$*, such that* $\mathbf{A}=\mathbf{U}\boldsymbol{\Lambda}\mathbf{U}'$*,* $\mathbf{B}=\mathbf{UMU}'$*, where* $\boldsymbol{\Lambda}=diag\left(\lambda_i\right)_{i=1,\ldots,n}$*,* $\mathbf{M}=diag\left(\mu_i\right)_{i=1,\ldots,n}$*. Then:*
(a) $tr\left(\mathbf{A}\wedge\mathbf{B}\right)=\sum\lambda_i^{\frac{1}{2}}\mu_i^{\frac{1}{2}}$*,*
(b) $tr\left(\mathbf{A}\vee\mathbf{B}\right)=\sum\left(\lambda_i+\mu_i-\lambda_i^{\frac{1}{2}}\mu_i^{\frac{1}{2}}\right).$

Proof. This is a consequence of (Pr. 1) applied to the set of eigenvalues of the intersection and union obtained in Lemma 2.4.5. □

Lemma 2.4.11 *Let* $\mathbf{A}\lesssim\mathbf{B}$ *be matrices satisfying the hypothesis of Proposition 2.4.10. Then* $tr\left(\mathbf{A}\right)\leq tr\left(\mathbf{B}\right).$

Proof. $\mathbf{A}\lesssim\mathbf{B}$ is equivalent to $\lambda_i\leq\mu_i$, $i=1,...,n$, so $\sum\lambda_i\leq\sum\mu_i$. □

Corollary 2.4.12 *Let* $[\mathbf{U}]$ *be the class of the* $n\times n$ *symmetric matrices sharing the eigenvectors. If the subclass* $[\mathbf{U}^1]\subset[\mathbf{U}]$ *consists of the matrices having their largest eigenvalue* $\lambda_1\leq 1$*, then any* $\mathbf{M}\in[\mathbf{U}^1]$ *satisfies:*

$$0\leq tr\left(\mathbf{M}\right)\leq n.$$

Proof. We have $\mathbf{0}\leq\mathbf{M}\leq\mathbf{I}$. Then apply Lemma 2.4.11. □

Theorem 2.4.13 *Let* $\mathbf{A},\mathbf{B}\in\mathcal{M}_n$*, with* $\mathbf{A}\lesssim\mathbf{B}$*. If* $\mathbf{A}=\mathbf{U}\boldsymbol{\Lambda}\mathbf{U}'$*,* $\mathbf{B}=\mathbf{UMU}'$*, where* $\boldsymbol{\Lambda}=diag\left(\lambda_i\right)$*,* $\mathbf{M}=diag\left(\mu_i\right)$*, then*

$$tr\left(\mathbf{A}\right)\leq tr\left(\mathbf{A}\wedge\mathbf{B}\right)\leq tr\left(\mathbf{A}\vee\mathbf{B}\right)\leq tr\left(\mathbf{B}\right).$$

Proof. From Proposition 2.4.6 we have

$$\mathbf{A}\lesssim\mathbf{A}\wedge\mathbf{B}\lesssim\mathbf{A}\vee\mathbf{B}\lesssim\mathbf{B}.$$

The proof follows from Lemma 2.4.11. □

2.5 An application to Related Metric Scaling

We conclude this chapter presenting some theoretical results of an extension of multidimensional scaling. *Multidimensional Scaling* is a well-known method designed to construct a configuration of n points on the Euclidean plane from a distance matrix related to n objects (see [14], [81]).

Related Metric Scaling (RMS) is an extension. This is a distance based method of multivariate analysis proposed in [15], [28]. Some applications can be found in [2], [3], [4], [37]. The present chapter on operations on matrices is in fact a framework where the use of this technique appears in a natural way.

RMS is useful when two distance matrices are available on the same objects, and a joint representation has interest. This method defines a joint distance matrix from two distance matrices. Let δ_A and δ_B be two related Euclidean distances on n objects and $\mathbf{\Delta}_A$ and $\mathbf{\Delta}_B$ the respective distance matrices. The inner product matrices associated with $\mathbf{\Delta}_A$ and $\mathbf{\Delta}_B$ are the $n \times n$ matrices

$$\mathbf{A} = \mathbf{X}\mathbf{X}', \qquad \mathbf{B} = \mathbf{Y}\mathbf{Y}',$$

where $\mathbf{X}$ and $\mathbf{Y}$ are the centred $n \times p$ matrices of coordinates, with $p = rank\,(\mathbf{X}) = rank\,(\mathbf{Y})$, and $p \leq n$.

Matrices $\mathbf{A}$ and $\mathbf{B}$ are the double-centred product matrices and can be obtained from $\mathbf{\Delta}_A$ and $\mathbf{\Delta}_B$, by,

$$\mathbf{A} = \mathbf{H}\left(-\frac{1}{2}\mathbf{\Delta}_A^{(2)}\right)\mathbf{H}, \qquad \mathbf{B} = \mathbf{H}\left(-\frac{1}{2}\mathbf{\Delta}_B^{(2)}\right)\mathbf{H} \tag{2.10}$$

being $\mathbf{H}$ the centring matrix and $\mathbf{\Delta}_A^{(2)} = (\delta_A^2\,(i,j))$, $\mathbf{\Delta}_B^{(2)} = (\delta_B^2\,(i,j))$. Consider the spectral decompositions

$$\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}', \qquad \mathbf{B} = \mathbf{V}\mathbf{M}\mathbf{V}',$$

and the square roots

$$\mathbf{A}^{\frac{1}{2}} = \mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{U}', \qquad \mathbf{B}^{\frac{1}{2}} = \mathbf{V}\mathbf{M}^{\frac{1}{2}}\mathbf{V}'.$$

Then the above matrices of coordinates are

$$\mathbf{X} = \mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}, \qquad \mathbf{Y} = \mathbf{V}\mathbf{M}^{\frac{1}{2}}.$$

$\mathbf{X}, \mathbf{Y}$ are the $n \times p$ matrices of principal coordinates of the n objects relative to the distances matrices $\mathbf{\Delta}_A$, $\mathbf{\Delta}_B$, respectively. That is, the row vectors $\mathbf{x}_i$ and $\mathbf{y}_i$ are the principal coordinates of the $i-th$ object relative to $\mathbf{\Delta}_A$ and $\mathbf{\Delta}_B$.

The inner product matrix associated with the joint distance

We define the joint distance δ_{AB} between the $i-th$ and $j-th$ objects by

$$\delta_{AB}^2(i,j) = \delta_A^2(i,j) + \delta_B^2(i,j) - \tau(i,j), \tag{2.11}$$

being

$$\tau(i,j) = (\mathbf{x}_i - \mathbf{x}_j)\,\mathbf{\Lambda}^{-\frac{1}{2}}\mathbf{X}'\mathbf{Y}\mathbf{M}^{-\frac{1}{2}}(\mathbf{y}_i - \mathbf{y}_j)'.$$

The quantity $\tau(i,j)$ encodes the redundance between $\mathbf{\Delta}_A$ and $\mathbf{\Delta}_B$.

Theorem 2.5.1 *Let* $\mathbf{A}$ *and* $\mathbf{B}$ *be the inner product matrices associated to the distances* δ_A *and* δ_B*, respectively. The inner product matrix associated with the joint distance* δ_{AB} *is the p.(s.)d. matrix*

$$\mathbf{G}_{AB} = \mathbf{A} + \mathbf{B} - \frac{1}{2}\left(\mathbf{A}^{\frac{1}{2}}\mathbf{B}^{\frac{1}{2}} + \mathbf{B}^{\frac{1}{2}}\mathbf{A}^{\frac{1}{2}}\right).$$

This matrix is obtained by applying principal coordinates analysis on δ_{AB} via diagonalization of $\mathbf{G}_{AB}$, (see (2.10)), on $\mathbf{\Delta}_{AB}^{(2)} = (\delta_{AB}^2(i,j))$. Clearly, $\mathbf{G}_{AB}$, the inner product matrix related to the joint distance (2.11) is the *union* of the inner product matrices related to the marginal distances, whereas the matrix related to the above redundance quantity is the *intersection*.

We next derive some properties of the joint distance in order to obtain analogous properties of the associated inner product matrices. We use the following definitions:

Definition 2.5.2 *Let* $\mathbf{\Delta}_A$ *and* $\mathbf{\Delta}_B$ *be two distance matrices and let* $\mathbf{X}$ *and* $\mathbf{Y}$ *be the corresponding principal coordinates. We say that* $\mathbf{\Delta}_A$ *and* $\mathbf{\Delta}_B$ *are orthogonal if* $\mathbf{X}'\mathbf{Y} = \mathbf{0}$.

Definition 2.5.3 *Let* $\mathbf{\Delta} = (\delta_{ij})$ *be a distance matrix. The rank order of* $\mathbf{\Delta}$ *is defined as*

$$\delta_{(i_1 j_1)} \leq \delta_{(i_2 j_2)} \leq \ldots \leq \delta_{(i_m j_m)} \qquad m = n(n-1)/2.$$

The following basic properties of the joint distance δ_{AB} hold:
(a) $\delta_{AB} = \delta_{BA}$.
(b) If δ_A and δ_B are orthogonal, i.e., $\mathbf{X}'\mathbf{Y} = \mathbf{0}$, then $\delta_{AB}^2 = \delta_A^2 + \delta_B^2$.
(c) If $\delta_A = \delta_B$ then $\delta_{AB} = \delta_A = \delta_B$.

(d) If $\delta_A = 0$ then $\delta_{AB} = \delta_B$, and if $\delta_B = 0$ then $\delta_{AB} = \delta_A$.
(d') If δ_A is any distance and δ_B is a constant (i.e., 0 if $i = j$; c if $i \neq j$) then δ_{AB} preserves the rank order associated with δ_A.
Observe that these properties show how δ_{AB} encodes the possible redundance between δ_A and δ_B. It remains invariant if distances are the same, it does not change if one of them is null. If one of them is not informative, then the joint distance is equivalent to the other one.

By means of (2.10), these properties of the joint distance give similar properties for the inner product matrix $\mathbf{G}_{AB}$, which can also be obtained by applying Proposition 2.3.4 directly to $\mathbf{G}_{AB}$. Here $\mathbf{A}, \mathbf{B}$ are the inner product matrices associated with δ_A, δ_B.

Proposition 2.5.4 *The inner product matrix for the joint distance satisfies:*
(a) $\mathbf{G}_{AB} = \mathbf{G}_{BA}$,
(b) *If δ_A and δ_B are orthogonal, that is $\mathbf{AB} = \mathbf{0}$, the redundance is zero and*

$$\mathbf{G}_{AB} = \mathbf{A} + \mathbf{B}.$$

(c) *If $\delta_A = \delta_B$, $\mathbf{A} = \mathbf{B}$, the redundance is total and*

$$\mathbf{G}_{AB} = \mathbf{A}.$$

(d) *If $\delta_B = 0$ then $\delta_{AB} = \delta_A$, the redundance is zero and*

$$\mathbf{G}_{A0} = \mathbf{A}.$$

Lemma 2.5.5 *If δ_A and δ_B share a common principal axis (eigenvector of $\mathbf{A}, \mathbf{B}$) with eigenvalues $\lambda_{(k)}(\mathbf{A})$, $\lambda_{(k)}(\mathbf{B})$, respectively, then:*
1. δ_{AB} shares the same principal axis (eigenvector).
2. The corresponding eigenvalue is

$$\lambda_{(k)}(\mathbf{G}_{AB}) = \lambda_{(k)}(\mathbf{A}) + \lambda_{(k)}(\mathbf{B}) - \left(\lambda_{(k)}(\mathbf{A})\,\lambda_{(k)}(\mathbf{B})\right)^{\frac{1}{2}}.$$

Proof. The proof follows from Lemma 2.4.1 to the common principal axis (eigenvector) of the RMS solution for $\mathbf{G}_{AB}$. □

Further properties are next presented. See [15].

Lemma 2.5.6 *If δ_A and δ_B are Euclidean distances with principal coordinates $\mathbf{X}$, $\mathbf{Y}$, then δ_{AB} is also Euclidean and does not depend on $\mathbf{X}$, $\mathbf{Y}$.*

Note that

$$\mathbf{A}^{\frac{1}{2}}\mathbf{B}^{\frac{1}{2}} = \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{U}'\mathbf{V}\mathbf{M}^{\frac{1}{2}}\mathbf{V}' = \mathbf{X}\mathbf{R}_{\mathbf{AB}}\mathbf{Y}'$$

where

$$\mathbf{R}_{AB} = \mathbf{U}'\mathbf{V} \tag{2.12}$$

is the $p_1 \times p_2$ matrix whose entries are the correlation coefficients between the p_1 columns of $\mathbf{X}$ and the p_2 columns of $\mathbf{Y}$.

Proposition 2.5.7 *Let $\mathbf{R}_{AB}$ be the correlation matrix in (2.12) and let $\mathbf{R}_{BA} = \mathbf{R}'_{AB}$. Then the inner product matrix associated with δ_{AB} can be written as*

$$\mathbf{G}_{AB} = \mathbf{X}\mathbf{X}' + \mathbf{Y}\mathbf{Y}' - \frac{1}{2}\left(\mathbf{X}\mathbf{R}_{\mathbf{AB}}\mathbf{Y}' + \mathbf{Y}\mathbf{R}_{\mathbf{BA}}\mathbf{X}'\right).$$

Remark 2.5.8 (1) *The correlation between the principal axes in the marginal distances is always* 0*. Moreover, zero correlation between principal axes in* $\mathbf{A}$ *and* $\mathbf{B}$ *is equivalent to* $\mathbf{R}_{\mathbf{AB}} = \mathbf{0}$*. Then* $\mathbf{G}_{AB} = \mathbf{X}\mathbf{X}' + \mathbf{Y}\mathbf{Y}'$ *(orthogonality).* (2) *Total correlation between the principal axes in both matrices is equivalent to* $\mathbf{R}_{\mathbf{AB}} = \mathbf{I}_p$*, being* $p = \min\{p_1, p_2\}$.

2.6 Concluding remarks

The intersection and union operations between symmetric matrices provide a theoretical framework useful in *Related Metric Scaling*. These operations have similar properties to the intersection and union of vector subspaces (for example, subspaces spanned by the columns of a matrix). We have defined a partial order relation between positive (semi)definite matrices with interesting results. Nevertheless, further research is required to understand some surprising results. For instance, according to this partial order, if a matrix $\mathbf{A}$ is lower or equal ($\lesssim$) to another matrix $\mathbf{B}$, then $\mathbf{A} \lesssim \mathbf{A} \wedge \mathbf{B} \lesssim \mathbf{A} \vee \mathbf{B} \lesssim \mathbf{B}$. The theoretical aspects developed in this chapter may facilitate the geometric interpretation of *RMS* and may open the way to future research on similar multivariate techniques. The operations bring out the dependence structure underlying the information presented in a multivariate context. A continuous version of this framework is the subject of the next chapter.

Chapter 3

Continuous extensions of some operations with matrices

The union and intersection operations of symmetric non-negative definite matrices introduced in [15] and studied in Chapter 2, are a way of studying multivariate dependences between data matrices of the same dimension, with applications in related multidimensional scaling. When we work with random variables, multidimensional scaling can be extended in a continuous way, (see [25]). Extending this approach from matrices to functions of two variables and using some tools of functional analysis, we study the redundant relation between functions.

3.1 Symmetric positive definite kernels

Let $(\Omega_1, \mathcal{F}_1, \mathcal{M}_1)$ and $(\Omega_2, \mathcal{F}_2, \mathcal{M}_2)$ be two measure spaces.

Definition 3.1.1 *A real valued function $K(s,t)$ which is measurable on the product space $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \times \mathcal{F}_2)$ is said to be a kernel. K is said to be symmetric if for every pair $(x,y) \in \Omega_1 \times \Omega_2$, $K(x,y) = K(y,x)$.*

It is evident from the definition that symmetry implies equality of sets Ω_1 and Ω_2.

As in the discrete case (finite matrices), we focus on symmetric kernels. Suppose that X, Y is a pair of identically distributed (i.d.) r.v.'s with common cdf $F(x)$ and finite variance. We can consider that $\mathcal{F}_k$, $k = 1, 2$, is the Borel $\sigma-$field and $\mathcal{M}_k$ is the space of bounded regular Borel signed measures on $(\Omega_k, \mathcal{F}_k)$, which contains the space of probability measures, say, $\mathbb{P}$. We denote by $(\mathcal{X}, \mathcal{B}(\mathbb{R}))$ this measurable space, and by F the probability measure induced by the cdf F. Let $K(x, y)$ be a symmetric real valued function defined on $\mathcal{X} \times \mathcal{X} = [a, b] \times [a, b] \subset \mathbb{R}^2$, being $[a, b]$ the range of X.

If K is a (symmetric) Hilbert-Schmidt kernel, i.e.

$$\int_a^b \int_a^b K^2(s,t)\, dF(s)\, dF(t) < \infty,$$

there exists an eigenfunction expansion of K, convergent in mean square with respect to the product measure $dF(s)\, dF(t)$. Thus there exists a complete orthonormal basis $\{\xi_i\}$ over $L^2([a, b], F)$, such that

$$K(x, y) = \sum_{i=0}^{\infty} \lambda_i \xi_i(x)\, \xi_i(y), \tag{3.1}$$

where the countable sets of eigenvalues and eigenfunctions $\{\lambda_i\}, \{\xi_i\}$ of K on F satisfy

$$\int_a^b \xi_i(x)\, K(x, y)\, dF(x) = \lambda_i \xi_i(y), \tag{3.2}$$

with $\sum_{i=0}^{\infty} \lambda_i^2 < \infty$. Notice that $E[\xi_i(X)\, K(X, y)] = \lambda_i \xi_i(y)$, and $E[K^2(X, Y)] = \sum_{i=0}^{\infty} \lambda_i^2$. See, for instance, [35].

An orthonormal sequence of functions $\{\xi_i\}$ on F satisfies

$$E(\xi_k(X)\, \xi_l(X)) = \int_a^b \xi_k(x)\, \xi_l(x)\, dF(x) = \delta_{kl} \tag{3.3}$$

where δ_{kl} is Kronecker's delta. If each ξ_k, $k > 0$, has 0 mean, i.e.,

$$E(\xi_k(X)) = \int_a^b \xi_k(x)\, dF(x) = 0, \qquad \forall k > 0,$$

then condition (3.3) is equivalent to pairwise zero correlation and unit variance. Therefore $\rho_{ij} \equiv Cor(\xi_i(X), \xi_j(X))$ is $\rho_{ij} = Cov(\xi_i(X), \xi_j(X)) \equiv \sigma_{ij} = \delta_{ij}$, and we denote $\rho_{ii} = \rho_i$, $\sigma_{ii} = \sigma_i$.

An orthonormal sequence $\{\xi_i\}$ is complete if, given any function g, the r.v. $g(X)$ with finite variance has an expansion, convergent in the mean square sense

$$g(X) = \sum_{i=0}^{\infty} a_i \xi_i(X),$$

where $a_i = E\left(g\left(X\right)\xi_i\left(X\right)\right)$. The expansion is in the sense that

$$\lim_{n\to\infty} E\left(g\left(X\right) - S_n\right)^2 = 0,$$

where

$$S_n = \sum_{i=0}^{n} a_i \xi_i\left(X\right).$$

An equivalent condition uses

$$Var\left(g\left(X\right)\right) = \sum_{i=0}^{\infty} a_i^2.$$

If $\sum_{i=0}^{\infty} a_i^2 < \infty$, the series converges with probability one.

Let us consider the quadratic integral form J associated to K

$$J\left(\varphi,\varphi\right) = \int_a^b \int_a^b K\left(s,t\right)\varphi\left(s\right)\varphi\left(t\right) dsdt$$

where φ is any continuous or piecewise continuous function in the basic domain $[a,b]$. A quadratic integral form J is said to be positive semidefinite (p.s.d.) if

$$\forall\varphi \in \mathcal{C}\left([a,b]\right), \quad J\left(\varphi,\varphi\right) \geq 0$$

and positive definite (p.d.) if

$$\varphi \neq 0 \Longrightarrow J\left(\varphi,\varphi\right) > 0.$$

A symmetric kernel K is said to be p.(s.)d. whenever its associated quadratic integral form J is either p.s.d. or p.d. (see [11]).

Let

$$\mathcal{K}_{\mathcal{X}\times\mathcal{X}} = \left\{\text{continuous, symmetric p.s.d. kernels } K \text{ on } \mathcal{X}\times\mathcal{X}\right\}.$$

Mercer's theorem (see, for instance, [11]) states that for some sets $\{\lambda_i\}, \{\xi_i\}$ expansion (3.1) holds for such kernels, and the series converges absolutely and uniformly. A kernel K is p.s.d. iff all the non-zero eigenvalues λ_i are > 0. It is p.d. iff all eigenvalues are > 0 and the corresponding eigenfunctions $\{\xi_i\}$ form a complete orthonormal system (see [63]).

We introduce a relevant, specific class of kernels which arises in the context of the asymptotic study of U-statistics.

Definition 3.1.2 *A square integrable symmetric kernel K is said to be degenerate if the mathematical expectation exists and*

$$E_Y(K(x,Y)) = \int_a^b K(x,y)\,dF(y)$$

is a constant a.e. in x.

From the symmetry of these kernels we can also define degeneracy by

$$E_X(K(X,y)) = c$$

where c is a constant a.e. in y.

The following result characterizes degenerate kernels (see [35] for a proof).

Theorem 3.1.3 *The following statements are equivalent:*

1. *The square integrable kernel K is degenerate.*
2. *One eigenfunction of K, say, ξ_0, is a constant.*
3. $E_X(\xi_k(X)) = 0, \quad \forall k > 0.$

This theorem has some relevant consequences as we show below: a) the complete orthonormal set of eigenfunctions in (3.1) starts with a constant function ξ_0, for example the constant 1, and the corresponding eigenvalue is $\lambda_0 = 0$; b) If the kernel is symmetric and degenerate, then it is double-centred (see below), and every eigenfunction ξ_k, $k > 0$, has zero expectation.

Example 3.1.4 *Let $K(x,x') = xx'$ and X, Y independent i.d. as X with cdf F, $E(X) = 0$, $Var(X) = \sigma^2$ and $[a,b]$ is the range of X. This kernel is degenerate*

$$E_X(K(X,x')) = \int_a^b xx'dF(x) = x'E(X) = 0.$$

The eigenexpansion is

$$K(x,x') = \sigma^2 \frac{x}{\sigma}\frac{x'}{\sigma},$$

since the only non-null eigenvalue is $\lambda_1 = \sigma^2$ with eigenvector $\xi_1(x) = x/\sigma$. Note that $\xi_0(x) = 1$ is also eigenvector with eigenvalue 0.

Example 3.1.5 *Let $c_1(x,y), c_2(x,y)$ be two symmetric bivariate densities with uniform marginal densities on $[0,1]$. Let us consider the (symmetric) kernel $K = c_1 - c_2$. Then*

$$E_X(K(X,y)) = \int_0^1 (c_1(x,y) - c_2(x,y))dx = 1 - 1 = 0,$$

for almost all $y \in [0,1]$, so K is degenerate.

3.2 The product of symmetric kernels

Our aim is to define the *intersection* and *union* operations of continuous, symmetric kernels, generalizing the above similar operations with finite matrices.

Let us denote by $\star$ the integral operator Φ with symmetric kernel $K(s,t)$, $t \in \mathcal{X}$ (see [29])

$$\Phi(u(\cdot)) = K(\cdot,t) \star u(t) = \int_{\mathcal{X}} K(\cdot,t)\, u(t)\, dF(t).$$

Thus $\star$ stands for the integration with respect to the repeated variable t.

Let (X,Y) be a random vector defined on $\mathcal{X} \times \mathcal{Y}$ with bivariate cdf H and marginal cdf's F, G, respectively. Thus H belongs to the Fréchet class $\mathcal{F}(F,G)$. The product of kernels $K_1 \in \mathcal{K}_{\mathcal{X}\times\mathcal{X}}$, and $K_2 \in \mathcal{K}_{\mathcal{Y}\times\mathcal{Y}}$ is next defined. Let us preserve the above notation and write $\star$ for double integration with respect to two repeated variables.

Definition 3.2.1 *Let $H \in \mathcal{F}(F,G)$ be a bivariate cdf absolutely continuous with respect to the product measure FG. Let $K_1 \in \mathcal{K}_{\mathcal{X}\times\mathcal{X}}$, $K_2 \in \mathcal{K}_{\mathcal{Y}\times\mathcal{Y}}$. The $H-$product $K_1 \star K_2$ is a kernel defined on $\mathcal{X} \times \mathcal{Y}$ by*

$$(K_1 \star K_2)_H(x,y) = Cov(K_1(x,X), K_2(Y,y)), \tag{3.4}$$

for every pair $(x,y) \in \mathcal{X} \times \mathcal{Y}$, with $(X,Y) \sim H$.

Remark 3.2.2

1. *Covariance (3.4) is*

$$\int_{\mathcal{X}}\int_{\mathcal{Y}} K_1(x,s)\, K_2(t,y)\, dH(s,t) - \int_{\mathcal{X}} K_1(x,s)\, dF(s) \int_{\mathcal{Y}} K_2(t,y)\, dG(t) \tag{3.5}$$

and can be computed by

$$\int_{\mathcal{X}}\int_{\mathcal{Y}}(H(s,t)-F(s)G(t))dK_1(\cdot,s)\,dK_2(t,\cdot)\,,$$

provided $K_1(\cdot,s)$ *is of bounded variation on* $\mathcal{X}$ *and* $K_2(t,\cdot)$ *is of bounded variation on* $\mathcal{Y}$*. See [18], Theorem 1.*

2. *The convergence of the first integral is ensured by assuming* H *absolutely continuous with respect to* FG*.*

Theorem 3.2.3 *Let* $K_1\in\mathcal{K}_{\mathcal{X}\times\mathcal{X}}$*,* $K_2\in\mathcal{K}_{\mathcal{Y}\times\mathcal{Y}}$ *satisfying conditions of Definition 3.2.1, and consider their eigenexpansions*

$$K_1=\sum_{i=0}^{\infty}\lambda_i\xi_i\otimes\xi_i,\qquad K_2=\sum_{i=0}^{\infty}\mu_i\eta_i\otimes\eta_i, \tag{3.6}$$

where $\xi_i\otimes\xi_i$ *stands for* $\xi_i(x)\,\xi_i(x')$ *for every pair* $(x,x')\in\mathcal{X}\times\mathcal{X}$ *and similarly* $\eta_i\otimes\eta_i$*. For a given random vector* (X,Y) *with bivariate cdf* H*, defined on* $\mathcal{X}\times\mathcal{Y}$*, then*

$$K_1\star K_2=\sum_{i,j=0}^{\infty}\lambda_i\mu_j Cov\left(\xi_i(X),\eta_j(Y)\right)\xi_i\otimes\eta_j. \tag{3.7}$$

Proof. Recall the orthonormality of the sets $\{\xi_i\},\{\eta_i\}$. Then for each pair $i,j\geq 0$

$$\begin{aligned} Cov\left(\xi_i(X),\eta_j(Y)\right) &= E_{X,Y}\left[\xi_i(X)\,\eta_j(Y)\right]-E_X\left[\xi_i(X)\right]E_Y\left[\eta_j(Y)\right]\\ &= \int_{\mathcal{X}}\int_{\mathcal{Y}}\xi_i(s)\,\eta_j(t)\,dH(s,t)\\ &\quad -\int_{\mathcal{X}}\xi_i(s)\,dF(s)\int_{\mathcal{Y}}\eta_j(t)\,dG(t)\,. \end{aligned} \tag{3.8}$$

By Mercer's theorem, the expansions of kernels (3.6) converge absolutely and uniformly. By the dominated convergence theorem we can exchange integration and infinite sums, and by Fubini's theorem:

$$\begin{aligned} &\int_{\mathcal{X}}\int_{\mathcal{Y}}K_1(x,s)\,K_2(t,y)\,dH(s,t)=\\ &=\int_{\mathcal{X}}\int_{\mathcal{Y}}\left[\sum_{i=0}^{\infty}\lambda_i\xi_i(x)\,\xi_i(s)\sum_{j=0}^{\infty}\mu_j\eta_j(t)\,\eta_j(y)\right]dH(s,t)\\ &=\sum_{i,j=0}^{\infty}\lambda_i\mu_j\xi_i(x)\,\eta_j(y)\int_{\mathcal{X}}\int_{\mathcal{Y}}\xi_i(s)\,\eta_j(t)\,dH(s,t)\,, \end{aligned}$$

$$
\begin{aligned}
\int_{\mathcal{X}} K_1(x,s)\,dF(s) &= \int_{\mathcal{X}} \sum_{i=0}^{\infty} \lambda_i \xi_i(x)\,\xi_i(s)\,dF(s) \\
&= \sum_{i=0}^{\infty} \lambda_i \xi_i(x) \int_{\mathcal{X}} \xi_i(s)\,dF(s), \\
\int_{\mathcal{Y}} K_2(t,y)\,dG(t) &= \int_{\mathcal{Y}} \sum_{j=0}^{\infty} \mu_j \eta_j(t)\,\eta_j(y)\,dG(t) \\
&= \sum_{j=0}^{\infty} \mu_j \eta_j(y) \int_{\mathcal{Y}} \eta_j(t)\,dG(t).
\end{aligned}
$$

Using (3.4) and substituting these results in (3.5)

$$
\begin{aligned}
(K_1 \star K_2)_H(x,y) &= \sum_{i,j=0}^{\infty} \lambda_i \mu_j \xi_i(x)\,\eta_j(y) \int_{\mathcal{X}} \int_{\mathcal{Y}} \xi_i(s)\,\eta_j(t)\,dH(s,t) \\
&\quad - \sum_{i,j=0}^{\infty} \lambda_i \xi_i(x)\,\mu_j \eta_j(y) \int_{\mathcal{X}} \xi_i(s)\,dF(s) \int_{\mathcal{Y}} \eta_j(t)\,dG(t) \\
&= \sum_{i,j=0}^{\infty} \lambda_i \mu_j \xi_i(x)\,\eta_j(y) \,[\int_{\mathcal{X}} \int_{\mathcal{Y}} \xi_i(s)\,\eta_j(t)\,dH(s,t) + \\
&\quad - \int_{\mathcal{X}} \xi_i(s)\,dF(s) \int_{\mathcal{Y}} \eta_j(t)\,dG(t)].
\end{aligned}
$$

Expansion (3.7) follows by substituting (3.8) in the last expression.

Note that since $\{\xi_i(X)\}, \{\eta_i(Y)\}$ have finite variances, Schwartz inequality

$$
\sum_{i,j=0}^{\infty} \lambda_i^2 \mu_j^2 Cov^2(\xi_i(X), \eta_j(Y)) \leq \sum_{i,j=0}^{\infty} \lambda_i^2 \mu_j^2 Var(\xi_i(X))\, Var(\eta_j(Y))
$$

provides a proof of the convergence of (3.7) in the mean-square sense. □

Corollary 3.2.4 *Let (X,Y) be a random vector on $\mathcal{X} \times \mathcal{Y}$, with $H = FG$, i.e., X and Y are independent. Let $K_1 \in K_{\mathcal{X}\times\mathcal{X}}$ and $K_2 \in K_{\mathcal{Y}\times\mathcal{Y}}$. Then $K_1 \star K_2 = 0$.*

Proof. The covariance between two functions ξ_i, η_j of the r.v.'s X, Y can be computed by

$$
Cov(\xi_i(X), \eta_j(Y)) = \int_{\mathcal{X}} \int_{\mathcal{Y}} (H(s,t) - F(s)\,G(t))\,d\xi_i(s)\,d\eta_j(t),
$$

as mentioned above (see [18]). If $H = FG$ all covariances in (3.7) are 0. □

Remark 3.2.5 *As the ordinary product of matrices, (3.4) may not give a symmetric kernel. Furthermore, the associative property does not hold.*

Example 3.2.6 *Let $K_X(s,t) = st$ with $(s,t) \in \mathcal{X} \times \mathcal{X}$, and similarly, $K_Y(s,t) = st$ with $(s,t) \in \mathcal{Y} \times \mathcal{Y}$. Consider two r.v.'s X, Y with ranges $\mathcal{X}, \mathcal{Y}$ and bivariate cdf $H \in \digamma(F,G)$, such that $E(X) = E(Y) = 0$, $Var(X) = \sigma_1^2$, $Var(Y) = \sigma_2^2$. The only non-zero eigenvalue of K_X is $\lambda_1 = \sigma_1^2$, with eigenfunction $\xi_1(x) = x/\sigma_1$, orthogonal to $\xi_0(x) = 1$, with eigenvalue $\lambda_0 = 0$; similarly for K_Y, $\mu_1 = \sigma_2^2$ is the non-zero eigenvalue, with eigenfunction $\eta_1(y) = y/\sigma_2$. Both kernels are degenerate for both variables and the $H-$product is*

$$\begin{aligned}(K_X \star K_Y)_H(x,y) &= \sigma_1^2\sigma_2^2\left(\frac{x}{\sigma_1}\cdot\frac{y}{\sigma_2}\right)Cov\left(\frac{X}{\sigma_1},\frac{Y}{\sigma_2}\right)\\ &= Cov(X,Y)\,xy,\\ &= E_{(X,Y)}[XY]\,xy.\end{aligned}$$

Notice that the product kernel $K_X \star K_Y$ does not depend on σ_1^2, σ_2^2. It only depends on the mathematical expectation of the variable $X \cdot Y$ (i.e., on its covariance). If X, Y are stochastically independent r.v.'s, $K_X \star K_Y = 0$ for every pair $(x,y) \in \mathcal{X} \times \mathcal{Y}$. $\sigma_1\sigma_2$ is an upper bound for the covariance; it follows from Cauchy-Schwartz inequality. Finally, the eigenexpansion of $K_X \star K_X$ is

$$(K_X \star K_X)_H(s,t) = \sigma_1^4\frac{s}{\sigma_1}\cdot\frac{t}{\sigma_1}.$$

$K_X \star K_X$ and K_X share their eigenfunctions but the eigenvalues of $K_X \star K_X$ are the squared of the eigenvalues of K_X.

The $H-$product of kernels has some properties.

Proposition 3.2.7 *Let $K_1 \in \mathcal{K}_{\mathcal{X}\times\mathcal{X}}$, $K_2 \in \mathcal{K}_{\mathcal{Y}\times\mathcal{Y}}$. For every $(x,y) \in \mathcal{X} \times \mathcal{Y}$, $K_1 \star K_2$ satisfies:*

$$(K_1 \star K_2)_H(x,y) = (K_2 \star K_1)_{H'}(y,x)$$

where H' is the bivariate cdf of the random vector (Y,X) such that $H'(y,x) = H(x,y)$ for every $x \in \mathcal{X}$ and for every $y \in \mathcal{Y}$.

Proof. The proof follows as an immediate consequence of the symmetry of the covariance and K_1, K_2. □

Proposition 3.2.8 *Let $K_1, K_2 \in \mathcal{K}_{\mathcal{X}\times\mathcal{X}}$. If $\alpha \geq 0$ is an eigenvalue with eigenfunction ξ of the $H-$product $K_1 \star K_2$ with respect to F, i.e.,*

$$\int_{\mathcal{X}} \xi(s)(K_1 \star K_2)_H(s,t)\,dF(s) = \alpha\xi(t), \tag{3.9}$$

where H is symmetric and $H \in \digamma(F,F)$, then α is also an eigenvalue of $K_2 \star K_1$ with respect to F.

PROOF. Proposition 3.2.7 allows us to substitute $(K_2 \star K_1)(t,s)$ for $(K_1 \star K_2)(s,t)$ in (3.9). □

We are now interested in the product of kernels defined on the same product space, i.e., $\mathcal{X} = \mathcal{Y}$. If, in addition, $X = Y$ almost surely $(a.s)$. some relevant simplifications hold. In this case, H attains H^+, the Fréchet upper bound (see Section 1.2 in Chapter 1). Then (3.4) can be obtained by integration with respect to F, as integrating $dH^+(s,t)$ on $s = t$ is the same as integrating $dF(s)$ on s. Thus (see [18])

$$\int_{\mathbb{R}^2} \alpha(s)\beta(t)dH^+(s,t) = \int_{\mathbb{R}} \alpha(s)\beta(s)dF(s). \tag{3.10}$$

It suffices to consider

$$K_1(\cdot,s) \equiv \alpha(s), \quad K_2(t,\cdot) \equiv \beta(t)$$

and apply (3.10). Then (3.4), (3.7) reduce to

$$(K_1 \star K_2)_H(x,y) = \int_{\mathcal{X}} K_1(x,s) K_2(s,y) dF(s) - E_X(K_1(x,X)) E_X(K_2(X,y)),$$

and

$$K_1 \star K_2 = \sum_{i,j=0}^{\infty} \lambda_i \mu_j Cov_{H^+}(\xi_i, \eta_j) \xi_i \otimes \eta_j, \tag{3.11}$$

where $Cov_{H^+}(\xi_i,\eta_j)$ stands for $Cov(\xi_i(X), \eta_j(X))$. Expansion (3.11) is derived directly from (3.7).

Recall that intersection and union operations involve square roots of matrices. In the context of metric multidimensional scaling, in finding principal coordinates from an $n \times n$ distance matrix $\mathbf{\Delta} = (\delta_{ij})$, the inner product matrix related to $\mathbf{\Delta}$ must be double-centred. Double-centred kernels are next defined.

Definition 3.2.9 *A kernel $G \in \mathcal{K}_{\mathcal{X}\times\mathcal{X}}$ is double-centred with respect to a cdf F if for every $x, x' \in \mathcal{X}$:*

$$\begin{aligned} E_X(G(X,x')) &= \int_{\mathcal{X}} G(x,x') dF(x) = 0, \\ E_{X'}(G(x,X')) &= \int_{\mathcal{X}} G(x,x') dF(x') = 0. \end{aligned}$$

The continuous version of the centring matrix **H** (see (2.10) in Chapter 2) is the double-centring of K (see [29])

$$G_K(x,x') = K(x,x') - E_X(K(X,x')) - E_{X'}(K(x,X')) + E_X[E_{X'}(K(X,X'))].$$

It is straightforward to see that $G_K \in \mathcal{K}_{\mathcal{X}\times\mathcal{X}}$, the double-centring of K, satisfies

$$E_X(G_K(X,x')) = E_{X'}(G_K(x,X')) = 0, \quad \forall\, x,x' \in \mathcal{X}.$$

Thus a double-centred kernel is always degenerate.

Proposition 3.2.10 *If a non-constant kernel $K \in \mathcal{K}_{\mathcal{X}\times\mathcal{X}}$ is degenerate then it is double-centred.*

PROOF. Theorem 3.1.3 establishes that a degenerate kernel K has one constant (non-null) eigenfunction, ξ_0, and $E_X[\xi_k(X)] = 0$, for every $k > 0$. Then

$$\begin{aligned} E_X[K(X,x')] &= E_X\left[\sum_{k=0}^{\infty} \lambda_k \xi_k(X)\,\xi_k(x')\right] \\ &= \lambda_0 \xi_0^2. \end{aligned}$$

On the other hand, ξ_0 is (a non-null, constant) eigenfunction of K with eigenvalue λ_0 iff

$$\int_{\mathcal{X}} \xi_0 K(x,x')\,dF(x) = \lambda_0 \xi_0$$

or, equivalently, $E_X[K(X,x')] = \lambda_0$. The only possible values of λ_0 are 0 or 1. If $\lambda_0 = 1$ then $K(x,y) = \xi_0$ (i.e., is a constant) for almost every $(x,y) \in \mathcal{X}\times\mathcal{X}$. Hence, if K is not a constant, it is double-centred. □

In the following, we only consider double-centred kernels on $\mathcal{K}_{\mathcal{X}\times\mathcal{X}}$ unless the contrary is stated.

Definition 3.2.11 *Let $K \in \mathcal{K}_{\mathcal{X}\times\mathcal{X}}$ be double-centred, such that $K = \sum_{i=1}^{\infty} \lambda_i \xi_i \otimes \xi_i$ for every pair of values $(x,x') \in \mathcal{X}\times\mathcal{X}$. The square root of K, say, $K^{\frac{1}{2}}$, is defined as the kernel of $\mathcal{K}_{\mathcal{X}\times\mathcal{X}}$*

$$K^{\frac{1}{2}} = \sum_{i=1}^{\infty} \lambda_i^{\frac{1}{2}} \xi_i \otimes \xi_i,$$

provided $\sum_{i=1}^{\infty} \lambda_i < \infty$. Any rational power $K^{\frac{m}{n}}$, $m < n$ is similarly defined.

Theorem 3.2.12 *$K^{\frac{1}{2}}$ is well defined as the square root of K, in the sense that the $H-$product satisfies $K^{\frac{1}{2}} \star K^{\frac{1}{2}} = K$, if, for all $i, j > 0$, $Cov\left(\xi_i\left(X\right), \xi_j\left(Y\right)\right) = \delta_{ij}$. In particular, $K^{\frac{1}{2}}$ is well defined if K is symmetric and $H = H^+$.*

PROOF. From (3.7)

$$\left(K^{\frac{1}{2}} \star K^{\frac{1}{2}}\right)_H (x, y) = \sum_{i,j=1}^{\infty} \lambda_i^{\frac{1}{2}} \lambda_j^{\frac{1}{2}} Cov\left(\xi_i\left(X\right), \xi_j\left(Y\right)\right) \xi_i\left(x\right) \xi_j\left(y\right),$$

assuming that, if $\lambda_0 = 0$ is an eigenvalue of K, this series starts at $i, j = 0$. Thus, the product $K^{\frac{1}{2}} \star K^{\frac{1}{2}} = K$ if $Cov\left(\xi_i\left(X\right), \xi_j\left(Y\right)\right) = \delta_{ij}$, for every $i, j > 0$.

If $H = H^+$, we have

$$\int_{\mathcal{X}} \int_{\mathcal{X}} \xi_i\left(s\right) \xi_j\left(t\right) dH^+\left(s, t\right) = \int_{\mathcal{X}} \xi_i\left(s\right) \xi_j\left(s\right) dF\left(s\right),$$

where F is the cdf of X, and the last integral is δ_{ij} for every $i, j > 0$ (see condition (3.3)). □

3.3 Intersection and union operations

3.3.1 Definitions

Let $K_1, K_2 \in \mathcal{K}_{\mathcal{X}\times\mathcal{X}}$. The definition of the $H-$product and the square root of kernels allow us to define the union and intersection operations of kernels.

Definition 3.3.1 *Let $H \in \digamma\left(F, F\right)$ be a symmetric bivariate cdf, each marginal variable with range $\mathcal{X}$. The intersection $K_1 \wedge K_2$ and the union $K_1 \vee K_2$ of $K_1, K_2 \in \mathcal{K}_{\mathcal{X}\times\mathcal{X}}$ are defined by*

$$K_1 \wedge K_2 = \frac{1}{2}\left(K_1^{\frac{1}{2}} \star K_2^{\frac{1}{2}} + K_2^{\frac{1}{2}} \star K_1^{\frac{1}{2}}\right), \quad (3.12)$$

$$K_1 \vee K_2 = K_1 + K_2 - K_1 \wedge K_2, \quad (3.13)$$

where $\star$ stands for the above $H-$product.

It is obvious that $K_1 \wedge K_2$, $K_1 \vee K_2$ are also symmetric kernels on $\mathcal{X} \times \mathcal{X}$.

Proposition 3.3.2 *Let $K_1, K_2 \in \mathcal{K}_{\mathcal{X}\times\mathcal{X}}$. The union $K_1 \vee K_2$ is a positive semidefinite kernel. However, $K_1 \wedge K_2$ is not positive semidefinite in general.*

PROOF. Consider the following well-known results on eigenvalues of kernels:

1. Let α be any real number, and let λ, φ be an eigenvalue and its corresponding eigenfunction of a kernel $K \in \mathcal{K}_{\mathcal{X}\times\mathcal{X}}$. Then λ is also an eigenvalue of αK, with eigenfunction $\varphi' = \alpha\varphi$.

2. Each positive eigenvalue of the sum $K + K^{+}$, where K^{+} is a positive definite kernel, is not lower than the corresponding eigenvalue of the kernel K (see, [11]).

Write $K_1 \vee K_2$ as

$$K_1 \vee K_2 = \frac{1}{2}\left(K_1 + K_2 + \left(K_1^{\frac{1}{2}} - K_2^{\frac{1}{2}}\right)^2\right),$$

where $\left(K_1^{\frac{1}{2}} - K_2^{\frac{1}{2}}\right)^2$ stands for $K_1 + K_2 - K_1^{\frac{1}{2}} \star K_2^{\frac{1}{2}} - K_2^{\frac{1}{2}} \star K_1^{\frac{1}{2}}$, which is positive. Thus $K_1 \vee K_2$ is also positive as a consequence of 2.
Next, let us consider the kernel $K \in \mathcal{K}_{\mathcal{X}\times\mathcal{X}}$

$$K(s,t) = \sigma^2 \frac{s}{\sigma} \cdot \frac{t}{\sigma},$$

and the r.v.'s X, Y, with range $\mathcal{X}$, identically distributed, with $E(X) = E(Y) = 0$, $Var(X) = Var(Y) = \sigma^2$ and $Cov(X,Y) = \sigma_{12} < 0$ for a fixed H. If K_X, K_Y denote the kernel K when defined on the random vectors (X,X), (Y,Y), respectively, both kernels are positive definite, double-centred, with eigenvalue σ^2 and corresponding eigenfunction $\xi(s) = \frac{s}{\sigma}$. Direct computations show that

$$(K_X \wedge K_Y)_H (x,y) = \sigma_{12} xy.$$

Thus $K_X \wedge K_Y$ is negative definite because the covariance is negative and the only non-null eigenvalue is $\sigma_{12}\sigma^2 < 0$. □

Remark 3.3.3 *K_1 and K_2, must be defined on the same spaces $\mathcal{X} = \mathcal{Y}$, in coherence with definitions (3.12) and (3.13).*

Example 3.3.4 *Consider the kernel $K = \sum_{i=0}^{\infty} \lambda_i \xi_i \otimes \xi_i$ with $X = Y$ (a.s.). Then the intersection of K and $K^{\frac{1}{2}}$ with respect to H^{+} is*

$$K \wedge K^{\frac{1}{2}} = \sum_{i=0}^{\infty} \lambda_i^{\frac{3}{4}} \xi_i \otimes \xi_i.$$

3.3.2 Basic properties

Let (X, Y) be measurable functions on the product space $\mathcal{X} \times \mathcal{X}$ with joint cdf H and common marginal cdf's F. Let us consider the expansions (3.6) of $K_1 \in K_{\mathcal{X}\times\mathcal{X}}$ and $K_2 \in K_{\mathcal{X}\times\mathcal{X}}$. Note that the symmetric property of the operations allows us to interchange Y and X when necessary. The following proposition summarizes the main properties of the intersection and union operations.

Proposition 3.3.5 *Let K_1, K_2 be symmetric p.s.d. kernels as defined above and let K_0 be the null kernel, i.e., $K_0(x, y) = 0$ for almost every $(x, y) \in \mathcal{X} \times \mathcal{X}$. The intersection and union operations satisfy the following properties:*

1. *Commutativity: for any $H \in \digamma(F, F)$*

$$K_1 \wedge K_2 = K_2 \wedge K_1, \quad K_1 \vee K_2 = K_2 \vee K_1.$$

2. *Zero element: for any $H \in \digamma(F, F)$, K_0 is the neutral element of the union of kernels,*

$$K_1 \wedge K_0 = K_0, \quad K_1 \vee K_0 = K_1.$$

3. *Equality: if H attains H^+ and $K_1 = K_2$, i.e., for almost every $(x, y) \in \mathcal{X} \times \mathcal{X}$, $K_1(x, y) = K_2(x, y)$, then*

$$K_1 \wedge K_2 = K_1 = K_2, \quad K_1 \vee K_2 = K_1 = K_2.$$

4. *Orthogonality: if X, Y are stochastically independent ($H(x, y) = F(x) F(y)$ for almost every $(x, y) \in \mathcal{X} \times \mathcal{X}$), then*

$$K_1 \wedge K_2 = K_0, \quad K_1 \vee K_2 = K_1 + K_2.$$

Proof. Commutativity is directly derived from the commutativity of the sum of measurable functions on the product space.The proof of 2 is also straightforward. That H attains H^+ is a necessary condition for $K^{\frac{1}{2}} \star K^{\frac{1}{2}} = K$ (see Theorem 3.2.12); then 3 is evident. Statement 4 follows from Corollary 3.2.4. □

Remark 3.3.6 *A neutral element for the intersection does not exist. In other words, there is no kernel K_E, such that*

$$K \wedge K_E = K \tag{3.14}$$

for every kernel K defined on (X, Y). We may consider a suitable class of kernels, for example those which share the same set of eigenfunctions. Let $K_E = \sum_{i=0}^{\infty} \nu_i \xi_i \otimes \xi_i$ denote the possible neutral kernel and let K_I be the kernel with eigenfunctions $\{\xi_i(x)\}$ and whose eigenvalues are all 1. Observe that

$$K_I = \sum_{i=0}^{\infty} 1 \xi_i \otimes \xi_i \tag{3.15}$$

is not convergent. Let us consider, finally, the subclass of degenerate symmetric p.d. kernels (a finite number of eigenvalues higher than 0) with eigenfunctions $\{\xi_i(x)\}_{i=0,\ldots N}$. If we denote $K_{IN} = S_{N+1}$, the partial sum of the $N+1$ first terms of (3.15), the intersection of K_E and K_{IN} is

$$(K_{IN} \wedge K_E)(x, y) = \sum_{i=0}^{N} \nu_i^{\frac{1}{2}} \xi_i(x) \xi_i(y)$$

which, by hypothesis (3.14), must be equal to K_{IN}. Then $\nu_i^{\frac{1}{2}} = 1$ and $K_{IN} = K_E$ is the possible neutral element. However, for any kernel $K \in \mathcal{K}_{\mathcal{X}\times\mathcal{X}}$

$$(K_{IN} \wedge K)(x, y) = \sum_{i=0}^{N} \lambda_i^{\frac{1}{2}} \xi_i(x) \xi_i(y)$$

with $\{\lambda_i\}$ being the (possibly infinite) set of eigenvalues of K. Thus $K_{IN} \wedge K \neq K$, in general. These results are similar to those obtained in Chapter 2 for the case of finite dimensional matrices but using the identity matrix.

Remark 3.3.7 *The intersection and union operations of symmetric kernels are not associative in general.*

Example 3.3.8 *Let us consider Example 3.2.6 with $\sigma_1^2 = \sigma_2^2 = 1$. Then $K(x, y) = xy$, $(K \star K)_H(x, y) = \rho xy$, where $\rho = Cor(X, Y)$ with respect to a given bivariate cdf H and*

$$(K \wedge K)_H(x, y) = \rho xy, \quad (K \vee K)_H(x, y) = 2xy - \rho xy.$$

All properties can be verified in this example. Thus, if $\rho = 0$ then $K \wedge K = K_0$, $K \vee K = 2K$, and if $\rho = 1$ (i.e., there is a linear relation between X and Y, as when $X = Y$ a.s.) then $K \wedge K = K \vee K = K$.

3.3.3 Dimension of kernels

The cardinal of the set of non-null eigenvalues of a kernel K determines its dimension, denoted by $\dim(K)$. First, let us introduce some relevant concepts and results, proved in [11].

Definition 3.3.9 *A kernel $K(s,t)$ which can be written as a finite sum of products of functions of s and functions of t*

$$K(s,t) = \sum_{i=0}^{n} \varsigma_i(s)\,\theta_i(t) \tag{3.16}$$

is called a finite dimensional kernel.

Theorem 3.3.10 *The following properties hold:*

1. *Every continuous symmetric non-null kernel possesses eigenvalues and eigenfunctions with cardinality $\varkappa_0$ iff the kernel cannot be written as (3.16).*
2. *If a kernel K has only a finite number of eigenvalues $\lambda_1, \lambda_2, ..., \lambda_n$ it must be finite dimensional and can be represented in the form*

   $$K(s,t) = \sum_{i=1}^{n} \lambda_i \xi_i(s)\,\xi_i(t)\,.$$

   Conversely, a finite dimensional kernel has only a finite number of eigenvalues.
3. *All eigenvalues of a real symmetric kernel are real. The number of non-null eigenvalues of a kernel is the dimension of the kernel. If a kernel is positive definite (all its eigenvalues are strictly positive) its dimension is infinite. If K is positive semidefinite, its dimension, say, N, is finite.*

Recall Example 3.1.4, with one positive eigenvalue. Observe that the null kernel is the only kernel with all its eigenvalues equal to 0; thus, its dimension is 0.

Now our interest is to study the dimension of the kernels resulting from the intersection and union operations of two kernels. The product $K_1 \star K_2$ defined on the range of two variables X and Y can be expanded on a double sum of functions of these two variables. But the coefficients of this sum do not coincide with the eigenvalues of the product. Furthermore, their computation is not easy, in general. In the following sections, we define a suitable class of kernels and obtain the eigenvalues of the product as well as the intersection and union, provided they are defined. In this case, their dimensions will be studied more extensively. Nevertheless, the following general results hold.

Proposition 3.3.11 *Let $K \in \mathcal{K}_{\mathcal{X}\times\mathcal{X}}$ and let $K^{\frac{1}{2}}$ be its square root. Then*

$$\dim(K) = \dim\left(K^{\frac{1}{2}}\right).$$

PROOF. The set of eigenvalues of the square root of a kernel is the set of (non-negative) square roots of its eigenvalues. □

Proposition 3.3.12 *Let K_1, $K_2 \in \mathcal{K}_{\mathcal{X}\times\mathcal{X}}$, not necessarily double-centred. The dimension of the $H-$product, where H is symmetric, satisfies*

$$\dim(K_1 \star K_2) = \dim(K_2 \star K_1).$$

PROOF. Proposition 3.2.8 ensures that the set of eigenvalues of $K_1 \star K_2$ and $K_2 \star K_1$ coincide. □

Proposition 3.3.13 *Let K_1, $K_2 \in \mathcal{K}_{\mathcal{X}\times\mathcal{X}}$. The dimension of the intersection satisfies*

$$\dim(K_1 \wedge K_2) = \dim\left(K_1^{\frac{1}{2}} \star K_2^{\frac{1}{2}}\right).$$

PROOF. Let α, ξ be eigenvalue, eigenvector of $K_1^{\frac{1}{2}} \star K_2^{\frac{1}{2}}$. Then

$$\int_{\mathcal{X}} \xi(s)\left(K_1^{\frac{1}{2}} \star K_2^{\frac{1}{2}}\right)(s,t)\,dF(s) = \alpha\xi(t).$$

Proposition 3.2.8 applied to the intersection gives

$$\begin{aligned}
\int_{\mathcal{X}} \xi(s)(K_1 \wedge K_2)(s,t)\,dF(s) &= \frac{1}{2}\int_{\mathcal{X}} \xi(s)\left(K_1^{\frac{1}{2}} \star K_2^{\frac{1}{2}}\right)(s,t)\,dF(s) + \\
&\quad + \frac{1}{2}\int_{\mathcal{X}} \xi(s)\left(K_2^{\frac{1}{2}} \star K_1^{\frac{1}{2}}\right)(s,t)\,dF(s) \\
&= \frac{1}{2}(\alpha\xi(t) + \alpha\xi(t)) \\
&= \alpha\xi(t).
\end{aligned}$$

Thus the intersection and the products $K_1^{\frac{1}{2}} \star K_2^{\frac{1}{2}}$ and $K_2^{\frac{1}{2}} \star K_1^{\frac{1}{2}}$ have the same eigenvalues. □

Two examples of symmetric kernels (two copulas) covering finite and infinite dimensions are next given.

Example 3.3.14 *Let $C(u,v) = uv(1+\theta(1-u)(1-v))$ be the FGM copula, and $C^{+}(u,v) = \min\{u,v\}$ the Fréchet upper bound. Then $C(u,v) - uv$ has dimension 1 whereas $C^{+}(u,v) - uv$ has dimension $\varkappa_0$. Let K and K^{+} denote the kernels*

$$\begin{aligned}
K(u,v) &= C(u,v) - uv \\
&= \theta u(1-u)v(1-v)
\end{aligned}$$

and

$$\begin{aligned} K^{+}(u,v) &= C^{+}(u,v) - uv \\ &= u(1-v)\,\chi_{[u<v]}(u,v) + v(1-u)\,\chi_{[u\geq v]}(u,v) \end{aligned}$$

where $\chi_{[u<v]}$ *stands for the indicator function of* $[u<v]$*. Note that* K *satisfies condition* (3.16)*, but* K^{+} *does not. Thus* K *is finite dimensional while the set of eigenfunctions of* K^{+} *is* $\left\{\sqrt{2}\sin(n\pi t)\right\}$, $n>0$*, the set of eigenvalues is* $\left\{\frac{1}{(n\pi)^2}\right\}$ *(see [99]) and has cardinality* $\varkappa_0$*. Moreover,* $\lambda_1 = \frac{\theta}{30}$, $\xi_1(s) = \sqrt{30}s(1-s)$ *satisfy condition* (3.2)*, and* $K(u,v) = \lambda_1\xi_1(u)\xi_1(v)$ *agrees with Theorem 3.3.10.*

The kernels in Example 3.3.14 are covariance kernels. In Chapter 5, a definition of the dimension of a bivariate distribution is given. The dimension of the covariance kernel is used to study this new definition.

3.3.4 Trace

Definition 3.3.15 *The trace of a linear operator* $\mathcal{K}$ *of kernel* K*, measurable with respect to a suitable measure* μ *is defined by*

$$tr\,(K_\mu) = \int_{\mathcal{X}} K(s,s)\,d\mu(s)\,.$$

We denote $tr\,(K_\mu) \equiv tr\,(K)$.

Proposition 3.3.16 *The trace of a symmetric kernel* K *with respect to* F *is the sum of the eigenvalues of the kernel with respect to the measure* F,

$$tr\,(K) = \sum_{i=0}^{\infty} \lambda_i,$$

where K *can be expanded as*

$$K = \sum_{i=0}^{\infty} \lambda_i \xi_i \otimes \xi_i,$$

for a suitable complete set of eigenfunctions $\{\xi_i\}$ *orthonormal with respect to* F.

PROOF. This is a consequence of the dominated convergence theorem applied to Definition 3.3.15 and the orthonormality (see (3.3)):

$$\begin{aligned} tr\left(K\right) &= \int_{\mathcal{X}} \sum_{i=0}^{\infty} \lambda_i \xi_i\left(s\right) \xi_i\left(s\right) dF\left(s\right) \\ &= \sum_{i=0}^{\infty} \lambda_i \int_{\mathcal{X}} \xi_i\left(s\right) \xi_i\left(s\right) dF\left(s\right) \\ &= \sum_{i=0}^{\infty} \lambda_i. \end{aligned}$$

□

Remark 3.3.17 *A generalized version of this result was given by Cuadras and Fortiana [26]. Let G_π be a kernel defined by means of a suitable distance δ and a measure π. The trace of the positive semidefinite linear operator $\mathfrak{G}_\pi$ on the space of square measurable functions $L^2\left(\mathcal{X},\pi\right)$ defined by the integral kernel G_π,*

$$\left(\mathfrak{G}_\pi \cdot \varphi\right)(x) = \int_{\mathcal{X}} G_\pi\left(x,x'\right)\varphi\left(x'\right) d\pi\left(x'\right), \qquad \varphi \in L^2\left(\mathcal{X},\pi\right),$$

is the sum of the eigenvalues of the kernel G with respect to the measure π. In this context this sum coincides with the so-called geometric variability of π with respect to δ.

3.4 Binary relations between kernels

3.4.1 An equivalence class

Suppose that the kernels $K_1, K_2 \in \mathcal{K}_{\mathcal{X}\times\mathcal{X}}$ share the countable set of orthonormal functions $\{\xi_i\}$ for some sets of eigenvalues $\{\lambda_i\}, \{\mu_j\}$. Assume that the subsets $\left\{\lambda_{(i)}\right\}_{i\in I}, \left\{\mu_{(j)}\right\}_{j\in J}$ are the sets of non-null eigenvalues with $I \subseteq J$, and $N_1 \leq N_2 \leq \infty$ are the respective dimensions, in the sense that $I = \{1,2,...,N_1\}$, $J = \{1,2,...,N_2\}$ and

$$K_1 = \sum_{i\in I} \lambda_{(i)}\xi_{(i)} \otimes \xi_{(i)}, \quad K_2 = \sum_{j\in J} \mu_{(j)}\xi_{(j)} \otimes \xi_{(j)}.$$

Notice that $\lambda_0 = \mu_0 = 0$ may not be a proper eigenvalue of these kernels. Then we say that K_1, K_2 are equivalent and write $K_1 \sim K_2$.

Of course, if $N_1 < N_2$ we can write the first expansion as

$$K_1 = \sum_{i=0}^{N_2} \lambda_i \xi_i \otimes \xi_i,$$

where $\lambda_{N_1+1} = \lambda_{N_1+2} = ... = \lambda_{N_2} = 0$.

The equivalence class of symmetric positive definite kernels on $\mathcal{X} \times \mathcal{X}$ sharing the eigenfunctions $\{\xi_i\}$, $i = 0, ..., N$, with $N \leq \infty$, is denoted by

$$[\xi_N] = \left\{ K \in \mathcal{K}_{\mathcal{X}\times\mathcal{X}} \mid K = \sum_{i=0}^{N} \gamma_i \xi_i \otimes \xi_i \right\}.$$

Thus $K_1 \sim K_2$ iff $K_1, K_2 \in [\xi_N]$.

3.4.2 Partial order relation

Here we suppose that λ_i, μ_i are related to the same eigenfunction for each i, i.e., we assume that the eigenfunctions ξ_i are conveniently arranged.

Definition 3.4.1 *Let us consider $K_1 \sim K_2$, such that* $\dim(K_1) = N_1 \leq \dim(K_2) = N_2$. *$K_1$ is said to be smaller than K_2, and we denote this partial order by $K_1 \lesssim K_2$, if*

$$\lambda_i \leq \mu_i, \qquad i = 0, 1, ..., N_1.$$

Notice that $\lambda_i \leq \mu_i$, $i = N_1 + 1, ..., N_2$ follows immediately from $N_1 \leq N_2$.

Example 3.4.2 *Extending Example 2.4.7 in Chapter 2, let ξ_1, ξ_2, ξ_3 be orthogonal functions on the (common) range of the random variables X, Y and define the kernel K with expansion*

$$K = \sum_{i=1}^{3} \lambda_i \xi_i \otimes \xi_i,$$

where $\lambda_1 > \lambda_2 > \lambda_3 > 0$. Define

$$K_1 = \lambda_1 \xi_1 \otimes \xi_1, \quad K_2 = \lambda_2 \xi_2 \otimes \xi_2, \quad K_3 = \lambda_1 \xi_1 \otimes \xi_1 + \lambda_3 \xi_3 \otimes \xi_3.$$

These three kernels share the same set of eigenfunctions, ξ_1, ξ_2, ξ_3. K_1 is one dimensional with eigenvalues $\{\lambda_1, 0, 0\}$. K_2 is one dimensional with eigenvalues $\{0, \lambda_2, 0\}$. K_3 has dimension two with eigenvalues $\{\lambda_1, 0, \lambda_3\}$. It is easy to check that

$$K_1 \lesssim K_3 \lesssim K, \qquad K_2 \lesssim K.$$

No other comparisons hold.

3.4.3 Properties of equivalent kernels

Intersection and union

Proposition 3.4.3 *Let* $K \in [\xi_N]$ *with eigenvalues* $\{\lambda_i\}$. *If* $\sum_{i=0}^{\infty} \lambda_i < \infty$, *then* $K^{\frac{1}{2}} \in [\xi_N]$.

PROOF. Hypothesis $\sum_{i=0}^{\infty} \lambda_i < \infty$ ensures that the series $\sum_{i=0}^{\infty} \lambda_i^{\frac{1}{2}} \xi_i \otimes \xi_i$ converges (see Definition 3.2.11). □

Proposition 3.4.4 *The* H^+*−product of two double-centred kernels with the same set of eigenfunctions is commutative.*

PROOF. The $H-$product is commutative if the kernels

$$(K_1 \star K_2)(x,y) = \sum_{i,j=0}^{\infty} \lambda_i \mu_j Cov(\xi_i(X), \xi_j(Y)) \xi_i(x) \xi_j(y),$$

and

$$(K_2 \star K_1)(x,y) = \sum_{i,j=0}^{\infty} \mu_j \lambda_i Cov(\xi_j(X), \xi_i(Y)) \xi_j(x) \xi_i(y),$$

are equal for almost every pair (x,y). If H is a symmetric and PQD (positive quadrant dependent, see Chapter 4) then $Cov(\xi_i(X), \xi_j(Y)) = Cov(\xi_j(X), \xi_i(Y))$. But condition $\xi_i(x)\xi_j(y) = \xi_j(x)\xi_i(y)$ for all i,j does not hold, in general.
The set of eigenfunctions $\{\xi_i\}$ is orthogonal with respect to the marginal distribution F. Thus $Cov(\xi_i(X), \xi_j(Y)) = \delta_{ij}$, if $H = H^+$ (see Theorem 3.2.12). In this particular case

$$\begin{aligned}
(K_1 \star K_2)(x,y) &= \sum_{i,j=0}^{\infty} \lambda_i \mu_j Cov(\xi_i(X), \xi_j(Y)) \xi_i(x) \xi_j(y) \\
&= \sum_{i=0}^{\infty} \lambda_i \mu_i Cov(\xi_i(X), \xi_i(Y)) \xi_i(x) \xi_i(y) \\
&= \sum_{i=0}^{\infty} \lambda_i \mu_i \xi_i(x) \xi_i(y) \\
&= (K_2 \star K_1)(x,y).
\end{aligned}$$

□

Proposition 3.4.5 *Let* $K_1, K_2 \in [\xi_N]$ *with eigenvalues* $\{\lambda_i\}, \{\mu_i\}$, *respectively, such that* $\sum\limits_{i=0}^{\infty} \lambda_i < \infty$, $\sum\limits_{i=0}^{\infty} \mu_i < \infty$. *Then*

$$K_1 \wedge K_2 \in [\xi_N] \quad \textit{with eigenvalues } \left\{\lambda_i^{\frac{1}{2}} \mu_i^{\frac{1}{2}}\right\},$$

and

$$K_1 \vee K_2 \in [\xi_N] \quad \textit{with eigenvalues } \left\{\lambda_i + \mu_i - \lambda_i^{\frac{1}{2}} \mu_i^{\frac{1}{2}}\right\}.$$

PROOF. From Proposition 3.4.4

$$\begin{aligned} K_1 \wedge K_2 &= K_1^{\frac{1}{2}} \star K_2^{\frac{1}{2}} \\ &= \sum_{i=0}^{\infty} \lambda_i^{\frac{1}{2}} \mu_i^{\frac{1}{2}} \xi_i \otimes \xi_i. \end{aligned} \tag{3.17}$$

Then

$$\begin{aligned} K_1 \vee K_2 &= \sum_{i=0}^{\infty} \lambda_i \xi_i \otimes \xi_i + \sum_{i=0}^{\infty} \mu_i \xi_i \otimes \xi_i - \sum_{i=0}^{\infty} \lambda_i^{\frac{1}{2}} \mu_i^{\frac{1}{2}} \xi_i \otimes \xi_i \\ &= \sum_{i=0}^{\infty} \left(\lambda_i + \mu_i - \lambda_i^{\frac{1}{2}} \mu_i^{\frac{1}{2}}\right) \xi_i \otimes \xi_i. \end{aligned} \tag{3.18}$$

The convergence in the mean square sense of the series (3.17) and (3.18) follows from the convergence of every set of eigenvalues. □

Remark 3.4.6 *The sets* $\left\{\lambda_i^{\frac{1}{2}} \mu_i^{\frac{1}{2}}\right\}$, $\left\{\lambda_i + \mu_i - \lambda_i^{\frac{1}{2}} \mu_i^{\frac{1}{2}}\right\}$ *are non-negative. If* $\lambda_i \geq 0$, $\mu_i \geq 0$, *then* $\lambda_i^{\frac{1}{2}} \mu_i^{\frac{1}{2}} \geq 0$ *and* $\lambda_i + \mu_i - \lambda_i^{\frac{1}{2}} \mu_i^{\frac{1}{2}}$ *can be expressed as:*

$$\lambda_i + \mu_i - \lambda_i^{\frac{1}{2}} \mu_i^{\frac{1}{2}} = \left(\lambda_i^{\frac{1}{2}} - \mu_i^{\frac{1}{2}}\right)^2 + \lambda_i^{\frac{1}{2}} \mu_i^{\frac{1}{2}} \geq 0.$$

Thus the class of ordered kernels is closed under the intersection and the union as established in Proposition 3.4.5.

Proposition 3.4.7 *Let* K_1, K_2 *be two kernels satisfying the hypothesis of Proposition 3.4.5, such that* $K_1 \lesssim K_2$. *Then*

$$K_1 \lesssim K_1 \wedge K_2 \lesssim K_1 \vee K_2 \lesssim K_2.$$

PROOF. $K_1 \lesssim K_2$ is equivalent to $\lambda_i \leq \mu_i$, $\forall i$. Proposition 3.4.5 shows that $\left\{\lambda_i^{\frac{1}{2}} \mu_i^{\frac{1}{2}}\right\}$, $\left\{\lambda_i + \mu_i - \lambda_i^{\frac{1}{2}} \mu_i^{\frac{1}{2}}\right\}$ are the sets of eigenvalues of $K_1 \wedge K_2, K_1 \vee K_2$, respectively. The relations are a consequence of

$$\lambda_i \leq \lambda_i^{\frac{1}{2}} \mu_i^{\frac{1}{2}} \leq \lambda_i + \mu_i - \lambda_i^{\frac{1}{2}} \mu_i^{\frac{1}{2}} \leq \mu_i.$$

□

Dimension

Proposition 3.4.5 provides the expression of the eigenvalues of $K_1 \wedge K_2$ and $K_1 \vee K_2$, whenever $K_1 \sim K_2$.

Proposition 3.4.8 *Let* $K_1 \sim K_2$ *be two kernels with dimensions* $\dim(K_1) = N_1, \dim(K_2) = N_2$. *The dimensions of* $K_1 \wedge K_2$ *and* $K_1 \vee K_2$ *satisfy*

$$\dim(K_1 \wedge K_2) = \min\{N_1, N_2\} \leq \dim(K_1 \vee K_2) = \max\{N_1, N_2\}.$$

PROOF. $\dim(K_1 \wedge K_2)$ is the number of positive products $\lambda_i^{\frac{1}{2}} \mu_i^{\frac{1}{2}}$. This product is 0 if λ_i or μ_i is zero. For $\dim(K_1 \vee K_2)$, we have 0 if both λ_i, μ_i are 0. □

Some examples illustrate these results. We start with the simple case, $K_1 = K_2 = \sum_{i=0}^{N} \lambda_i \xi_i \otimes \xi_i$. The product

$$K_1 \star K_2 = \sum_{i=0}^{N} \lambda_i^2 \xi_i \otimes \xi_i$$

has the same dimension as K_1 and K_2.

Suppose now $N_2 < N,\ K_2 = \sum_{i=0}^{N_2} \lambda_i \xi_i \otimes \xi_i$. Then

$$K_1 \star K_2 = \sum_{i=0}^{N_2} \lambda_i^2 \xi_i \otimes \xi_i$$

and the dimension of the product is $N_2 = \dim(K_2)$. Observe that this result holds for every $N \leq \infty$ and $N_2 \geq 0$.

If $K_2 = \sum_{i=0}^{N_2} \mu_i \xi_i \otimes \xi_i$ the dimension does not change as the product is

$$K_1 \star K_2 = \sum_{i=0}^{N_2} \lambda_i \mu_i \xi_i \otimes \xi_i.$$

Trace

Proposition 3.3.16 applied to the sets of eigenvalues of the intersection and union of equivalent matrices provide immediate proofs for the following results concerning two equivalent kernels $K_1 \sim K_2$.

Proposition 3.4.9 *Let* $\{\lambda_i\}, \{\mu_i\}$ *be the set of eigenvalues. The traces are:*

$$\begin{aligned} tr\,(K_1 \wedge K_2) &= \sum_{i=0}^{\infty} \lambda_i^{\frac{1}{2}} \mu_i^{\frac{1}{2}}, \\ tr\,(K_1 \vee K_2) &= tr\,(K_1) + tr\,(K_2) - tr\,(K_1 \wedge K_2)\,. \end{aligned}$$

Proposition 3.4.10 *If* $K_1 \lesssim K_2$*, then*

$$tr\,(K_1) \leq tr\,(K_1 \wedge K_2) \leq tr\,(K_1 \vee K_2) \leq tr\,(K_2)\,.$$

PROOF. Use Proposition 3.4.7. □

3.5 Concluding remarks

We have obtained some definitions and results as a theoretical framework for a continuous extension of the so-called Related Metric Scaling and related topics. The following table summarizes the main definitions and properties.

Operations with symmetric kernels

$H-$product $\star$ with H the cdf of (X, Y)

Definition:	$(K_1 \star K_2)_H (\cdot,\cdot) = Cov\,(K_1(\cdot, X), K_2(Y, \cdot))$
Properties:	1. $(K_1 \star K_2)_H (x, y) = (K_2 \star K_1)_{H'} (y, x)$
	2. If $K = \sum \lambda_i \xi_i \otimes \xi_i$, $K^{\frac{1}{2}} = \sum \lambda_i^{\frac{1}{2}} \xi_i \otimes \xi_i$ satisfies $K^{\frac{1}{2}} \star K^{\frac{1}{2}} = K$

Intersection $\wedge$ and union $\vee$

Definitions:	$K_1 \wedge K_2 = \frac{1}{2}\left(K_1^{\frac{1}{2}} \star K_2^{\frac{1}{2}} + K_2^{\frac{1}{2}} \star K_1^{\frac{1}{2}}\right)$
	$K_1 \vee K_2 = K_1 + K_2 - K_1 \wedge K_2$
Properties:	1. Commutativity: $K_1 \wedge K_2 = K_2 \wedge K_1, \quad K_1 \vee K_2 = K_2 \vee K_1$
	2. Orthogonality: if X, Y are stochastically independent,
	$K_1 \wedge K_2 = K_0, \quad K_1 \vee K_2 = K_1 + K_2$
	3. Equality: if $H = H^+$ then $\quad K_1 \wedge K_1 = K_1 \vee K_1 = K_1$
	4. Null element: $K_1 \wedge K_0 = K_0, \quad K_1 \vee K_0 = K_1$
	5. If $K = \sum \lambda_i \xi_i \otimes \xi_i$, $tr\,(K) = \sum \lambda_i$

Partial order $\lesssim$

Definition:	$K_1 \lesssim K_2$ if $K_1 = \sum \lambda_i \xi_i \otimes \xi_i$, $K_2 = \sum \mu_i \xi_i \otimes \xi_i$ and $\lambda_i \leq \mu_i \; \forall i$
Properties:	1. If $K_1 \lesssim K_2$, $tr\,(K_1) \leq tr\,(K_2)$
	2. If $K_1 \lesssim K_2$, $tr\,(K_1) \leq tr\,(K_1 \wedge K_2) \leq tr\,(K_2 \wedge K_1) \leq tr\,(K_2)$

Chapter 4

Covariance and affinity between functions

Cuadras (2002) generalized Hoeffding's lemma (1940), which gives the covariance in terms of the cumulative distribution functions, to the covariance between functions of bounded variation. Further consequences are obtained, especially some relevant relations between this generalized covariance and the Fréchet bounds.

4.1 Covariance and distributions

Hoeffding's lemma [52] gives an expression for the covariance using the joint and marginal cdf's.

Theorem 4.1.1 *(Hoeffding, 1940) Let X, Y be two square integrable random variables, with supports $[a,b]$, $[c,d] \subset \mathbb{R}$, bivariate cdf H and univariate marginals F and G, respectively. The covariance between X and Y in terms of their cdf's is given by*

$$Cov\,(X,Y) = \int_a^b \int_c^d (H(x,y) - F(x)G(y))\,dxdy.$$

The relevance of this formula for the covariance has been proved by several authors in studying the covariance kernel $H(x,y) - F(x)G(y)$ and obtaining, in an easier way, some measures of dependence between r.v.'s: Spearman's correlation coefficient, Pearson's correlation coefficient, Kendall's tau, etc. (see Section 4.4 below). Some authors have generalized this result to the covariance between functions of r.v.'s. Thus Mardia and Thompson [80] proved that

$$Cov\left(X^r, Y^s\right) = \int_a^b \int_c^d \left(H(x,y) - F(x)G(y)\right) r x^{r-1} s y^{s-1} dx dy,$$

and Cuadras [18] gave a generalized Hoeffding's lemma to the wider class of functions of bounded variation.

Let X, Y be two r.v.'s which satisfy the hypotheses of Theorem 4.1.1. The following theorem is an extension. Hypotheses in the following theorem are assumed to be satisfied by all the variables and functions that appear along this chapter. See [18] for another proof. Here we present a completely different integral proof. $BV\left([a,b]\right)$ stands for the set of functions of bounded variation on $[a,b]$.

Theorem 4.1.2 *If $\alpha(x)$ and $\beta(y)$ are two functions defined on $[a,b]$, $[c,d]$, respectively, such that:*

1. *Both functions are of bounded variation, $\alpha \in BV\left([a,b]\right), \beta \in BV\left([c,d]\right),$*
2. *$E\left(|\alpha(X)\beta(Y)|\right), E\left(|\alpha(X)|\right), E\left(|\beta(Y)|\right) < \infty,$*

then

$$Cov\left(\alpha(X), \beta(Y)\right) = \int_a^b \int_c^d \left(H(x,y) - F(x)G(y)\right) d\alpha(x) d\beta(y). \tag{4.1}$$

PROOF. By Fubini's theorem, (4.1) is

$$\int_a^b \int_c^d H(x,y) d\alpha(x) d\beta(y) - \int_a^b F(x) d\alpha(x) \int_c^d G(y) d\beta(y).$$

Let us denote these three integrals as I_H, I_F, I_G, respectively. The covariance $E(\alpha(X)\beta(Y)) - E(\alpha(X))E(\beta(Y))$ is

$$Q = \int_S \alpha(x)\beta(y) dH(x,y) - \int_a^b \alpha(x) dF(x) \int_c^d \beta(y) dG(y),$$

where $S = [a,b] \times [c,d]$. Integration by parts gives

$$A = \int_a^b \alpha(x) dF(x) = \alpha(b) - \int_a^b F(x) d\alpha(x) = \alpha(b) - I_F,$$

$$B = \int_c^d \beta(y) dG(y) = \beta(d) - \int_c^d G(y) d\beta(y) = \beta(d) - I_G.$$

By Fubini's theorem for transition probabilities

$$E[\phi(X,Y)] = \int_a^b \left(\int_c^d \phi(x,y) dG_x(y) \right) dF(x),$$

where $G_x(y)$ is the cdf of Y given $X = x$, and we can write

$$C = \int_S \alpha(x)\beta(y) dH(x,y) = \int_a^b \alpha(x) \left(\int_c^d \beta(y) dG_x(y) \right) dF(x).$$

We first integrate with respect to y. Setting $u = \beta(y)$, $dv = dG_x(y)$ (and $v = \int_c^y dG_x(t)$), integration by parts gives

$$\begin{aligned} \int_c^d \beta(y) dG_x(y) &= \beta(d) \int_c^d dG_x(t) - \int_c^d \int_c^y dG_x(t) d\beta(y) \\ &= \beta(d) - \int_c^d \int_c^y dG_x(t) d\beta(y). \end{aligned}$$

Since $dG_x(t)dF(x) = dH(x,t)$, we have that

$$\begin{aligned} C &= \beta(d) \int_a^b \alpha(x) dF(x) - \int_c^d \left(\int_a^b \alpha(x) \int_c^y dG_x(t) dF(x) \right) d\beta(y) \\ &= \beta(d) \left[\alpha(b) - I_F \right] - \int_c^d \left(\int_a^b \alpha(x) \int_c^y dH(x,t) \right) d\beta(y). \end{aligned}$$

Now we integrate with respect to x. Setting $u = \alpha(x)$, $dv = \int_c^y dH(x,t)$ (so that $v = \int_a^x \int_c^y dH(s,t) = H(x,y)$), we find

$$\begin{aligned} \int_a^b \alpha(x) \int_c^y dH(x,t) &= \alpha(x) H(x,y)]_a^b - \int_a^b H(x,y) \, d\alpha(x) \\ &= \alpha(b) G(y) - \int_a^b H(x,y) \, d\alpha(x). \end{aligned}$$

Finally, we integrate with respect to y:

$$\begin{aligned} \int_c^d \left(\int_a^b \alpha(x) \int_c^y dH(x,t) \right) d\beta(y) &= \int_c^d \left(\alpha(b) G(y) - \int_a^b H(x,y) \, d\alpha(x) \right) d\beta(y) \\ &= \alpha(b) I_G - I_H. \end{aligned}$$

Hence,

$$C = \beta(d)\alpha(b) - \beta(d) I_F - \alpha(b) I_G + I_H.$$

Therefore the covariance $Q = C - A \cdot B$ is

$$Q = \beta(d)\alpha(b) - \beta(d) I_F - \alpha(b) I_G + I_H - (\alpha(b) - I_F)(\beta(d) - I_G),$$

and a final simplification shows that $Q = I_H - I_F \cdot I_G$. □

4.2 Properties of the covariance between functions

In this section, we introduce an inner product and study some properties of the covariance between functions obtained from (4.1), most of them studied and proved in [18]. Let us write $H - F \otimes F$ for $H(x, y) - F(x)F(y)$. These properties are derived from the fact that the continuous, symmetric, positive kernel $H - F \otimes F$ is Riemann-Stieltjes integrable with respect to α and β, i.e.,

$$\begin{aligned} \int_a^b (H(x, y) - F(x) F(y)) \; d\beta(y) &\in RS(\alpha), \\ \int_a^b (H(x, y) - F(x) F(y)) \; d\alpha(x) &\in RS(\beta) \end{aligned}$$

or, simply, $H - F \otimes F \in RS(\alpha \times \beta)$ (see, for instance, [7], [8]).

4.2.1 Defining an inner product

The above covariance between functions motivates the following "inner product" between two functions α, β. This operation depends on H.

Definition 4.2.1 *On the set of functions of bounded variation we define the inner product of two functions by*

$$\langle \alpha, \beta \rangle_H = Cov(\alpha(X), \beta(Y)). \tag{4.2}$$

Note that this covariance is defined on (X, Y) with bivariate cdf H fixed. However the obvious equality $Cov(\alpha(X), \beta(Y)) = Cov(\beta(Y), \alpha(X))$ does not imply $\langle \alpha, \beta \rangle_H = \langle \beta, \alpha \rangle_H$ since $\langle \beta, \alpha \rangle_H = Cov(\beta(X), \alpha(Y))$. Cuadras ([18]) proved some properties of $\langle \alpha, \beta \rangle_H$, regarded as an inner product, when H satisfies:

- H is symmetric, i.e., $H(x, y) = H(y, x)$ for every $(x, y) \in \mathbb{R}^2$ and hence, $F = G$,
- H is positive quadrant dependent (PQD), i.e., $H(x, y) \geq F(x) F(y)$.

See [77], [56] for this and other related concepts of dependence.

The next theorem justifies that $\langle \cdot, \cdot \rangle_H$ can be considered as an inner product on the vector space of the real functions of bounded variation on an interval $[a, b]$, $BV([a, b])$, if H is symmetric and PQD. Observe that the Riemann-Stieltjes integral of a function of $BV([a,b])$, with respect to a given function $\alpha \in BV([a,b])$,

$$\begin{aligned} \langle \alpha, \cdot \rangle : \quad & BV([a,b]) \to \mathbb{R} \\ & \beta \to \langle \alpha, \beta \rangle_H \end{aligned}$$

defines a linear form on the set of functions of bounded variation on a suitable interval. Analogously $\langle \cdot, \alpha \rangle$.

Theorem 4.2.2 *Suppose that $H(x,y)$ is symmetric in x, y and PQD. Let $\alpha, \beta \in BV([a,b])$, $[a,b] \subseteq \mathbb{R}$. Then $\langle \cdot, \cdot \rangle_H \equiv \langle \cdot, \cdot \rangle$ satisfies:*

1. *For every $\alpha \in BV([a,b])$, $\langle \cdot, \alpha \rangle$ and $\langle \alpha, \cdot \rangle$ are linear:*

$$\langle r\alpha_1 + s\alpha_2, l\beta_1 + q\beta_2 \rangle = rl \langle \alpha_1, \beta_1 \rangle + rq \langle \alpha_1, \beta_2 \rangle + sl \langle \alpha_2, \beta_1 \rangle + sq \langle \alpha_2, \beta_2 \rangle,$$

where $r, s, l, q \in \mathbb{R}$. If $\mathbf{0}$ is the zero constant function, then $\langle \alpha, \mathbf{0} \rangle = \langle \mathbf{0}, \alpha \rangle = 0$.

2. *$\langle \alpha, \beta \rangle = \langle \beta, \alpha \rangle$ (symmetry).*

3. *$\langle \alpha, \alpha \rangle \geq 0$ (non negativity).*

PROOF. Symmetry and non negativity were proved in [18]. The first statement in 1 is an obvious property derived from properties of Riemann-Stieltjes integral (see, for instance, [58]). To prove the second statement, it also suffices to apply linearity of Riemann-Stieltjes integral (see, for instance, [7]). □

Remark 4.2.3 *When a symmetric bivariate cdf H satisfies property 3, i.e.,*

$$Cov(\alpha(X), \alpha(Y)) \geq 0$$

for every real-valued function α, then (X,Y) or H is positive function dependent, [56].

These properties provide the following corollary.

Corollary 4.2.4 *The vector space $BV([a,b])$ is a pre-Hilbert space with the inner product $\langle \cdot, \cdot \rangle_H$.*

Remark 4.2.5 *It is worth noting that positivity of the inner product does not hold. Thus $\langle \alpha, \alpha \rangle_H = 0$, with $\alpha \neq \mathbf{0}$ is possible. This property is analogous to that presented by the covariance between r.v.'s. Some authors use the term quasi-inner product in such a case. For example, if U and V are $[0,1]$ uniform with bivariate cdf belonging to the Farlie-Gumbel-Morgersten (FGM) family of distributions, then the copula $C(u,v)$ is*

$$C(u,v) = uv(1 + \theta(1-u)(1-v)), \quad \theta \in [-1,1]. \tag{4.3}$$

Take as α and β the second shifted Legendre polynomial $\alpha(z) = \sqrt{5}\,(6z^2 - 6z + 1)$, $\beta = \alpha$. Clearly $C(u,v) - uv$ is bounded on $[0,1]$. Let α' be the first derivative of α. Since $\alpha, \alpha' \in \mathcal{C}^1$, $(C(u,v) - uv)\alpha'(u)$ is Riemann integrable on $[0,1]$ so that

$$\int_0^1 \int_0^1 (C(u,v) - uv) d\alpha(u) d\alpha(v) = \int_0^1 \int_0^1 (C(u,v) - uv)\alpha'(u)\, du\; \alpha'(v)\, dv$$

(see Example 3.2. in [7]). Then $\alpha \neq \mathbf{0}$, but the covariance between $\alpha(U)$ and $\alpha(V)$ with respect to C is

$$\begin{aligned} \langle \alpha, \alpha \rangle_C &= \int_0^1 \int_0^1 (C(u,v) - uv) d\alpha(u) d\alpha(v) \\ &= 5\theta \int_0^1 \int_0^1 uv(1-u)(1-v)(12u-6)(12v-6) du dv \\ &= 0. \end{aligned}$$

We have established that $\langle \cdot, \cdot \rangle_H$ is a quasi-inner product of the vector space $BV([a,b])$, $[a,b] \subset \bar{\mathbb{R}}$, or, equivalently, $BV([a,b])$ is a pre-Hilbert space with the product $\langle \cdot, \cdot \rangle_H$. Its norm is given by

$$||\alpha||_H^2 = \langle \alpha, \alpha \rangle_H.$$

Orthogonality of α, β is naturally defined by $\langle \alpha, \beta \rangle_H = 0$. Example in Remark 4.2.5 provides a verification of orthogonality of $\alpha(z) = \sqrt{5}\,(6z^2 - 6z + 1)$ with respect to itself relative to the symmetric, PQD, bivariate copula with the FGM distribution.

Proposition 4.2.6 *Let $\alpha, \beta \in BV([a,b])$ and $H = F \otimes F$ be the independent bivariate cdf. Then*

$$\langle \alpha, \beta \rangle_{F \otimes F} = 0,$$

i.e., all functions belonging to $BV([a,b])$ are orthogonal relative to $F \otimes F$.

Proof. Note that $H - F \otimes F = 0$ for $H = F \otimes F$. Then $\langle \alpha, \beta \rangle_{F \otimes F}$ is the (double) Riemann-Stieltjes integral of 0. □

Proposition 4.2.7 *Let* $\alpha \in BV([a,b])$ *and* κ *a real constant. Then for any cdf* H *with support in* $[a,b] \times [a,b]$,

$$\langle \alpha, \kappa \rangle_H = \langle \kappa, \alpha \rangle_H = 0.$$

PROOF. This follows from $d\kappa = 0$. □

Proposition 4.2.8 $\langle \alpha, \beta \rangle_H$ *satisfies*

$$\langle \alpha, \beta \rangle_H^2 \leq \langle \alpha, \alpha \rangle_H \langle \beta, \beta \rangle_H \tag{4.4}$$

with equality if and only if

$$\langle \alpha + t\beta, \alpha + t\beta \rangle_H = 0 \tag{4.5}$$

for some constant t.

PROOF. Expanding $\langle \alpha + t\beta, \alpha + t\beta \rangle_H$ we obtain the inequality

$$\langle \alpha, \alpha \rangle_H + t^2 \langle \beta, \beta \rangle_H + 2t \langle \alpha, \beta \rangle_H \geq 0,$$

which holds only if the discriminant satisfies

$$4 \langle \alpha, \beta \rangle_H^2 - 4 \langle \alpha, \alpha \rangle_H \langle \beta, \beta \rangle_H \leq 0,$$

with $\langle \alpha + t\beta, \alpha + t\beta \rangle_H = 0$ iff this discriminant is 0. Moreover, $t = -\frac{\langle \alpha, \beta \rangle_H}{\langle \beta, \beta \rangle_H}$ is the constant satisfying (4.5), assuming $\langle \beta, \beta \rangle_H > 0$. □

Notice that (4.4) is the Cauchy-Schwartz inequality, an expected property since $BV([a,b])$ is a pre-Hilbert space with $\langle \cdot, \cdot \rangle_H$.

Example 4.2.9 *Let* C *denote the FGM copula (4.3) and let* $\alpha, \beta \in BV([0,1])$ *such that*

$$\alpha(u) = u, \quad \beta(v) = \frac{1}{3} v^3.$$

Straightforward computations show that

$$\langle \alpha, \alpha \rangle_C = \frac{\theta}{36}, \quad \langle \beta, \beta \rangle_C = \frac{\theta}{400}, \quad \langle \alpha, \beta \rangle_C = \frac{\theta}{120}.$$

Thus inequality (4.4) holds but with equality, i.e.,

$$\langle \alpha, \beta \rangle_H^2 = \langle \alpha, \alpha \rangle_H \langle \beta, \beta \rangle_H,$$

in spite of $\alpha \neq \beta$. *The value* t *satisfying (4.5) is* $t = -\frac{10}{3}$.

4.2.2 Two equivalence classes

Some properties of the covariance between the functions α and β can be obtained more easily by centring the variables $\alpha(X), \beta(Y)$. In general, the covariance between r.v.'s is not influenced by changes in the location parameter but is not invariant under changes of scale. Here we study the action of some usual transformations on the space of functions of bounded variation defined on the range of a r.v., with special attention to their influence on the covariance between these functions.

Definition 4.2.10 *Let $\alpha \in BV([a,b])$. We say that $\gamma \in BV([a,b])$ is equivalent to α and write $\gamma \sim \alpha$ if there exists a real number g such that $\gamma(x) = \alpha(x) + g$ (for almost every $x \in [a,b]$). The equivalence class containing α is noted by $[\alpha]$.*

Remark 4.2.11 *Observe that for $\gamma \sim \alpha$:*

$$d\gamma(x) = d(\alpha(x) + g) = d\alpha(x).$$

Consequently, for any $\beta \in BV([a,b])$,

$$\begin{aligned} \langle \gamma, \beta \rangle_H &= \int\int (H - F \otimes G)\, d\gamma d\beta \\ &= \int\int (H - F \otimes G)\, d\alpha d\beta \\ &= \langle \alpha, \beta \rangle_H . \end{aligned}$$

This is a well-known result (see [8]).

Example 4.2.12 *Let $\alpha \in BV([a,b])$, $\beta \in BV([c,d])$, and consider the centred variables $\alpha_0(X) \equiv \alpha(X) - E(\alpha(X))$, $\beta_0(Y) \equiv \beta(Y) - E(\beta(Y))$. Thus $\alpha_0 \in [\alpha]$, $\beta_0 \in [\beta]$ and $E(\alpha_0(X)) = E(\beta_0(Y)) = 0$. Then*

$$\langle \alpha_0, \beta_0 \rangle_H = E_{(X,Y)}(\alpha_0(X)\beta_0(Y)).$$

Corollary 4.2.13 *The covariance between $\alpha(X)$, $\beta(Y)$, where α, β are functions of bounded variation defined on the ranges of the r.v.'s X and Y, respectively, is invariant under changes of position.*

Proof. This result follows from Remark 4.2.11. It can also be derived from Property 1 of Theorem 4.2.2 and Proposition 4.2.7. □

Some immediate consequences are presented. Corollary 4.2.13 states that the covariance between elements of two equivalence classes $[\alpha]$ and $[\beta]$ is the same, and any result must be considered valid for almost every element. Moreover, the computation can be made simpler by centring the r.v.'s, i.e., by considering α_0 and β_0.

Corollary 4.2.14 *The covariance between two functions of bounded variation $\gamma, \delta \in BV([a,b])$, defined on the range of the r.v.'s X, Y with symmetric, PQD bivariate cdf H, is*

$$Cov\,(\gamma(X), \delta(Y)) = \int_a^b \int_a^b \alpha_0(x)\beta_0(y)dH(x,y)$$

for almost every $\gamma \in [\alpha]$, $\delta \in [\beta]$, where $\alpha_0 \in [\alpha]$, and $\beta_0 \in [\beta]$ satisfy $E(\alpha_0(X)) = E(\beta_0(Y)) = 0$.

Proof. Take $\gamma(X) = \alpha_0(X)$, $\delta(Y) = \beta_0(Y)$ and apply Corollary 4.2.13. □

A wider equivalence class including both translation and change of scale is next defined.

Definition 4.2.15 *Let α be a function of bounded variation. We say that ξ is equivalent to α ($\xi \overset{*}{\sim} \alpha$) if there exist real numbers h, g, $h \neq 0$, such that $\xi(x) = h\alpha(x) + g$ a.e. This equivalence class of α is noted by $[\alpha]^*$.*

Remark 4.2.16 *Observe that for any two members of a class $[\alpha]^*$, $\xi \overset{*}{\sim} \alpha$:*

$$d\xi(x) = d(h\alpha(x) + g) = hd\alpha(x)$$

where h, g are suitable constants. Consequently, for any $\beta \in BV([a,b])$,

$$\begin{aligned} \langle \xi, \beta \rangle_H &= \int\int (H - F \otimes G)\, d\xi d\beta \\ &= \int\int (H - F \otimes G)\, hd\alpha d\beta \\ &= h \langle \alpha, \beta \rangle_H . \end{aligned}$$

4.2.3 An affinity measure between functions

Mardia and Thompson [80] used Theorem 4.1.1 to extend the concept of covariance, even when the ordinary covariance does not exist, e.g., when the variances are zero.

Cuadras [18] extended the correlation coefficient between functions in the same way

$$Cor\left(\alpha\left(X\right),\beta\left(Y\right)\right)=\frac{Cov\left(\alpha\left(X\right),\beta\left(Y\right)\right)}{\left(Var(\alpha(X))Var\left(\beta\left(Y\right)\right)\right)^{\frac{1}{2}}} \tag{4.6}$$

where

$$Var(\alpha(X)=\int_a^b\int_a^b\left(\min\left\{F\left(x\right),F\left(y\right)\right\}-F\left(x\right)F\left(y\right)\right)d\alpha\left(x\right)d\alpha\left(y\right).$$

This correlation can exist even when the variance is 0. Also note that (4.6) is Pearson's correlation coefficient when α and β are the identity function.

From inequality (4.4) in Proposition 4.2.8 we can define another coefficient $A_H\left(\alpha,\beta\right)$, such that $A_H^2\left(\alpha,\beta\right)$ ranges between 0 and 1. This may provide a measure of affinity between the functions α and β with respect to H.

Definition 4.2.17 *Let $H(x,y)$ be a bivariate cdf, symmetric in x,y, and PQD. Let $\alpha,\beta\in BV\left([a,b]\right)$. Then $A_H\left(\alpha,\beta\right)$, defined by*

$$A_H\left(\alpha,\beta\right)=\frac{\left\langle\alpha,\beta\right\rangle_H}{\left(\left\langle\alpha,\alpha\right\rangle_H\left\langle\beta,\beta\right\rangle_H\right)^{\frac{1}{2}}}$$

is the $H-$affinity between the functions α and β or affinity with respect to H.

Notice that the correlation coefficient (4.6) does not coincide, in general, with $A_H\left(\alpha,\beta\right)$ since

$$Cov(\alpha\left(X\right),\alpha\left(Y\right))\neq Var(\alpha(X))$$

if $H\neq H^+$.

Every measure of dependence should satisfy certain conditions, and these conditions have been proposed and studied by various authors. There is an early work by Rényi in 1959 [93], whose conclusions have been reviewed by Lancaster [72], [75], Schweizer and Wolf [98] and Hutchinson and Lai [54], among others (see Section 1.1 in Chapter 1). However A_H measures the concordance between functions rather than r.v.'s. The following result summarizes the main properties of A_H.

Proposition 4.2.18 *The H-affinity measure A_H satisfies:*

1. $0\leq A_H^2\leq 1$.
2. *If the r.v. X and Y are stochastically independent, then* $A_H^2=0$.

3. *If α, β are orthogonal with respect to $\langle \cdot, \cdot \rangle_H$, then $A_H^2 = 0$.*

4. *Let X, Y be two non independent random variables. If $\alpha = h\beta + g$, for some constants $h, g \in \mathbb{R}$, $h \neq 0$, and α, β are not orthogonal with respect to $\langle \cdot, \cdot \rangle_H$, then $A_H^2 = 1$; otherwise, $A_H^2 = 0$.*

5. *The $H-$affinity between functions of r.v.'s is invariant under the group of linear transformations.*

Proof. Statement 1 follows from inequality (4.4) and Definition 4.2.17. Statements 2 and 3 follow from the analogous properties for the covariance: if X and Y are stochastically independent, then $H = F \otimes F$, as mentioned above, and obviously $H - F \otimes F = 0$. Then $\langle \alpha, \beta \rangle_{F \otimes F} = 0$, i.e., α, β are always orthogonal with respect to the independence distribution $F \otimes F$. The proof of 4 is as follows. If $\alpha(x) = h\beta(x) + g$, as a consequence of Remark 4.2.16 we have $\langle \alpha, \beta \rangle_H = h \langle \beta, \beta \rangle_H$, and $\langle \alpha, \alpha \rangle_H = h^2 \langle \beta, \beta \rangle_H$. Then

$$\begin{aligned} A_H^2(\alpha, \beta) &= \frac{\langle \alpha, \beta \rangle_H^2}{\langle \alpha, \alpha \rangle_H \langle \beta, \beta \rangle_H} \\ &= 1. \end{aligned}$$

The proof of 5 is easily derived from the analogous properties of the covariance. Invariance under linear transformations means that if $\gamma(x) = h\alpha(x) + g \quad a.e.$ and, analogously, $\delta(y) = l\beta(y) + d \quad a.e.$, then

$$A_H(\gamma, \delta) = A_H(\alpha, \beta).$$

Recall that covariance is not invariant under changes of scale but is invariant under translations:

$$\begin{aligned} \langle \gamma, \delta \rangle_H &= hl \langle \alpha, \beta \rangle_H, \\ \langle \gamma, \gamma \rangle_H &= h^2 \langle \alpha, \alpha \rangle_H, \\ \langle \delta, \delta \rangle_H &= l^2 \langle \beta, \beta \rangle_H. \end{aligned}$$

Thus, substituting these equalities in Definition 4.2.17, the invariance under linear transformations is proved. □

This result justifies the choice of normalized functions to compute A_H, in order to obtain simpler expressions.

Remark 4.2.19 *Observe that α and β in Example 4.2.9 are not related by a linear transformation while $A_C(\alpha,\beta) = 1$. This result is generalized below.*

Corollary 4.2.20 $A_H(\alpha,\alpha) = 1$ *if* $H \neq F \otimes F$, *and* $A_H(\alpha,\alpha) = 0$, *if* $H = F \otimes F$.

PROOF. This is a particular case of Proposition 4.2.18, with $h = 1$, $g = 0$. □

4.3 Fréchet classes and bounds

Covariance and the $H-$affinity A_H are functions defined on the (product) vector space of functions of bounded variation, for a fixed, symmetric, PQD, bivariate cdf H. It is well known that if $H \in \digamma(F,F)$, the class of bivariate cdf's with marginals F, the lower and upper bounds of this Fréchet class satisfy

$$H^{-}(x,y) \le H(x,y) \le H^{+}(x,y).$$

Note that H^- is not PQD, as $H^-(x,y) \le F(x)F(y)$, so (4.4) does not apply, in general. Thus we can define A_{H^+} but not A_{H^-}. In the following, we summarize and obtain further results on the covariance and the affinity A_H for some particular elements of a given Fréchet class (the upper bound and the independence bivariate cdf) when α and β are fixed. Finally, we obtain bounds for the covariance and A_H.

4.3.1 Covariance and bounds

This proposition is in fact a corollary of the main theorem 4.1.2.

Proposition 4.3.1 *The covariance between two functions $\alpha \in BV([a,b])$, $\beta \in BV([c,d])$ when the bivariate cdf H attains the Fréchet upper bound H^+ is given by*

$$\langle\alpha,\beta\rangle_{H^+} = \int_a^b \int_c^d [\min\{F(x), G(y)\} - F(x)G(y)] d\alpha(x)\, d\beta(y).$$

Proposition 4.3.2 *Let X, Y be two r.v.'s with range $[a,b]$, symmetric cdf H and common marginal cdf F. Let $\alpha \in BV([a,b])$ and $\beta \in BV([c,d])$. Consider $\alpha_0 \equiv \alpha - E(\alpha(X))$ and $\beta_0 \equiv \beta - E(\beta(Y))$. Then*

$$\langle\alpha,\beta\rangle_{H^+} = \int_a^b \alpha_0(s)\beta_0(s) dF(s).$$

PROOF. From Corollary 4.2.14 and Proposition 4.3.1,

$$\langle \alpha, \beta \rangle_{H^+} = \int_a^b \int_a^b [\min\{F(x), F(y)\} - F(x)F(y)] d\alpha_0(x) d\beta_0(y).$$

Also note that if $H = H^+$, then $F(X) = G(Y)$ (a.s.). Since $F = G$ are continuous cdf's, $Y = F^{-1}(F(X)) = X$ (a.s.). Hence, taking the diagonal of the distribution

$$H^+(s, s) = \min\{F(s), F(s)\} = F(s),$$

the result is directly obtained from Corollary 4.2.14. □

Cuadras and Lahlou ([30]) stated the following general result:

Theorem 4.3.3 *Let X be a r.v. with range $[a, b]$ and cdf F. Suppose that α, β are functions of bounded variation defined on $[a, b]$ such that $\alpha(a)F(a) = \beta(a)F(a) = 0$. Then*

$$Cov(\alpha(X), \beta(X)) = \int_a^b \int_a^b [\min\{F(x), F(y)\} - F(x)F(y)] d\alpha(x) d\beta(y).$$

PROOF. This is a direct consequence of Theorem 4.1.2. □

Corollary 4.3.4 *Suppose that $\alpha \in BV([a, b])$ satisfies $\alpha(a)F(a) = 0$. Then*

$$Var(\alpha(X)) = Var(\alpha(Y)) = \langle \alpha, \alpha \rangle_{H^+}$$

for any r.v. Y, such that $X = Y$ (a.s.).

Proposition 4.3.5 *The following bounds for $\langle \alpha, \alpha \rangle_H$, when H is symmetric, PQD, hold:*

$$0 = \langle \alpha, \alpha \rangle_{F \otimes F} \leq \langle \alpha, \alpha \rangle_H \leq \langle \alpha, \alpha \rangle_{H^+}.$$

PROOF. We have already proved that $\langle \cdot, \cdot \rangle_{F \otimes F} = 0$ (see Proposition 4.2.18) and that $\langle \alpha, \alpha \rangle_H \geq 0$ for every symmetric PQD cdf H and any α (see 3 in Theorem 4.2.2). The second inequality is straightforward if α is increasing, as $H \leq H^+$ and H is PQD. Thus $0 \leq H - F \otimes F \leq H^+ - F \otimes F$.

For any α, let us consider the positive, continuous and bounded function K

$$K \equiv H^+ - F \otimes F - (H - F \otimes F) \geq 0$$

It follows that $K \in RS(\alpha \times \alpha)$ and let us use an obvious notation to define

$$\langle \alpha, \alpha \rangle_K = \int_a^b \int_a^b K(x,y) d\alpha(x) d\alpha(y) \geq 0.$$

The proof that $\langle \alpha, \alpha \rangle_K \geq 0$ is similar to the proof of $\langle \alpha, \alpha \rangle_H \geq 0$, see [18]. Now it suffices to apply linearity of Riemann-Stieltjes integration to have

$$0 \leq \langle \alpha, \alpha \rangle_K = \langle \alpha, \alpha \rangle_{H^+} - \langle \alpha, \alpha \rangle_H ,$$

so $\langle \alpha, \alpha \rangle_H \leq \langle \alpha, \alpha \rangle_{H^+}$. □

Example 4.3.6 *Suppose U, V are uniform on $[0,1]$, with bivariate cdf belonging to the FGM family of distributions. Let $C(u,v)$ be their copula (4.3), and take the second shifted Legendre polynomial $\alpha(z) = \sqrt{5}(6z^2 - 6z + 1)$ as in the example given in Remark 4.2.5. We have obtained $\langle \alpha, \alpha \rangle_C = 0$. Some computations show that $\langle \alpha, \alpha \rangle_{C^+} = 19$ and the inequalities given by Proposition 4.3.5 hold. Note that the Fréchet bound $C^+(u,v) = \min\{u,v\}$ is not a member of the FGM family.*

Corollary 4.3.7 *Let H be the symmetric, PQD cdf of (X,Y) with common marginal F and $\alpha \in BV([a,b])$, where $[a,b]$ is the support of X, Y. If $\langle \alpha, \alpha \rangle_{H^+} = 0$ then*

1. $\langle \alpha, \alpha \rangle_H = A_H^2(\alpha, \alpha) = 0, \quad \forall H \in \mathcal{F}(F,F)$,
2. $\alpha \in [0]$, *i.e., α is a constant (a.s.).*

Proof. 1 is directly derived from Proposition 4.3.5. To prove 2 recall that $\langle \alpha, \alpha \rangle_{H^+} = Var(\alpha(X))$. If the variance is zero the r.v. is a constant a.s. □

Note that $\langle \alpha, \beta \rangle_H$ may be negative though H is PQD. Hence, it is not possible to obtain similar results for $\langle \alpha, \beta \rangle_H$.

Example 4.3.8 *Let α, β and C as in Example 4.2.9. Then $\langle \alpha, \beta \rangle_C = \frac{\theta}{120}$, $\theta \in [-1,1]$ and C is PQD if $\theta \geq 0$. As $\langle -\alpha, \beta \rangle_C = -\langle \alpha, \beta \rangle_C$ we see that the affinity $A_C(-\alpha, \beta)$ is negative for $0 < \theta \leq 1$.*

This example is not contradictory with the following characterization of the PQD bivariate distributions:

Theorem 4.3.9 *(Lehmann, [77]) Let H be the joint cdf of a random vector (X,Y). H is PQD iff for every pair of increasing functions f, g*

$$Cov(f(X), g(Y)) \geq 0.$$

4.3.2 H-affinity and bounds

The affinity measure between two functions of bounded variation $\alpha, \beta \in BV([a,b])$ relative to H defined in Definition 4.2.17 is bounded in absolute value by 0 and 1, as has already been proved. The following result states that A_H^2 is an upper bound of the correlation coefficient between functions (see (4.6)).

Proposition 4.3.10 *Let $Cor(\alpha(X), \beta(Y))$ be the correlation coefficient between two functions $\alpha(X), \beta(Y)$, and let $A_H^2(\alpha,\beta)$ be the $H-$affinity measure. Then $A_{H^+}(\alpha,\beta) = Cor(\alpha(X), \beta(Y))$ and*

$$A_{H^+}^2(\alpha,\beta) \leq A_H^2(\alpha,\beta).$$

PROOF. $\langle \alpha, \beta \rangle_H = Cov(\alpha(X), \beta(Y))$. Proposition 4.3.5 shows that

$$0 = \langle \alpha, \alpha \rangle_{F \otimes F} \leq \langle \alpha, \alpha \rangle_H \leq \langle \alpha, \alpha \rangle_{H^+},$$

and Corollary 4.3.4 stated that $Var(\alpha(X)) = \langle \alpha, \alpha \rangle_{H^+}$. Hence,

$$\frac{\langle \alpha, \beta \rangle_H^2}{\langle \alpha, \alpha \rangle_{H^+} \langle \beta, \beta \rangle_{H^+}} \leq \frac{\langle \alpha, \beta \rangle_H^2}{\langle \alpha, \alpha \rangle_H \langle \beta, \beta \rangle_H}.$$

□

Example 4.3.11 *Let C be the FGM copula and $\alpha(u) = u$, $\beta(v) = \frac{1}{3}v^3$ as in Example 4.2.9. We find*

$$\langle \alpha, \beta \rangle_C = \frac{\theta}{120}, \quad A_C(\alpha,\beta) = 1 \times sign(\theta).$$

If $C^+(u,v) = \min\{u,v\}$ is the Fréchet upper bound for any copula, then

$$\langle \alpha, \alpha \rangle_{C^+} = \frac{1}{12}, \quad \langle \beta, \beta \rangle_{C^+} = \frac{1}{112},$$

and the correlation coefficient (4.6) is

$$\rho(\alpha(X), \beta(Y)) = \frac{\sqrt{21}}{15}\theta.$$

Since $\theta \in [-1,1]$,

$$\rho^2(\alpha(X), \beta(Y)) < A_C^2(\alpha,\beta) = 1,$$

uniformly in θ.

4.4 Concordance measures and H-affinity

Spearman's ρ_S and Kendall's τ (see Section 1.1 in Chapter 1) are examples of concordance measures between two r.v.'s. These measures are based on the integrals

$$\int\int H_1(x,y)dH_2(x,y),$$

where H_1, H_2 are cdf's, and can be defined by using the respective copulas C_1, C_2.

The concordance measure between C_1, C_2 is

$$Q(C_1,C_2) = 4\int_{I^2} C_2(u,v)dC_1(u,v) - 1.$$

Spearman's ρ_S (1.2) is based on H and FG and is also defined as

$$\rho_S = \frac{Cov\left(F\left(X\right),G\left(Y\right)\right)}{\sqrt{Var\left(F\left(X\right)\right)}\sqrt{Var\left(G\left(Y\right)\right)}}.$$

Thus

$$\begin{aligned}\rho_S &= 12Cov\left(F\left(X\right),G\left(Y\right)\right)\\ &= 12\int\int(H(x,y)-F(x)G(y))dF(x)dG(y)\\ &= 12\int_{I^2}(C(u,v)-uv)dudv\\ &= 3Q(C,\Pi),\end{aligned}$$

where $\Pi(u,v)=uv$ is the independence copula. This coefficient ranges between -1 and 1.

Kendall's τ (1.3) is also defined by

$$\begin{aligned}\tau &= 4\int\int(H(x,y)-F(x)G(y))dH(x,y)\\ &= 4\int_{I^2} C(u,v)dC(u,v)-1\\ &= Q(C,C),\end{aligned}$$

where C is the copula for H. Note that Kendall's τ cannot be expressed as a covariance between functions, as $dH(x,y)$ is not $d\alpha(x)d\beta(y)$ in general.

The $H-$affinity between F and G, is

$$A_H(F(X),G(Y)) = \frac{Cov\left(F\left(X\right),G\left(Y\right)\right)}{\sqrt{Cov\left(F\left(X\right),F\left(Y\right)\right)}\sqrt{Cov\left(G\left(X\right),G\left(Y\right)\right)}}.$$

Assuming $F=G$ we find the obvious result

$$\begin{aligned}A_H(F(X),F(Y)) &= 1 \quad \text{if } H\neq F\otimes F,\\ A_H(F(X),F(Y)) &= 0 \quad \text{if } H= F\otimes F.\end{aligned}$$

However for some cdf's H, the value 1 is also reached for other functions.

Chapter 5

Affinities for some parametric families

The affinity or $H-$affinity measure $A_H(\alpha, \beta)$ between two functions of bounded variation $\alpha, \beta \in BV([a, b])$ relative to the symmetric, PQD bivariate cdf H, has been defined in Chapter 4 by

$$A_H^2(\alpha, \beta) = \frac{\langle \alpha, \beta \rangle_H^2}{\langle \alpha, \alpha \rangle_H \langle \beta, \beta \rangle_H},$$

the inner product $\langle \alpha, \beta \rangle_H$ being

$$\begin{aligned} \langle \alpha, \beta \rangle_H &= Cov(\alpha(X), \beta(Y)) \\ &= \int_a^b \int_a^b (H(x, y) - F(x)G(y))\, d\alpha(x) d\beta(y), \end{aligned}$$

where, as H is symmetric, $F = G$. We have proved that:

a) The correlation coefficient $\rho(\alpha(X), \beta(Y))$ satisfies

$$\rho^2(\alpha(X), \beta(Y)) \leq A_H^2(\alpha, \beta).$$

b) If $A_H^2(\alpha, \beta) = 1$ then $\langle \alpha + t\beta, \alpha + t\beta \rangle_H = 0$ for some real t.

In this chapter we relate the affinity to the correlation coefficient and some concordance measures, and study orthogonality and dimensionality for some families of distributions (see Section 1.2 in Chapter 1 for further details on these parametric families). These results are part of ongoing research, and are an application of the $H-$affinity measure defined in Chapter 4. See [21] for some complete proofs.

5.1 Bivariate dimensionality

Let us define the dimension of H in terms of the above inner product and affinity.

Definition 5.1.1 *Let $\Phi_H = (\varphi_\imath, \imath \in \Im)$ be a set of real functions of $L^2([a,b], F)$ where F is the probability measure induced by the marginal cdf F. The dimension of H is the cardinal $\#(\Phi_H)$, if these functions satisfy:*

1. $A_H^2(\varphi_\imath, \varphi_\jmath) = 0$, $\imath \neq \jmath$, *and* $A_H^2(\varphi_\imath, \varphi_\imath) = 1$, $\varphi_\imath, \varphi_\jmath \in \Phi_H$.

2. $A_H^2(\alpha, \beta) = 0$ *if* $\alpha, \beta \in \Phi_H^\perp$ *where the orthogonality is with respect to* $\langle \cdot, \cdot \rangle_H$.

If the dimension is finite or countable, then

$$\#(\Phi_H) \leq \#(\Phi_{H^+}),$$

as $\langle \alpha, \alpha \rangle_H \leq \langle \alpha, \alpha \rangle_{H^+}$ if H^+ is the upper Fréchet bound (see Chapter 4, Proposition 4.3.5). Thus we may have $\langle \alpha, \alpha \rangle_H = 0$ but $\langle \alpha, \alpha \rangle_{H^+} \neq 0$ for some $\alpha \neq 0$.

Examples of dimensions are:

1. Dimension 0 in the case of stochastic independence.

2. Finite dimension $n > 0$ if H is the generalized FGM family.

3. Countable dimension $\varkappa_0$ if H is the upper Fréchet bound, the regression family or the Ali-Mikhail-Haq family.

4. Continuous dimension $\varkappa_1$ if H is the Cuadras-Augé family.

When $H(x,y) = F(x)G(y)$ there is no function $\varphi \neq 0$ such that $\langle \varphi, \varphi \rangle_H > 0$ and the dimension is 0. The cases of finite, countable and continuous dimension are given below.

Complete orthonormal sets of functions which appear in some expansions for bivariate distributions might satisfy conditions of Definition 5.1.1. For instance, suppose that H is a general bivariate cdf with marginals F, G (possibly $F \neq G$), the measure

$dH(x, y)$ is absolutely continuous with respect to $dF(x)dG(y)$ and that Pearson's contingency coefficient ϕ^2 defined by

$$\phi^2 + 1 = \int_a^b \int_c^d (dH(x, y))^2/(dF(x)dG(y))$$

is finite. Then the following expansion holds

$$dH(x, y) - dF(x)dG(y) = \sum_{n\geq 1} \rho_n a_n(x) b_n(y) dF(x)dG(y), \tag{5.1}$$

where ρ_n are canonical correlations, ordered in descending order, and $a_n(x), b_n(y)$ are the canonical functions (see [74]). Thus $\rho_1 = Cor(a_1(X), b_1(Y))$ is the maximum correlation between a function of X and a function of Y, $\rho_2 = Cor(a_2(X), b_2(Y))$ is also a maximal correlation given that the functions are uncorrelated with $a_1(X)$ and $b_1(Y)$, etc.

Cuadras (see [17]) expresses (5.1) in terms of cdf's

$$H(x, y) - F(x)G(y) = \sum_{n\geq 1} \rho_n \int_a^b L(x, s) da_n(s) \int_c^d M(t, y) db_n(t), \tag{5.2}$$

where $L(x, s) = \min\{F(x), F(s)\} - F(s)F(t)$, $M(t, y) = \min\{G(t), G(y)\} - G(t)G(y)$. In general, there is no relation between the eigenexpansion of the covariance kernel $K = H - F \otimes G$ and the above expansion. An exception is given in Subsection 5.3.1. The set of eigenfunctions of the covariance kernel might also satisfy conditions of Definition 5.1.1.

Proposition 5.1.2 *If the above diagonal expansion exists, the dimension of H is determined by the number of canonical variables with positive canonical correlations.*

Proof. The set of canonical functions $\{a_n\}_{n\in\mathbb{N}}$ is a complete orthogonal system of functions over the interval $[a, b]$ (the range of X). An orthogonal set on $L^2([a, b])$ is complete iff for all $\alpha, \beta \in L^2([a, b])$ we have

$$\langle \alpha, \beta \rangle = \sum_{n=1}^{\infty} \langle \alpha, a_n \rangle \langle \beta, a_n \rangle$$

(see, for instance, Proposition 1 in [99]). Furthermore, there exist real coefficients $\{\alpha_n\}$, $\{\beta_n\}$, say, such that

$$\alpha = \sum_{n=1}^{\infty} \alpha_n a_n, \qquad \beta = \sum_{n=1}^{\infty} \beta_n a_n,$$

i.e., $\{a_n\}_{n\in\mathbb{N}}$ is a basis of $L^2([a,b])$. Suppose $\rho_1 \geq \ldots \geq \rho_N > \rho_{N+1} = 0$. Then $Cov(a_i(X), a_i(Y)) > 0$ if $i \leq N$, $Cov(a_i(X), a_j(Y)) = 0$ if $i \neq j$ and if $i = j > N+1$,

$$\begin{aligned} \langle \alpha, \beta \rangle &= \sum_{i=1}^{\infty} \left\langle \sum_{n=1}^{\infty} \alpha_n a_n, a_i \right\rangle \left\langle \sum_{m=1}^{\infty} \beta_m a_m, a_i \right\rangle \\ &= \sum_{i=1}^{\infty} \left(\sum_{n,m=1}^{\infty} \alpha_n \beta_m \langle a_n, a_i \rangle \langle a_m, a_i \rangle \right) \\ &= \sum_{i=1}^{\infty} \alpha_i \beta_i \langle a_i, a_i \rangle^2 \\ &= \sum_{i=1}^{N} \alpha_i \beta_i \langle a_i, a_i \rangle^2 , \end{aligned} \tag{5.3}$$

where N is finite. Define $\Phi_N = \{a_1, \cdots, a_N\}$. If $\alpha, \beta \in \Phi_N^{\perp}$, for each $i = 1, ..., N$, we have

$$\begin{aligned} \langle \alpha, a_i \rangle &= \left\langle \sum_{n=1}^{\infty} \alpha_n a_n, a_i \right\rangle \\ &= \sum_{n=1}^{\infty} \alpha_n \langle a_n, a_i \rangle \\ &= \alpha_i \langle a_i, a_i \rangle = 0, \end{aligned}$$

and analogously for β. Substituting this result in (5.3), we obtain $\langle \alpha, \beta \rangle = 0$ if $\alpha, \beta \in \Phi_N^{\perp}$. Obviously, the $H-$affinity $A_H^2(a_i, a_j) = \delta_{ij}$, if $a_i, a_j \in \Phi_N$. □

In general, when a distribution H can be expanded by means of a complete orthogonal system of functions, the number of these funtions with positive covariance will give the dimension of H.

5.2 Finite dimension: Generalized FGM family

The Farlie-Gumbel-Morgenstern (FGM) family (see Section 1.2 in Chapter 1) is defined by

$$H_\theta(x,y) = F(x)G(y)(1 + \theta(1 - F(x))(1 - G(y)), \quad -1 \leq \theta \leq 1.$$

The corresponding copula is

$$C_\theta(u,v) = uv(1 + \theta(1-u))(1-v)), \quad -1 \leq \theta \leq 1.$$

Thus, for this copula

$$Cov(\alpha(U), \beta(V)) = \theta I_\alpha I_\beta$$
$$Cov(\alpha(U), \alpha(V)) = \theta I_\alpha I_\alpha$$
$$Cov(\beta(U), \beta(V)) = \theta I_\beta I_\beta$$

where

$$I_\alpha = \int_0^1 u(1-u)d\alpha(u), \quad I_\beta = \int_0^1 v(1-v)d\beta(v).$$

The correlation coefficient is

$$Cor(\alpha(X), \beta(Y)) = \theta I_\alpha I_\beta / \sigma(\alpha)\sigma(\beta) \leq 1/3. \tag{5.4}$$

However we obtain the surprising result for any $\theta \neq 0$

$$\begin{aligned} A_C(\alpha,\beta) &= A_C(\alpha,\alpha) = A_C(\beta,\beta) = 0, && \text{if } \alpha' \text{ or } \beta' \text{ is orthogonal to } u(1-u), \\ A_C(\alpha,\beta) &= (I_\alpha I_\beta / I_\alpha I_\beta) \times sign(\theta) = \pm 1, && \text{otherwise.} \end{aligned}$$

Noting that for the FGM copula the covariance kernel $K(u,v) = C_\theta(u,v) - uv$ is

$$K(u,v) = \theta u(1-u)v(1-v),$$

the above result can be generalized to any one-dimensional kernel.

Proposition 5.2.1 *Suppose that a copula C is such that $K(u,v) = C(u,v) - uv$ satisfies*

$$K(u,v) = \theta\varphi(u)\varphi(v), \quad \textit{with } \varphi(0) = \varphi(1) = 0.$$

Then for any $\theta \neq 0$:

1. *If $\int_0^1 \varphi(u)d\alpha(u) = 0$ then $A_H(\alpha,\beta) = A_H(\alpha,\alpha) = 0$.*
2. *If both $\int_0^1 \varphi(u)d\alpha(u) \neq 0$ and $\int_0^1 \varphi(u)d\beta(u) \neq 0$ then*

$$A_H(\alpha,\beta) = \pm 1, \quad \textit{and } \langle \alpha + t\beta, \alpha + t\beta \rangle_H = 0,$$

for some real t.

Proof. The affinity coefficient depends on the integrals $I_\alpha = \int_0^1 \varphi(u)d\alpha(u), I_\beta = \int_0^1 \varphi(u)d\beta(u)$ and gives 1 if both integrals are $\neq 0$. On the other hand

$$\langle \gamma, \gamma \rangle_H = \theta \left(\int_0^1 \varphi(u)d\gamma(u) \right)^2 = 0$$

if γ satisfies $\int_0^1 \varphi(u)d\gamma(u) = 0$, e.g., if γ' exists and is orthogonal to φ. Then we see that $\langle \alpha + t\beta, \alpha + t\beta \rangle_H = 0$ for

$$t = -\frac{\int_0^1 \varphi(u)d\alpha(u)}{\int_0^1 \varphi(u)d\beta(u)}.$$

□

Next let us consider the Generalized FGM family proposed by Cuadras et al. (see [29])

$$H(x,y) - F(x)G(y) = \sum_{k=1}^{n} \gamma_k L_k^*(F(x))L_k^*(G(y)),$$

where $L_1(x) = 3^{1/2}(2x-1)$, $L_2(x) = 5^{1/2}(6x^2 - 6x + 1), ...$ are the shifted Legendre polynomials on $[0,1]$ and $L_k^*(u) = \int_0^u L_k(t)dt$ is given by

$$L_k^*(u) = \frac{1}{2}\{L_{k+1}(u)/(2k+3)^{1/2} - L_{k-1}(u)/(2k-1)^{1/2}\}/(2k+1)^{1/2}. \qquad (5.5)$$

Then L_i^*, L_j^* are orthogonal with respect to the measure of probability induced by the cdf $F \circ X$, if $\mid i - j \mid \neq 2$; so the expansion above is an eigendecomposition if we take, for instance, $\gamma_1 \geq \gamma_2 > \gamma_3 = \ldots = 0$.

Proposition 5.2.2 *Consider the Generalized FGM copula defined by*

$$C_\gamma(u,v) - uv = \sum_{k=1}^{n} \gamma_k L_k^*(u)L_k^*(v).$$

Suppose that the first derivatives of α, β exist. Then a sufficient condition for $A_{C_\gamma}(\alpha,\beta) = 1$ is the orthogonality of α', β' to the space generated by $L_1^*, \ldots, L_n^*$.

Proof. Condition $A_{C_\gamma}(\alpha,\beta) = 1$ is equivalent to $\langle \alpha + t\beta, \alpha + t\beta \rangle_{C_\gamma} = 0$ for some t. But

$$\begin{aligned}\langle \alpha + t\beta, \alpha + t\beta \rangle_{C_\gamma} &= \int_{[0,1]^2} \sum_{k=1}^{n} \gamma_k L_k^*(u)L_k^*(v)\ d\left(\alpha(u) + t\beta(u)\right) d\left(\alpha(v) + t\beta(v)\right), \\ &= \sum_{k=1}^{n} \gamma_k P_k^2,\end{aligned}$$

where

$$P_k = \int_0^1 L_k^*(u)d\alpha(u) + t\int_0^1 L_k^*(u)d\beta(u),$$

which cancels if $\int_0^1 L_k^*(u)d\alpha(u) = \int_0^1 L_k^*(u)d\beta(u) = 0$. □

Note that the functions of the set $\{L_1^*, \ldots, L_n^*\}$ satisfy conditions of Definition 5.1.1. Thus the dimension of this FGM generalized family is n. In a sense, this parametric copula is similar to the Ali-Mikhail-Haq copula introduced below.

5.3 Countable dimension

5.3.1 The bivariate upper bound

Suppose that X is a r.v. with continuous cdf F and range the interval $[a, b]$. Let $K(x, y) = \min\{F(x), F(y)\} - F(x)F(y)$, where $\min\{F(x), F(y)\}$ is the bivariate upper bound. Let us consider the eigenexpansion

$$K(s,t) = \sum_{n\geq 1} \lambda_n \psi_n(s)\psi_n(t).$$

Note that $\{\psi_n\}$ must be a countable set of continuous functions. If not, we could write

$$\min\{F(x), F(y)\} = F(x)F(y) + \sum_{n=1}^{N} \lambda_n \psi_n(x)\psi_n(y),$$

with $N < \infty$. But this is not possible, as the upper bound $\min\{F(x), F(y)\}$ has a singular component. Thus, if there exists a probability density $f(x)$ (Lebesgue measure), we would have

$$\frac{\partial^2}{\partial x \partial y} \min\{F(x), F(y)\} = 0 \quad \text{if } F(x) \neq F(y).$$

However, the derivative of the right-hand side of the above equation is

$$f(x)f(y) + \sum_{n=1}^{N} \lambda_n \psi_n'(x)\psi_n'(y) \neq 0.$$

Let us define $a_n = b_n = \lambda_n^{-1/2} f_n$, $n \geq 1$, where $f_n(x) = \int_a^x \psi_n(t)dt$. Clearly L, M in section 5.1, and K are the same kernels and

$$\begin{aligned}
\int_a^b L(x,s)da_k(s) &= \int_a^b L(x,s)\lambda_k^{-1/2}\psi_k(s)ds \\
&= \sum_{n\geq 1} \lambda_n\psi_n(x) \int_a^b \psi_n(s)\lambda_k^{-1/2}\psi_k(s)ds \\
&= \lambda_k^{1/2}\psi_k(x),
\end{aligned}$$

and similarly for b_k. The canonical correlations are $\rho_n = 1$, $n \geq 1$; $\{f_n(X)\}$, $\{f_n(Y)\}$ are canonical variables, and (5.2) reduces to (5.1). We have proved the following result.

Proposition 5.3.1 *The dimension of* $H^+(x,y) = \min\{F(x), F(y)\}$ *is* $\varkappa_0$.

5.3.2 Regression family

If the ranges of X, Y are the intervals $[a,b], [c,d]$, and $\varphi : [a,b] \to [c,d]$ is an increasing function, the family

$$H_\theta(x,y) = \theta F(\min\{x, \varphi^{-1}(y)\}) + (1-\theta)F(x)\,J_\theta(y), \quad 0 \le \theta < \theta^+, \tag{5.6}$$

is a bivariate cdf with marginals F, G, provided that

$$J_\theta(y) = (G(y) - \theta F(\varphi^{-1}(y)))/(1-\theta)$$

is a cdf.

The simplest version of this family appears with $\varphi = G^{-1} \circ F$:

$$R_\theta(x,y) = \theta \min\{F(x), G(y)\} + (1-\theta)F(x)G(y), \quad 0 \le \theta \le 1.$$

For this family, and assuming $F = G$

$$\begin{aligned} Cov_{R_\theta}(X,Y) &= \textstyle\int_a^b \int_c^d (\theta \min\{F(x), G(y)\} + (1-\theta)F(x)G(y) - F(x)G(y))dxdy \\ &= \textstyle\int_a^b \int_c^d \theta(\min\{F(x), G(y)\} - F(x)G(y))dxdy \\ &= \theta Var(X). \end{aligned}$$

Thus the correlation is $\rho(X,Y) = \theta$.

Considering the corresponding copula

$$C_\theta(u,v) = \theta \min\{u,v\} + (1-\theta)uv, \quad 0 \le \theta \le 1,$$

we have

$$\begin{aligned} Cov(\alpha(U), \beta(V)) &= \theta J_{\alpha\beta}, \\ Cov(\alpha(U), \alpha(V)) &= \theta J_{\alpha\alpha}, \\ Cov(\beta(U), \beta(V)) &= \theta J_{\beta\beta}, \end{aligned}$$

where

$$J_{\gamma\delta} = \int_0^1 \int_0^1 (\min\{u,v\} - uv)d\gamma(u)d\delta(v).$$

Thus the affinity is given by

$$A_{C_\theta}(\alpha,\beta) = \frac{J_{\alpha\beta}}{\sqrt{J_{\alpha\alpha}J_{\beta\beta}}}, \tag{5.7}$$

which does not depend on $\theta \neq 0$. Then $A_{C_\theta}(\alpha,\beta) = A_{C^+}(\alpha,\beta)$. In other words, the affinity only depends on the upper Fréchet bound $C^+(u,v)$.

Proposition 5.3.2 *The regression family has the same dimension $\varkappa_0$ as the Fréchet upper bound.*

PROOF. We have proved that the upper Fréchet bound has a countable dimension. An alternative proof is as follows. Let us consider the complete orthogonal system $\{\sqrt{2}\sin(n\pi x),\ n \geq 1\}$ on $L^2([0,1])$. This is the set of eigenfunctions of C^+ (see [99]), and

$$\langle \sin(m\pi U), \sin(n\pi V)\rangle_{C^+} = mn\pi^2 \int_0^1 \int_0^1 (\min\{u,v\} - uv)\cos(m\pi u)\cos(n\pi v)dudv,$$

which is 0 if $m \neq n$, and $\neq 0$ otherwise. Define $\Phi_H = \{\sin(n\pi x)\}_{n\geq 1}$. We have that $A_{C_\theta}(\sin(m\pi U), \sin(n\pi V)) = \delta_{m,n}$. Since Φ_H is complete, there is no function $\alpha \neq 0$ orthogonal to all functions in Φ_H (see [54]). Hence, if $\alpha, \beta \in \Phi_H^{\perp}$, then $\alpha \equiv \beta \equiv 0$ and $\langle \alpha, \beta\rangle_{C^+} = 0$. □

5.3.3 AMH family

The Ali-Mikhail-Haq distribution (1978, see [1]) is defined as

$$H_\theta(x,y) = F(x)G(y)/[1 - \theta(1 - F(x))(1 - G(y))], \quad -1 \leq \theta \leq 1.$$

The corresponding copula is

$$C_\theta(u,v) = uv/[1 - \theta(1-u)(1-v)], \quad -1 \leq \theta \leq 1.$$

We can express this copula as

$$C_\theta(u,v) = uv + \theta u(1-u)v(1-v) + \sum_{k=2}^{\infty} \theta^k u(1-u)^k v(1-v)^k, \quad -1 \leq \theta \leq 1. \quad (5.8)$$

The first term of this diagonal expansion is the FGM copula. However note that

$$\int_0^1 u(1-u)^k u(1-u)^{k'} du = B(3, k+k'+1) \neq 0,$$

where $B(.,.)$ is the beta function, so we do not obtain for $C_\theta(u,v) - uv$ an eigenexpansion, as these functions are not orthogonal.

Define the functional

$$Z(k,\phi) = \int_0^1 (1-u)^k (1 + (k-1)u)\phi(u)du. \quad (5.9)$$

Assuming $[u(1-u)^{k+1}\phi(u)]_0^1 = 0$, integration by parts gives

$$Z(k, \phi) = -\int_0^1 u(1-u)^{k+1} d\phi(u).$$

Therefore

$$\begin{aligned} Cov(\alpha(U), \alpha(V)) &= \textstyle\sum_{k=1}^{\infty} \theta^k Z(k-1, \alpha)^2, \\ Cov(\alpha(U), \beta(V)) &= \textstyle\sum_{k=1}^{\infty} \theta^k Z(k-1, \alpha) Z(k-1, \beta). \end{aligned}$$

Thus the affinity is given by

$$A_{C_\theta}(\alpha, \beta) = \frac{\sum_{k=1}^{\infty} \theta^k Z(k-1, \alpha) Z(k-1, \beta)}{\sqrt{\sum_{k=1}^{\infty} \theta^k Z(k-1, \alpha)^2} \sqrt{\sum_{k=1}^{\infty} \theta^k Z(k-1, \beta)^2}}. \tag{5.10}$$

Proposition 5.3.3 *The AMH family has dimension $\varkappa_0$*

PROOF. $\langle \phi, \phi \rangle = 0$ implies

$$\sum_{k=1}^{\infty} \theta^k Z(k-1, \phi)^2 = 0.$$

If $\theta \neq 0$, this is only possible for $Z(k-1, \phi) = 0$, $k \geq 1$. But the system $\{(1-u)^k(1+(k-1)u)\}$ is complete in $L^2([0,1])$, so $\phi \equiv c$ (a constant). Take $\Phi_H = \{(1-u)^k(1+(k-1)u)\}_{k \geq 1}$ and the result is proved. □

5.4 Continuous dimension: Cuadras-Augé family

5.4.1 Definition

The Cuadras-Augé bivariate distribution, 1981 (see [22]) is defined by

$$H_\theta(x, y) = \min\{F(x), G(y)\}^\theta (F(x)G(y))^{1-\theta}, \quad 0 \leq \theta \leq 1.$$

The copula is

$$C_\theta(u, v) = \min\{u, v\}^\theta (uv)^{1-\theta}, \quad 0 \leq \theta \leq 1.$$

If $\mathcal{H}_1$ is the Heaviside distribution

$$\mathcal{H}_1(x) = 0 \quad \text{if } x < 1, \;\; \mathcal{H}_1(x) = 1 \quad \text{if } x \geq 1,$$

it can be proved (see [18]) that

$$Cor(\mathcal{H}_1(X), \mathcal{H}_1(Y)) = \max_{\varphi} Cor(\varphi(X), \varphi(Y)).$$

Cuadras, 2004 ([21]) generalizes this result by finding the canonical correlations for this copula. We present the main results of this ongoing research, without proofs.

5.4.2 Eigenanalysis

Let us consider the covariance kernels

$$K_\theta(u,v) = \min\{u,v\}^\theta (uv)^{1-\theta} - uv, \quad L(u,v) = \min\{u,v\} - uv.$$

Given two functions $\phi_1, \phi_2 : I = [0,1] \to \mathbb{R}$, the squared correlation between $\phi_1(U), \phi_2(V)$ is

$$\begin{aligned} \rho^2 &= \frac{(Cov(\phi_1(U),\phi_2(V)))^2}{Var(\phi_1(U))Var(\phi_2(V))} \\ &= \frac{(\int_{I^2} K_\theta(u,v)d\phi_1(u)d\phi_2(v))^2}{\int_{I^2} L(u,v)d\phi_1(u)d\phi_1(v)\int_{I^2} L(u,v)d\phi_2(u)d\phi_2(v)}. \end{aligned}$$

A function $\phi : I \to \mathbb{R}$ is an eigenfunction of K_θ with respect to L with eigenvalue λ if

$$\int_0^1 K_\theta(u,v)d\phi(v) = \lambda \int_0^1 L(u,v)d\phi(v).$$

Let us define

$$H_{\gamma,\varepsilon}(x) = \mathcal{H}_{\gamma^-}(x) - \frac{\gamma}{\gamma+\varepsilon}\mathcal{H}_{(\gamma+\varepsilon)^+}(x),$$

where

$$\begin{aligned} &\mathcal{H}_{\gamma^-}(x) = 0 \quad \text{if } x < \gamma, \ \mathcal{H}_{\gamma^-}(x) = 1 \quad \text{if } x \geq \gamma, \\ &\mathcal{H}_{\gamma^+}(x) = 0 \quad \text{if } x \leq \gamma, \ \mathcal{H}_{\gamma^+}(x) = 1 \quad \text{if } x > \gamma. \end{aligned}$$

Theorem 5.4.1 *The set $(\phi_\gamma, \lambda_\gamma)$ of eigenfunctions and eigenvalues of K_θ with respect to L is given by*

$$\phi_\gamma = \lim_{\varepsilon\to 0} H_{\gamma,\varepsilon}, \qquad \lambda_\gamma = \theta\gamma^{1-\theta},$$

where ϕ_γ is the indicator of γ, i.e., $\phi_\gamma(x) = 0$ if $x \neq \gamma$, and $\phi_\gamma(\gamma) = 1$, for $\gamma \in [0,1]$.

Proof. See [21]. □

The definition of the canonical correlations (see Section 1.1 in Chapter 1) is adapted to the continuous case. The following result is derived:

Theorem 5.4.2 *The set $(\phi_\gamma, \lambda_\gamma)$, $\gamma \in [0,1]$, is the set of canonical functions and canonical correlations.*

PROOF. See [21]. □

Proposition 5.4.3 *The dimension of the Cuadras-Augé copula is* $\varkappa_1$.

PROOF. The cardinality depends on the functions ϕ_γ where $\gamma \in [0,1]$. Then

$$Cov(\phi_{\gamma_1}(U), \phi_{\gamma_2}(V)) = 0 \quad \text{if } \gamma_1 \neq \gamma_2.$$

However, it can be proved that $Cov(H_{\gamma,\varepsilon}(U), H_{\gamma,\varepsilon}(V)) \neq 0$. Clearly, the affinity $A_H^2(H_{\gamma,\varepsilon}, H_{\gamma,\varepsilon})$ is

$$\frac{Cov^2(H_{\gamma,\varepsilon}(U), H_{\gamma,\varepsilon}(V))}{Cov(H_{\gamma,\varepsilon}(U), H_{\gamma,\varepsilon}(V))Cov(H_{\gamma,\varepsilon}(U), H_{\gamma,\varepsilon}(V))},$$

whose limit is $A_H^2(\phi_\gamma, \phi_\gamma) = 1$. □

5.5 Generalized Cuadras-Augé family

Let C_0, C_1 be two continuous copulas. A generalization of the above family is

$$C_\theta(u,v) = C_1(u,v)^\theta C_0(u,v)^{1-\theta}, \quad \theta, u, v \in [0,1].$$

Clearly C_θ reduces to C_0 for $\theta = 0$ and to C_1 for $\theta = 1$. The above Cuadras-Augé copula is reached for $C_0(u,v) = uv$ and $C_1(u,v) = \min\{u,v\}$.

Let us study the dimension of C_θ when it is constructed with the FGM and the independence copulas.

Proposition 5.5.1 *If* $\theta, \alpha \neq 0$*, the dimension of* C_θ *with* C_1 *the FGM copula and* C_0 *the independence copula, is* $\varkappa_0$.

PROOF. The copula is

$$\begin{aligned} C_\theta(u,v) &= (uv[1+\alpha(1-u)(1-v)])^\theta \, [uv]^{1-\theta} \\ &= [1+\alpha(1-u)(1-v)]^\theta uv, \end{aligned}$$

where $\theta, u, v \in [0,1]$, $-1 \leq \alpha \leq 1$.

The Taylor series

$$(1+x)^\theta = 1 + \theta x + \frac{\theta(\theta-1)}{2!}x^2 + \ldots$$

gives the following expression for the copula C_θ:

$$\begin{aligned} C_\theta(u,v) &= uv + \theta\alpha u(1-u)v(1-v) + \frac{\theta(\theta-1)}{2!}\alpha^2 u(1-u)^2 v(1-v)^2 + \dots \\ &= uv + \theta\alpha u(1-u)v(1-v) + \sum_{k=2}^{\infty} Q_{k,\theta}\alpha^k u(1-u)^k v(1-v)^k, \end{aligned}$$

where

$$Q_{k,\theta} = \frac{\theta(\theta-1)\cdots(\theta-k+1)}{k!}.$$

Compare this expression with (5.8), the AMH copula. Both expressions are formally similar and we can adapt to this family the proof given for the AMH. □

5.6 Examples of H-affinities

We conclude presenting H-affinities for some functions and distributions. It is worth noting that the affinity, in some cases, is equal to 0 or 1, independently of the functions α, β.

Functions	Bivariate cdf	Affinity	Correlation	Name
x, y	$H \neq FG$	1	ρ	Pearson
α, β	FG	0	0	Pearson
$F = G$	H	1	ρ_S	Spearman
$F = G$	FGM	1	$\theta/3$	Spearman
α, β $(I_\alpha = 0)$	FGM	0	0	Pearson
α, β $(I_\alpha, I_\beta \neq 0)$	FGM	1	(5.4)	Pearson
x, y	Regression	1	θ	Pearson
α, β	Regression	(5.7)	θ	Pearson
α, β	Upper bound	(5.7)	1	Pearson
α, β	AMH	(5.10)	(*)	Pearson
x, y	Cuadras-Augé	1	$2\theta/(3-\theta)$	Pearson
$\mathcal{H}_1, \mathcal{H}_1$	Cuadras-Augé	1	θ	Max. correl.

Notes: 1) We always suppose $\theta \neq 0$. 2) (*) not in closed form.

Chapter 6

Construction of diagonal distributions via principal components

Let X, Y be two random variables with range $[a,b]\,,[c,d]$, absolutely continuous bivariate cdf H and marginals F, G. Let h, f, g be their densities, respectively. Let $\{a_i\}\,,\{b_j\}$ be complete sets of orthonormal functions defined on the marginal distributions $F\,(x)$ and $G\,(y)$, respectively, by

$$\int_a^b a_i(x)a_j(x)dF\,(x) = \int_c^d b_i(y)b_j(y)dG\,(y) = \delta_{ij},$$

δ_{ij} being Kronecker's delta. Let

$$\rho_{ij} = Cor\,(a_i\,(X)\,,b_j\,(Y)) = \int_a^b\int_c^d a_i(x)b_j(y)dH\,(x,y)$$

be Pearson's correlation coefficient.

Lancaster ([71], [72]) showed that, if $\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\rho_{ij}^2 < \infty$, then

$$dH\,(x,y) = dF\,(x)\,dG\,(y)\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\rho_{ij}a_i(x)b_j(y),$$

where double summation is convergent in the mean square sense. It is also worth noting that Pearson's contingency coefficient ϕ^2, defined by

$$\phi^2 + 1 = \int_a^b\int_c^d (dH\,(x,y))^2\,/\,(dF\,(x)\,dG\,(y))\,,$$

satisfies

$$\phi^2 = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \rho_{ij}^2.$$

We say that H is ϕ^2-bounded if $\phi^2 < \infty$. This condition is satisfied by most of the distributions used in the applications.

The sets $\{a_i\}, \{b_j\}$ can be chosen so that the correlation matrix $\mathbf{R} = (\rho_{ij}), \, i, j \geq 1$ is diagonal. Lancaster used the canonical variables, i.e., two sets $\{a_i\}, \{b_j\}$, of orthonormal functions defined on the marginal distributions in a recursive manner such that the correlation between corresponding members of the two sets is maximal, given the preceding canonical variables. The $\rho_i \equiv \rho_{ii}$ are the canonical correlations and can be assumed positive. These sets satisfy the so-called biorthogonal property,

$$E[a_i(X) b_j(Y)] = \delta_{ij} \rho_i,$$

and H can be expanded as

$$dH(x, y) = dF(x) \, dG(y) \sum_{i=0}^{\infty} \rho_i a_i(x) b_j(y), \tag{6.1}$$

the diagonal expansion of H. Lancaster proves in [72] that $a_i(X)$ is also orthogonal to every square summable function of Y, orthogonal to the canonical variables $b_j(Y)$ and, similarly, $b_j(Y)$ with respect to X. Hence, orthonormal polynomials with respect to the marginals have been used as an extension of this method (see [34], and Hutchinson & Lai [54], Chapter 14, for a general description).

Notice that expansion (6.1) is equivalent to

$$h(x, y) = f(x) \, g(y) \left[1 + \sum_{i=1}^{\infty} \rho_i a_i(x) b_i(y)\right], \tag{6.2}$$

since 1 can be understood as a member of zero$-th$ order.

Cuadras and Fortiana (see [26]) showed that this expansion can be seen as a particular instance of continuous weighted scaling. Cuadras (see [17]) proved that this diagonal expansion can be expressed in terms of the cdf's and introduced another extension of this method consisting in using the principal directions of each marginal variable as orthogonal sets of functions (see [16]). Principal components (or directions) were obtained for X uniform on $[0, 1]$, X exponential, X logistic and, finally, when X is Pareto (α, θ), $\alpha > 2$, $\theta \geq 0$. The following section introduces some definitions and results (see, for instance, [25]) related to the method of obtaining these principal directions.

6.1 Principal components of random variables

Let X be a r.v. on a probability space $(\Omega, A, \mathcal{P})$, with values on an interval $I = [a, b]$, absolutely continuous cdf F and probability density f, with respect to the Lebesgue measure. Let $\delta : I \times I \to \mathbb{R}_+$ be a distance function. A continuous Euclidean configuration representing X with respect to δ is defined as a stochastic process $\mathbf{X} = \{X_t\}_{t \in I}$ such that for all $\omega_1, \omega_2 \in \Omega$, the Euclidean distance between $X_t(\omega_1), X_t(\omega_2)$ defined as

$$D_E(\omega_1, \omega_2) \equiv \left(\int_I (X_t(\omega_1) - X_t(\omega_2))^2 \, dt \right)^{\frac{1}{2}}$$

equals $\delta(X(\omega_1), X(\omega_2))$.

Consider the function $u : I \times I \to [0, 1]$ defined by

$$u(t, x) = \begin{cases} 1 \text{ if } t < x, \\ 0 \text{ if } t \geq x. \end{cases}$$

Let $\mathbf{X} = \{X_t\}_{t \in I}$ be defined as $X_t = u(t, x)$, for $t \in I$. The stochastic process $\mathbf{X}$ is a continuous Euclidean representation of X, and the following results hold:

1. The covariance function of $\mathbf{X}$ is given by

$$K(s, t) = \min\{F(s), F(t)\} - F(s)F(t), \quad s, t \in I. \tag{6.3}$$

2. The trace of this kernel $\int_I K(s, s) \, ds$ equals the geometric variability V_δ of X with respect to the distance function $\delta = \sqrt{|x - y|}$, defined by

$$V_\delta(X) = \frac{1}{2} \int_{\mathbb{R}^2} \delta^2(s, t) \, dF(s) \, dF(t),$$

   provided that this integral exists.

3. If $E[X] < \infty$, $\lim_{s \to -\infty} sF(s) = 0$, and $V_\delta(X) < +\infty$, from Mercer's theorem, the expansion

$$K(s, t) = \sum_{i=1}^{\infty} \lambda_i \varphi_i(s) \varphi_i(t)$$

   is absolutely and uniformly convergent in both s and t (see, for instance, [11], vol I), where $\{\varphi_i\}_{i \in \mathbb{N}}$ is a complete orthonormal set (over $\mathcal{L}^2(I)$) of solutions of

$$\int_I \varphi_i(s) K(s, t) \, ds = \lambda_i \varphi_i(t).$$

Finally, from the theorem of Kac and Siegert (see [61]), the following decomposition is obtained:

$$X_t = \sum_{j=1}^{\infty} Z_j \varphi_j (t),$$

where $\{Z_j\}_{j \in \mathbb{N}}$ is an orthogonal set (with respect to the covariance) of square integrable r.v.'s defined by

$$Z_j = \int_I X_t \varphi_j (t) \, dt, \quad j \in \mathbb{N},$$

satisfying $Var(Z_j) = \lambda_j$.

Each Z_j is called a principal component of $\mathbf{X}$. The following theorem (Theorem 1 in [25]) shows how to obtain these components.

Theorem 6.1.1 *Let φ_j be an eigenfunction of K with eigenvalue λ_j, and consider the function*

$$h_j (s) = \int_a^s \varphi_j (t) \, dt, \quad s \in (a, b).$$

Then:

1. *The principal component Z_j corresponding to λ_j is given by $Z_j = h_j(X)$.*
2. *$\mu_j = E[Z_j] = \int_I [1 - F(t)] \varphi_j (t) \, dt$.*
3. *h_j is a solution of the differential equation*

$$\lambda_j h_j'' + (h_j - \mu_j) f = 0, \qquad h_j (a) = 0, \quad h_j' (a) = 0,$$

where f is the density of X.

PROOF. See [25]. □

Some principal components Z_j have been obtained using this method by Cuadras and Fortiana, 1995 [25], Cuadras and Lahlou, 2000 [30], and Cuadras and Lahlou, 2002 [31].

Examples of eigenfunctions φ_n and eigenvalues $\lambda_n = Var(Z_n) = Var(h_n(X))$ (see [25], [30]) are:

1. $\varphi_n(x) = \sqrt{2} \sin (n\pi x)$, $\lambda_n = 1/(n\pi)^2$, if X is $[0, 1]$ uniform.

2. $\varphi_n(x) = \exp(-x/2)\left[J_1(\xi_n \exp(-x/2))\right]/J_0(\xi_n)$, $\lambda_n = 4/\xi_n^2$, if X is exponential with unit mean, where ξ_n is the $n-th$ positive root of J_1 and J_0, J_1 are the Bessel functions of the first kind, of order 0 and 1, respectively.

3. $\varphi_n(x) = \sqrt{\frac{2n+1}{n(n+1)}} L_n'\left(F\left(x\right)\right) f\left(x\right)$, $\lambda_n = 1/(n(n+1))$, if X is standard logistic, where L_n' are the first derivatives of the shifted Legendre polynomials on $[0,1]$, $F(x) = 1/(1+\exp(-x))$ is the cdf of X and $f\left(x\right) = F'\left(x\right)$

4. $\varphi_n(x) = a_n\left(\sin\left(\eta_n/x\right) - \cos\left(\eta_n/x\right)/x\right)$, $\lambda_n = 3/\eta_n^2$, if X is Pareto with $F(x) = 1-x^{-3}$, $x > 1$, where $a_n = 2\eta_n^{-\frac{1}{2}}\left(2\eta_n - \sin\left(2\eta_n\right)\right)^{-\frac{1}{2}}$ and η_n is the $n-th$ positive root of equation $x = \tan(x)$.

6.2 Diagonal expansions via principal components

Let $\{f_j(X)\}, \{g_j(Y)\}$ be the sets of principal components of two r.v.'s X and Y respectively. The sets of standardized principal components $\{F_j(X)\}$, $\{G_j(Y)\}$

$$F_j(X) = \frac{f_j(X) - E\left[f_j(X)\right]}{\sqrt{Var\left(f_j(X)\right)}},$$

and similarly, G_j , are orthonormal sets of square integrable functions. Then F_j and G_j may play the role of canonical functions a_j, b_j in expansion (6.2). Thus

$$h(x,y) = f(x)\,g(y)\left[1 + \sum_{j=1}^{\infty} \rho_j F_j(x)\,G_j(y)\right] \tag{6.4}$$

represents a bivariate distribution with marginal densities f, g ([16], [17]). The infinite dimensional vector of correlation coefficients $(\rho_1, \rho_2, ...)$ can be chosen to construct different distributions with this representation. The dimension of h is the number of non-null correlation coeficients (see Proposition 5.1.2 in Chapter 5). For instance, by taking $\rho_N > \rho_{N+1} = 0$, we have that (6.4) is the representation of a nested $N-$parametric diagonal family of distributions. Hence, this representation provides a method to construct bivariate distributions with given marginals via principal components. When $N = 1$ this family reduces to the Sarmanov family (see [76]).

By construction, the following properties hold:

1. $\{F_j(X)\}$, $\{G_j(Y)\}$ are sequences of centered and uncorrelated random variables.

2. $Cor\left(F_i(x), G_j(y)\right) = \rho_j \delta_{ij}$.

3. $\{\rho_j\}_j$ is the sequence of canonical correlations.

4. $\{F_j(X), G_j(Y)\}$ is the sequence of canonical variables.

Remark 6.2.1 *Expansion (6.4) allows us to obtain the density $c(u,v)$ of the corresponding copula. Since*

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

it suffices to apply the change of variables $F(x) = u$, $G(y) = v$. Thus:

$$c(u,v) = 1 + \sum_{j=1}^{\infty} \rho_j F_j\left(F^{-1}(u)\right) G_j\left(G^{-1}(v)\right).$$

6.3 Stochastic independence and association

6.3.1 Characterization of Independence

X, Y are independent r.v.'s iff the law of (X,Y) is equal to the product of the marginal laws. Thus

$$H(x,y) = F(x)G(y),$$

where H and F, G are the joint and marginal cdf's, respectively. If X, Y are independent and integrable, then

$$E(XY) = E(X)\, E(Y).$$

This condition is necessary but not sufficient for stochastic independence.

A stronger criteria was put forward in [19]: let $\{f_j(X)\}, \{g_j(Y)\}$ be the principal components of X, Y respectively. Then X, Y are stochastically independent iff all the correlations between the principal components are zero:

$$Cor\left(f_m(X), g_n(Y)\right) = 0 \qquad m, n > 0.$$

As a consequence, if the density of (X,Y) can be expanded as (6.4), X and Y are independent iff

$$Cov\left(F_m(X), G_n(Y)\right) = 0, \qquad m, n \geq 1.$$

Furthermore canonical correlations $\rho_{mn} = 0$, $\forall\ m, n \geq 1$ and, hence, $h(x,y) = f(x)\, g(y)$.

In particular, if the marginals are uniform on $[0,1]$, then $c(u,v) = 1$. These characterizations of independence also hold when expansions are diagonal (i.e., $F_n = G_n$, $\rho_{mn} = \rho_n \delta_{mn}$).

6.3.2 One expansion for the correlation

Cuadras, 2002 ([18]) proved that the covariance between two functions can be expanded by using diagonal expansions. In terms of the cdf's, if $Cov(\alpha(X), \beta(Y))$ and expansion (6.1) exist, and we can integrate termwise, then

$$Cov(\alpha(X), \beta(Y)) = \sum_{i=1}^{\infty} \rho_i Cov(a_i(X), \alpha(X)) \, Cov(b_i(Y), \beta(Y)), \tag{6.5}$$

where $\{a_i\}, \{b_i\}$ are canonical variables and $\{\rho_i\}$ canonical correlations (see Theorem 2 in [18]). Since the canonical variables satisfy $E[a_i(X)] = E[b_i(Y)] = 0$, $Var[a_i(X)] = Var[b_i(y)] = 1$, (6.5) can be written in terms of the correlations as

$$Cor(\alpha(X), \beta(Y)) = \sum_{i=1}^{\infty} \rho_i Cor(a_i(X), \alpha(X)) \, Cor(b_i(Y), \beta(Y)).$$

If $\alpha = F$, $\beta = G$, then this correlation is Spearman's correlation coefficient ρ_S, which can be expanded as

$$\rho_S(X,Y) = \sum_{i=1}^{\infty} \rho_i Cor(a_i(X), F(X)) \, Cor(b_i(Y), G(Y)). \tag{6.6}$$

In these expansions, we can replace the canonical variables by the principal components of the marginals X and Y, to obtain expansions for suitable diagonal families, as defined above. Moreover, if the principal components are linearly related with the marginal distributions, the computation of ρ_S is very simple from expansion (6.6), since the correlation is invariant under linear transformations.

Chapter 7

Construction of some specific diagonal distributions

7.1 Preliminaries

We apply the method described in the previous chapter (see Section 6.2) to build families of distributions when the margins are fixed and the principal components related to the marginals are known. We can construct bivariate distributions with marginals being uniform, exponential, logistic, and Pareto by using the standardized principal components $F_j(X)$ in expansion (6.4),

$$h(x,y) = f(x)\,g(y)\left[1 + \sum_{j=1}^{\infty} \rho_j F_j(x)\,G_j(y)\right].$$

We denote $h \in \mathcal{F}(f,g)$ when the joint density h has marginals f,g. Next on we assume that either all the canonical correlations are positive (in this case, they constitute a monotone decreasing sequence of positive real numbers), or just a finite number of them are positive. We study the conditions that the constructed $h(x,y)$ must satisfy to be a probability density, the upper bound for the first canonical correlation, and finally, the maximum correlation for each family. A necessary, and also sufficient condition is found for $N=1$. Otherwise, necessary conditions are given.

By construction, the canonical correlations ρ_j are the correlations between the $j-th$ standardized principal components, which play the role of canonical variates.

This gives a diagonal expansion

$$h(x,y) = f(x)\,g(y)\left[1 + \sum_{i=1}^{\infty} \rho_i a_i(x)\, b_i(y)\right],$$

as it was stated by Lancaster (see [71]). See also Cuadras ([25]), and Chapter 6.

We start with the following general result.

Theorem 7.1.1 *Let X, Y be two identically distributed (i.d.) absolutely continuous r.v.'s with marginal cdf F. With the change of variables $F(x) = u$, $F(y) = v$, let us write the diagonal family as the density of the corresponding copula*

$$c(u,v) = 1 + \sum_{j=1}^{N} \rho_j F_j^*(u)\, F_j^*(v),$$

where $F_j^*(\cdot)$ stands for $F_j \circ F^{-1}(\cdot)$. Then:

1. *$c(u,v)$, as well as the corresponding $h(x,y)$, is a density if and only if (iff)*

$$\sup_{(u,v)\in[0,1]^2}\left[-\sum_{j=1}^{N} \rho_j F_j^*(u)\, F_j^*(v)\right] \leq 1. \tag{7.1}$$

2. *The canonical correlations must satisfy the necessary condition*

$$-\sum_{j=1}^{N} \rho_j F_j^*(u)\, F_j^*(1) \leq 1, \tag{7.2}$$

   uniformly in $u \in [0,1]$.

3. *Pearson's correlation coefficient is given by*

$$Cor(X,Y) = \left(\sum_{j=1}^{N} \rho_j I_j^2\right) / Var(X),$$

   where

$$I_j = \int_0^1 F^{-1}(u)\, F_j^*(u)\, du. \tag{7.3}$$

Proof. 1) follows directly from the positivity of $c(u,v)$ for all pairs $(u,v) \in [0,1]^2$. If $c(u,v) \geq 0$ then it is Riemann-integrable. Therefore, integration with respect to each variable gives the uniform density, and integration with respect to both variables over the unit square gives 1.
2) follows from 1) taking $v = 1$.

Next, as X, Y are i.d. r.v.'s, $Cor(X,Y) = Cov(X,Y)/Var(X)$. From (6.4), and by Fubini's theorem, the covariance is given by

$$\begin{aligned} Cov(X,Y) &= \int_{\mathbb{R}^2} xy\ h(x,y)\,dxdy - \int_{\mathbb{R}} xf(x)\,dx \int_{\mathbb{R}} yf(y)\,dy \\ &= \int_{\mathbb{R}^2} xyf(x)\,f(y)\left(\sum_{j=1}^{N} \rho_j F_j(x)\,F_j(y)\right) dxdy \\ &= \sum_{j=1}^{N} \rho_j \int_{\mathbb{R}} xF_j(x)\,f(x)\,dx \int_{\mathbb{R}} yF_j(y)\,f(y)\,dy. \end{aligned}$$

Since $dF(x) = f(x)\,dx$, the change of variables $F(x) = u$ proves 3). □

In the following sections we apply this general result to some specific distributions. If $N = 1$ then the supremum of $\{-\rho_1 F_1^*(u)\,F_1^*(v)\}$ is attained at $u = 0, v = 1$. Further research is needed to obtain the pair (u,v) for which supremum is attained if $N > 1$. With the choice of $v = 1$ we obtain a necessary (but possibly non sufficient) condition for the canonical correlations.

7.2 Diagonal family with uniform marginals

In this section we construct a bivariate family with uniform marginals, i.e., we construct a parametric copula.

Proposition 7.2.1 *Let U, V be uniform on $[0,1]$, i.e., with density $f_U(u) = \mathbf{1}_{[0,1]}(u)$. Let $\{\rho_j\}$ be a decreasing sequence of non-negative real numbers such that $1 \geq \rho_1 > \rho_2 > ... > 0$. Then:*

1. *The function*

$$c(u,v) = 1 + \sum_{j=1}^{N} \rho_j\ 2\cos(j\pi u)\cos(j\pi v), \qquad u, v \in [0,1] \tag{7.4}$$

is the density of a diagonal family with uniform marginals on $[0,1]$ iff

$$\sup_{(u,v)\in[0,1]^2} \left[-2\sum_{j=1}^{N} \rho_j \cos(j\pi u)\cos(j\pi v)\right] \leq 1. \tag{7.5}$$

Each ρ_j is the correlation coefficient between the $j-th$ standardized principal components $F_j(U) = -\sqrt{2}\cos(j\pi U)$, $F_j(V)$,

$$\rho_j = Cor(\cos(j\pi U), \cos(j\pi V)).$$

2. *These coefficients must satisfy the condition*

$$-2\sum_{j=1}^{N}\rho_j \cos(j\pi u)(-1)^j \leq 1, \tag{7.6}$$

uniformly in $u \in [0,1]$.

3. *The correlation coefficient of (U,V) is given by*

$$Cor(U,V) = \frac{96}{\pi^4}\sum_{k=1}^{N}\frac{\rho_{2k-1}}{(2k-1)^4}. \tag{7.7}$$

Proof. For uniform marginals the covariance kernel (6.3) is

$$K(s,t) = \min(s,t) - st, \quad s,t \in [0,1],$$

and the eigenfunctions are $\varphi_j(t) = \sqrt{2}\sin(j\pi t)$. Thus the $j-th$ principal component is

$$\begin{aligned} f_j(U) &= \int_0^U \sqrt{2}\sin(j\pi t)\,dt \\ &= \frac{\sqrt{2}}{j\pi}(1-\cos(j\pi U)). \end{aligned}$$

From $E[f_j(U)] = \frac{\sqrt{2}}{j\pi}$, $Var(f_j(U)) = \frac{1}{(j\pi)^2}$, we obtain the $j-th$ standardized principal component. Substitution of F_j in (6.4) gives (7.4), and we obtain conditions (7.5), (7.6) from direct application of Theorem 7.1.1.
To prove 3) we compute I_j (7.3) as in Theorem 7.1.1:

$$\begin{aligned} I_j &= -\sqrt{2}\int_0^1 u\cos(j\pi u)\,du \\ &= \frac{\sqrt{2}}{j^2\pi^2}\left(1-(-1)^j\right), \end{aligned}$$

which is null when j is even. Writing $j = 2k-1$, only odd terms are non-null, and we have

$$\begin{aligned} Cor(U,V) &= \left(\sum_{k=1}^{N}\rho_{2k-1}I_{2k-1}^2\right)/Var(U), \\ &= 12\sum_{k=1}^{N}\rho_{2k-1}\left(\frac{2\sqrt{2}}{(2k-1)^2\pi^2}\right)^2. \end{aligned}$$

A final simplification gives 3). □

7.2.1 The copula

The copula with density $c(u,v)$ is:

$$\begin{aligned} C(u,v) &= \int_0^u \int_0^v c(s,t)\ dsdt \\ &= uv + \sum_{j=1}^{N} \frac{2\rho_j}{(j\pi)^2} \sin(j\pi u)\sin(j\pi v), \quad (u,v) \in [0,1]^2. \end{aligned}$$

The independence copula $C^0(u,v) = uv$ belongs to this family (take $\rho_j = 0,\ j \geq 1$).

7.2.2 Some finite dimensional densities

Some examples illustrate different cases which appear when considering a finite set of strictly positive canonical correlations.

One-dimensional

If only the first canonical correlation is non-null, (7.4) attains its minimum value at $u = 0, v = 1$, and it is a density iff $\rho_1 \leq \frac{1}{2}$. Then

$$c(u,v) = 1 + \rho_1 2\cos(\pi u)\cos(\pi v), \qquad u,v \in [0,1]$$

is the density of a $1-$parametric copula.

Pearson's correlation coefficient is given by

$$Cor(U,V) = \frac{96}{\pi^4}\rho_1.$$

We can compute the correlation corresponding to the constructed bivariate densities. Moreover, since $\rho_1 \leq \frac{1}{2}$, $Cor(U,V)$ is bounded by $\frac{96}{2\pi^4} = 0.49277$.

Example 7.2.2 *Suppose that* $\rho_1 > \rho_2 = \ldots = 0$. *Figure 7.1 shows this* $1-$*dimensional density for* $\rho_1 = \frac{1}{2}$.

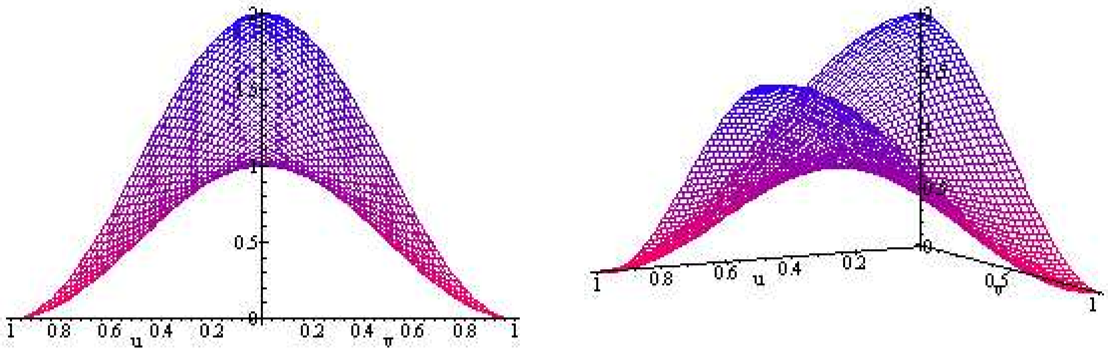

Figure 7.1: Bivariate density with uniform margins: $\rho_1 = \frac{1}{2}$

Two-dimensional

If we consider two non-null canonical correlations, condition (7.5) holds iff

$$-\rho_1 \cos(\pi u) \cos(\pi v) - \rho_2 \cos(2\pi u) \cos(2\pi v) \leq \frac{1}{2},$$

uniformly in $0 \leq u \leq v \leq 1$. The second term, $\rho_2 \cos(2\pi u) \cos(2\pi v)$, reachs extreme values for some pair $(u, v) \neq (0, 1)$, so (7.5) is difficult to verify. The necessary condition (7.6) is equivalent to

$$\rho_1 \cos(\pi u) - \rho_2 \cos(2\pi u) \leq \frac{1}{2},$$

uniformly in $0 \leq u \leq 1$. Since $\cos(2\pi u) = 2\cos^2(\pi u) - 1$, this reduces to

$$-2\rho_2 \cos^2(\pi u) + \rho_1 \cos(\pi u) + \rho_2 - \frac{1}{2} \leq 0,$$

which is a second degree polynomial in the variable $\cos(\pi u)$. If this polynomial takes only negative values then the discriminant of the equation must be either null or negative, i.e.,

$$\rho_1^2 + 8\rho_2 \left(\rho_2 - \frac{1}{2}\right) \leq 0 \Longleftrightarrow \rho_1^2 + \frac{\left(\rho_2 - \frac{1}{4}\right)^2}{1/8} \leq \frac{1}{2}.$$

This condition for $\rho_1 > \rho_2$, is illustrated by Figure 7.2.

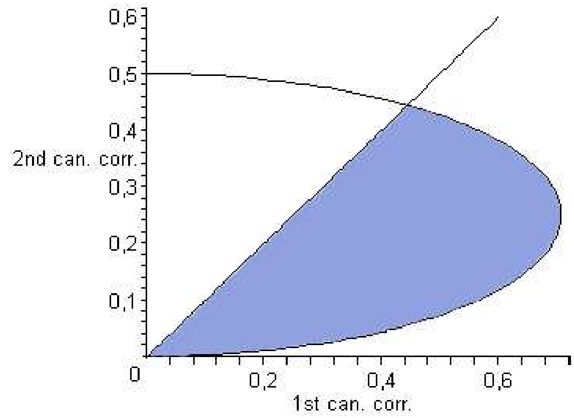

Figure 7.2: First and second canonical correlations for uniform margins

The first correlation attains its maximum at $\rho_1 = \frac{1}{\sqrt{2}}$, when $\rho_2 = \frac{1}{4}$.

Pearson's correlation coefficient is also given by $Cor\,(U, V) = \frac{96}{\pi^4}\rho_1$ since the second term in (7.7) is null. The bound for the correlation is higher:

$$Cor\,(U, V) \leq \frac{96}{\sqrt{2}\pi^4} = 0.69688.$$

Example 7.2.3 *Let $c \in \digamma\,(f_U, f_U)$ be the function given by*

$$c\,(u, v) = 1 + 2\left(\frac{1}{\sqrt{2}}\cos(\pi u)\cos(\pi v) + \frac{1}{4}\cos(2\pi u)\cos(2\pi v)\right), \qquad u, v \in [0, 1]\,.$$

Figure 7.3 shows this density.

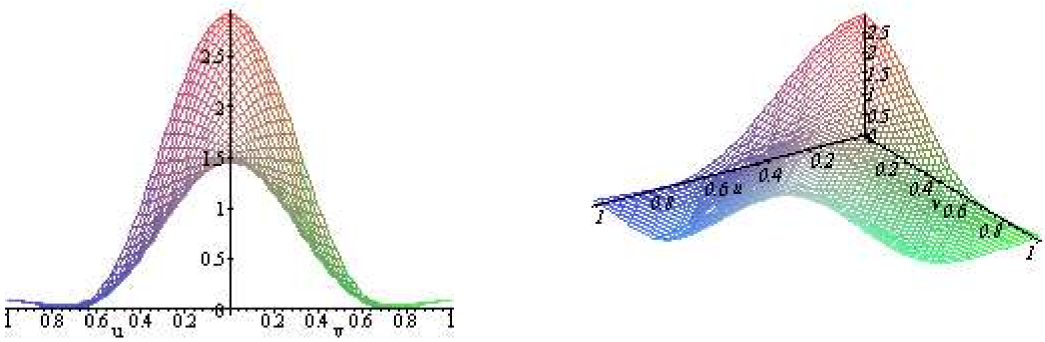

Figure 7.3: Density with $\rho_1 = \frac{1}{\sqrt{2}}, \rho_2 = \frac{1}{4}$

If we take $\rho_1 < \frac{1}{\sqrt{2}}$, but (ρ_1, ρ_2) lays out of the area represented by Figure 7.2, then $c(u,v)$ takes negative values in the unit square. Hence it is not a density. For instance, $c(u,v)$ when $\rho_1 = 0.6$, $\rho_2 = 0.5$ is shown in Figure 7.4:

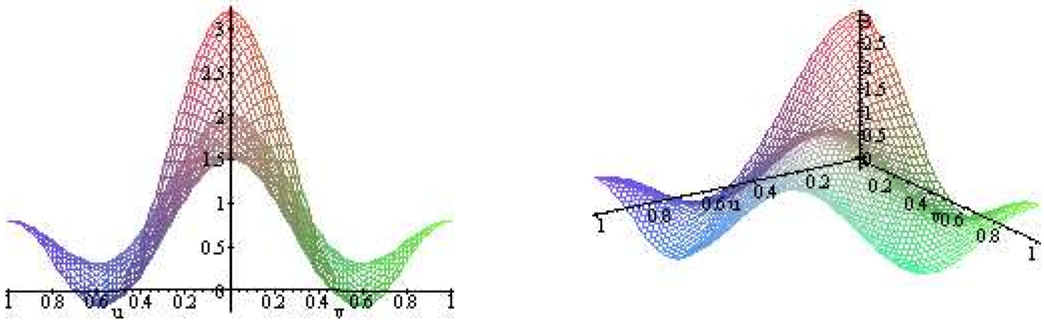

Figure 7.4: Function $c(u,v)$ with $\rho_1 = 0.6, \rho_2 = 0.5$

7.3 Diagonal family with exponential marginals

In this section we construct a bivariate family with exponential marginals.

Proposition 7.3.1 *Let X, Y be exponential with unit mean, i.e., with common density given by $f(x) = \exp(-x)\,\mathbf{1}_{[0,\infty]}(x)$. Let ξ_j be the $j-th$ positive root of J_1, where J_0, J_1 are the Bessel functions of the first kind*[1]*, with $0 \le \rho_j \le 1$. Then:*

1. *The function*

$$h(x,y) = f(x)\, f(y) \left[1 + \sum_{j=1}^{N} \frac{\rho_j}{J_0(\xi_j)^2}\; J_0\left(\xi_j \exp\left(\frac{-x}{2}\right)\right) J_0\left(\xi_j \exp\left(\frac{-y}{2}\right)\right)\right], \tag{7.8}$$

$x, y \ge 0$, is the density of a diagonal family $h \in (f,f)$ iff

$$\sup_{(u,v)\in[0,1]^2} \left[-\sum_{j=1}^{N} \frac{\rho_j}{J_0(\xi_j)^2} J_0\left(\xi_j\sqrt{1-u}\right) J_0\left(\xi_j\sqrt{1-v}\right)\right] \le 1. \tag{7.9}$$

Each ρ_j is the correlation between the $j-th$ standardized principal components $F_j(X) = J_0(\xi_j \exp(-X/2))/J_0(\xi_j)$, $F_j(Y)$,

$$\rho_j = Cor\left(J_0(\xi_j \exp(-X/2)), J_0(\xi_j \exp(-Y/2))\right).$$

[1] The Bessel function of the first kind and $m-th$ order is given by $J_m(x) = \sum_{l=0}^{\infty} \frac{(-1)^l}{(l+m)!\,l!}\left(\frac{x}{2}\right)^{2l}$.

2. *These coefficients must satisfy the condition*

$$-\sum_{j=1}^{N}\rho_j\frac{J_0\left(\xi_j\sqrt{1-u}\right)}{J_0\left(\xi_j\right)^2}\leq 1, \tag{7.10}$$

uniformly in $0\leq u\leq 1$.

3. *The correlation coefficient of* (X,Y) *is given by*

$$Cor\left(X,Y\right)=\sum_{j=1}^{N}\rho_j\left(\frac{4\left(1-J_0\left(\xi_j\right)\right)}{\xi_j^2J_0\left(\xi_j\right)}\right)^2. \tag{7.11}$$

Proof. The proof follows from Theorem 7.1.1. The eigenfunctions of the covariance kernel (6.3) when the distribution is exponential with unit mean are

$$\varphi_j\left(t\right)=\exp\left(-t/2\right)J_1\left(\xi_j\exp\left(-t/2\right)\right)/J_0\left(\xi_j\right).$$

Thus the $j-th$ principal component is

$$\begin{aligned} f_j\left(X\right) &= \int_0^X\exp\left(-t/2\right)J_1\left(\xi_j\exp\left(-t/2\right)\right)/J_0\left(\xi_j\right)\ dt \\ &= 2\left(J_0\left(\xi_j\exp\left(-X/2\right)\right)-J_0\left(\xi_j\right)\right)/\xi_jJ_0\left(\xi_j\right). \end{aligned}$$

Some computations show that $\mu_j=-2/\xi_j$, $\lambda_j=4/\xi_j^2$, giving the $j-th$ standardized principal components. By construction, the correlation between the $j-th$ principal components $F_j(X),F_j(Y)$, is ρ_j.

Substitution of F_j in (6.4) gives (7.8). Of course if h is non-negative, then $h\in\mathcal{F}\left(f,f\right)$. Since $f\left(x\right)$ is a density, non-negativity of $h\left(x,y\right)$ requires that

$$1+\sum_{j=1}^{\infty}\frac{\rho_j}{J_0\left(\xi_j\right)^2}\ J_0\left(\xi_j\exp\left(-x/2\right)\right)J_0\left(\xi_j\exp\left(-y/2\right)\right)\geq 0$$

for each $x,y\geq 0$. The change of variables $F\left(x\right)=u,F\left(y\right)=v$ gives the equivalence with condition (7.9). Take $v=1$; since $J_0\left(0\right)=1$ (see Figure 7.5 below), condition (7.10) follows.

Finally, since the variances are 1,

$$Cor\left(X,Y\right)=\sum_{j=1}^{N}\rho_jI_j^2, \tag{7.12}$$

where

$$\begin{aligned} I_j &= \frac{-1}{J_0\left(\xi_j\right)}\int_0^1\log\left(1-u\right)J_0\left(\xi_j\sqrt{1-u}\right)du \\ &= \frac{-1}{J_0\left(\xi_j\right)}\sum_{l=0}^{\infty}\left[\frac{\left(-1\right)^l}{2^{2l}\left(l!\right)^2}\xi_j^{2l}\int_0^1\left(1-u\right)^l\log\left(1-u\right)du\right] \\ &= \frac{1}{J_0\left(\xi_j\right)}\sum_{l=0}^{\infty}\left[\frac{\left(-1\right)^l}{2^{2l}\left(l!\right)^2}\xi_j^{2l}\left(\frac{1}{l+1}\right)^2\right]. \end{aligned}$$

Taking $k = l + 1$, and adding and substracting 1 (the zero-th term of this series),

$$\begin{aligned} I_j &= \frac{1}{J_0(\xi_j)} \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{2^{2k-2}(k!)^2} \xi_j^{2k-2} \\ &= -\frac{1}{J_0(\xi_j)} \left(\frac{\xi_j}{2}\right)^{-2} \left(\sum_{k=0}^{\infty} \frac{(-1)^k}{2^{2k}(k!)^2} \xi_j^{2k} - 1 \right) \\ &= \left(\frac{1 - J_0(\xi_j)}{J_0(\xi_j)} \right) \left(\frac{2}{\xi_j} \right)^2 . \end{aligned}$$

Substitution of I_j in equation (7.12) completes the proof. □

7.3.1 The cumulative distribution function

The bivariate cdf $H \in \mathcal{F}(F, F)$, $F(x) = 1 - \exp(-x)$, $x \geq 0$, can be obtained via the corresponding copula $C(u, v)$:

$$\begin{aligned} C(u,v) &= \int_0^u \int_0^v c(s,t)\, ds dt \\ &= \int_0^u \int_0^v \left(1 + \sum_{j=1}^{N} \frac{\rho_j}{J_0(\xi_j)^2} J_0\left(\xi_j \sqrt{1-s}\right) J_0\left(\xi_j \sqrt{1-t}\right) \right) ds dt \\ &= uv + \sum_{j=1}^{N} \frac{4\rho_j}{\xi_j^2 J_0(\xi_j)^2} \sqrt{1-u} J_1\left(\xi_j \sqrt{1-u}\right) \sqrt{1-v} J_1\left(\xi_j \sqrt{1-v}\right), \end{aligned}$$

for $0 \leq u, v \leq 1$. Since $\sqrt{1-u} = \sqrt{1 - F(x)} = \exp(-x/2)$,

$$H(x,y) = F(x) F(y) + \sum_{j=1}^{N} \frac{4\rho_j}{\xi_j^2 J_0(\xi_j)^2} \frac{J_1(\xi_j \exp(-x/2))}{\exp(x/2)} \frac{J_1(\xi_j \exp(-y/2))}{\exp(y/2)},$$

for $x, y \geq 0$.

7.3.2 Some finite dimensional densities

The density and correlations when only the first few terms are positive can be evaluated by using the following table showing the five first positive roots ν_j, ξ_j of J_0, J_1, and some approximate values of $J_0(\xi_j)$, $\zeta_j = \left(\frac{1 - J_0(\xi_j)}{J_0(\xi_j)}\right)^2 \left(\frac{2}{\xi_j}\right)^4$.

j	ν_j	ξ_j	$J_0(\xi_j)$	ζ_j
1	2.4048	3.8317	−0.40276	0.9003846
2	5.5201	7.0156	0.30012	0.0359185
3	8.6537	10.1735	−0.24970	0.0374122
4	11.7915	13.3237	0.21836	0.0065056
5	14.9309	16.4706	−0.19647	0.0080629

Table 7.1: First roots of the Bessel function of the first kind and other related values

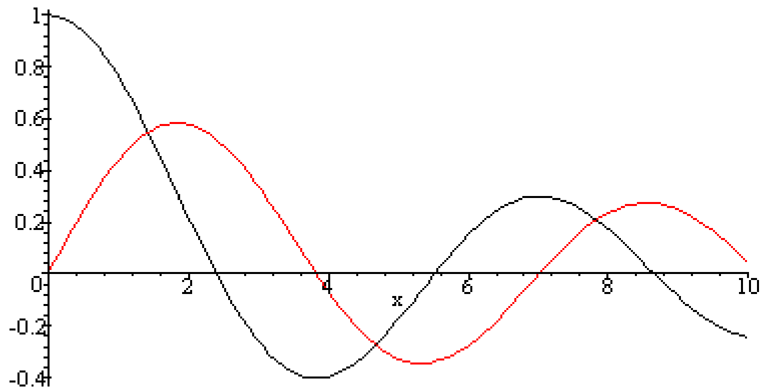

Figure 7.5: Bessel functions of the first kind, J_0 (black line), J_1 (red line)

One-dimensional

When only the first correlation ρ_1 is positive, this family reduces to

$$h(x,y) = f(x)\,f(y)\left[1 + \frac{\rho_1}{J_0(\xi_1)^2} J_0\left(\xi_1 \exp\left(\frac{-x}{2}\right)\right) J_0\left(\xi_1 \exp\left(\frac{-y}{2}\right)\right)\right],$$

for $x, y \geq 0$. The density of the corresponding copula is

$$c(u,v) = 1 + \frac{\rho_1}{J_0(\xi_1)^2} J_0\left(\xi_1\sqrt{1-u}\right) J_0\left(\xi_1\sqrt{1-v}\right), \quad (u,v) \in [0,1]^2.$$

It is easy to check that

$$\sup_{(u,v)\in[0,1]^2}\left[-\frac{\rho_1}{J_0(\xi_1)^2} J_0\left(\xi_1\sqrt{1-u}\right) J_0\left(\xi_1\sqrt{1-v}\right)\right] = \frac{-\rho_1}{J_0(\xi_1)}$$

is attained at $u = 0, v = 1$. Then $c(u,v)$, as well as $h(x,y)$, is a density iff $\rho_1 \leq -J_0(\xi_1) = 0.40276$.

Example 7.3.2 *The function*

$$h(x,y) = f(x) f(y) \left[1 - J_0(\xi_1) J_0\left(\xi_1 \exp\left(\frac{-x}{2}\right)\right) J_0\left(\xi_1 \exp\left(\frac{-y}{2}\right)\right)\right],$$

for $x, y \geq 0$ *is a density* $h \in \mathcal{F}(f, f)$. *The correlation of* (X, Y) *is*

$$Cor(X,Y) = \rho_1 \zeta_1 = 0.36264.$$

(see Table 7.1).

Two-dimensional

Unfortunately, for $N = 2$, conditions (7.9), (7.10) are computationally difficult. In particular, (7.10) is equivalent to

$$-\frac{\rho_1}{J_0(\xi_1)^2} J_0\left(\xi_1\sqrt{1-u}\right) - \frac{\rho_2}{J_0(\xi_2)^2} J_0\left(\xi_2\sqrt{1-u}\right) \leq 1,$$

uniformly in $0 \leq u \leq 1$. Since $J_0\left(\xi_1\sqrt{1-u}\right)$ is monotonic and decreasing in the interval $[0, \xi_1]$, while $J_0\left(\xi_2\sqrt{1-u}\right)$ is not, we have chosen as a candidate for extremum the value attained at $u \in [0,1]$ such that $J_0\left(\xi_2\sqrt{1-u}\right) = 0$, i.e., at $u = 1 - \left(\frac{\nu_2}{\xi_2}\right)^2$, being ν_2 the second positive root of J_0 (see Table 7.1). The first root ν_1 has been neglected because the first term $-\frac{\rho_1}{J_0(\xi_1)^2} J_0\left(\xi_1\sqrt{1-u}\right)$ is negative when $u = 1 - \left(\frac{\nu_2}{\xi_2}\right)^2$, so the condition is satisfied without restrictions. Hence, a necessary condition derived from the general condition (7.10) is

$$-\frac{\rho_1}{J_0(\xi_1)^2} J_0\left(\xi_1 \frac{\nu_2}{\xi_2}\right) \leq 1 \iff \rho_1 \leq \frac{J_0(\xi_1)^2}{J_0\left(\xi_1 \frac{\nu_2}{\xi_2}\right)} = 0.612.$$

7.4 Diagonal family with logistic marginals

In this section we construct a bivariate family with logistic marginals.

Proposition 7.4.1 *Let* X, Y *be r.v.'s with standard logistic distribution, i.e., with density given by*

$$f(x) = \frac{\exp(-x)}{(1 + \exp(-x))^2}, \quad x \in \mathbb{R}.$$

Let $L_j(t)$, $t \in [0,1]$ *be the* $j-th$ *shifted Legendre polynomial on* $[0,1]$, *and* $F(x) = 1/(1+\exp(-x))$ *the cdf. Let* $\{\rho_j\}$ *be a decreasing sequence of non-negative real numbers such that* $1 \geq \rho_1 > \rho_2 > ... > 0$. *Then*

1. *The function*

$$h(x,y) = f(x) f(y) \left[1 + \sum_{j=1}^{N} \rho_j \; L_j(F(x)) L_j(F(y)) \right], \qquad x, y \in \mathbb{R} \quad (7.13)$$

is the density of a diagonal family with logistic marginals iff

$$\sup_{(u,v)\in[0,1]^2} \left[-\sum_{j=1}^{N} \rho_j L_j(u) L_j(v) \right] \leq 1. \quad (7.14)$$

Each ρ_j *is the correlation between the* $j-th$ *standardized principal components* $F_j(X) = L_j(F(X))$, $F_j(Y)$,

$$\rho_j = Cor(L_j(F(X)), L_j(F(Y))).$$

2. *These coefficients must satisfy the condition*

$$-\sum_{j=1}^{N} \rho_j \sqrt{2j+1} L_j(u) \leq 1, \quad (7.15)$$

uniformly in $u \in [0,1]$.

3. *The correlation coefficient of* (X,Y) *is given by*

$$Cor(X,Y) = \frac{3}{\pi^2} \sum_{k=1}^{N} \frac{4k-1}{k^2(2k-1)^2} \rho_{2k-1}. \quad (7.16)$$

PROOF. The eigenfunctions of the covariance kernel (6.3) when the marginals are logistic are

$$\varphi_j(t) = c_j L_j'(F(t)) f(t),$$

where $c_j = \sqrt{\frac{2j+1}{j(j+1)}}$. Thus the $j-th$ principal component is

$$\begin{aligned} f_j(X) &= \int_0^X c_j L_j'(F(t)) f(t) \; dt \\ &= c_j L_j(F(X)) \\ &= c_j \left[P_j(2F(X)-1) - (-1)^j \right], \end{aligned}$$

where P_j is the $j-th$ Legendre polynomial on $[-1,1]$. The mean and variances are $\mu_j = -(-1)^j c_j$, $\lambda_j = \frac{1}{j(j+1)}$ and, hence, the $j-th$ standardized principal component is

$$\begin{aligned} F_j(X) &= \sqrt{2j+1} P_j(2F(X)-1) \\ &= L_j(F(X)). \end{aligned}$$

Substitution of F_j in (6.4) gives (7.13). Conditions (7.14) and (7.15) are obtained from direct application of Theorem 7.1.1.
To prove 3) we compute I_j as in Theorem 7.1.1: let $F(x) = u$, $x = \log\left(\frac{u}{1-u}\right)$. Then

$$I_j = \int_0^1 \log\left(\frac{u}{1-u}\right) L_j(u)\, du.$$

In computing this integral we find $I_1 = \sqrt{3}$, $I_2 = 0$, $I_3 = \frac{\sqrt{7}}{6}$, $I_4 = 0$, $I_5 = \frac{\sqrt{11}}{15}$, $I_6 = 0$, etc. In general we have

$$I_j = \sqrt{2j+1}\left(\frac{1-(-1)^j}{j(j+1)}\right) = \begin{cases} \frac{2\sqrt{2j+1}}{j(j+1)}, & \text{if } j \text{ is odd,} \\ 0 & \text{, if } j \text{ is even.} \end{cases}$$

By taking $j = 2k-1$ only the odd terms take part:

$$\begin{aligned} Cov(X,Y) &= \sum_{j=1}^{N}\left(\sqrt{2j+1}\left(\frac{1-(-1)^j}{j(j+1)}\right)\right)^2 \rho_j \\ &= \sum_{k=1}^{N} \frac{4k-1}{k^2(2k-1)^2}\rho_{2k-1}, \end{aligned}$$

and since $Var(X) = \frac{\pi^2}{3}$, equation (7.16) holds. □

7.4.1 The cumulative distribution function

The integrals of the Legendre polynomials can be expressed in closed form for all $j \geq 1$ (see 5.5). Let us denote

$$L_j^*(F(x)) = \int_{-\infty}^{x} L_j(F(s))\, dF(s),$$

or, equivalently,

$$L_j{}^*(u) = \int_0^u L_j(t)\, ds.$$

The cumulative distribution function for the density (7.13) is

$$H(x,y) = F(x)F(y) + \sum_{j=1}^{N} \rho_j\ L_j^*(F(x))\, L_j^*(F(y)), \qquad x, y \in \mathbb{R},$$

and the corresponding copula is

$$C(u,v) = uv + \sum_{j=1}^{N} \rho_j\ L_j^*(u)\, L_j^*(v), \qquad u, v \in [0,1].$$

That is, this is the Generalized FGM family (see Section 5.2), Chapter 5). For instance, if $N=1$

$$C(u,v)=uv\left[1+3\rho_1(u-1)(v-1)\right] \ , \qquad u,v\in[0,1],$$

and if $N=2$

$$C(u,v)=uv\left[1+3\rho_1(u-1)(v-1)+5\rho_2\left(2u^2-3u+1\right)\left(2v^2-3v+1\right)\right],$$

$u,v\in[0,1]$.

7.4.2 Some finite dimensional densities

The shifted Legendre polynomials are defined on $[0,1]$. The first three are

$$\begin{aligned} L_1(u) &= \sqrt{3}(2u-1), \ L_2(u)=\sqrt{5}\left(6u^2-6u+1\right), \\ L_3(u) &= \sqrt{7}\left(20u^3-30u^2+12u-1\right). \end{aligned}$$

Figure 7.6 shows these polynomials:

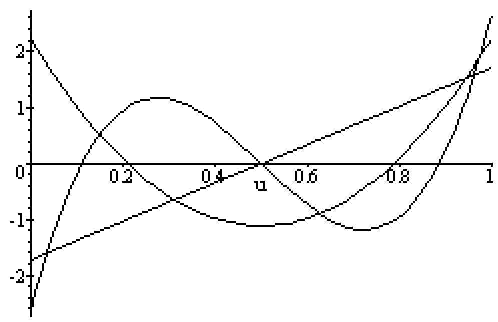

Figure 7.6: First three shifted Legendre polynomials

One-dimensional

If only the first canonical correlation is strictly positive, we obtain the well known FGM family of distributions. Let $F'(x)=f(x)$ be the common density of two r.v.'s X,Y which follow a standard logistic distribution. The function

$$h(x,y)=f(x)f(y)\left[1+3\rho_1\ (2F(x)-1)(2F(y)-1)\right], \qquad x,y\in\mathbb{R}$$

with parameter $\theta = 3\rho_1$ is a bivariate density $h \in \mathcal{F}(f, f)$ iff $\rho_1 \leq \frac{1}{3}$. This function is obtained from (7.13) taking $N = 1$. Then it is obvious that

$$\sup_{(u,v)\in[0,1]^2} [-3\rho_1 \ (2u-1)(2v-1)] = 3\rho_1,$$

which is attained at $u = 0$, $v = 1$. Then (7.14) holds iff

$$\rho_1 = \frac{|\theta|}{3} \leq \frac{1}{3}, \quad \theta \in [-1, 1].$$

The density of the corresponding copula is

$$c(u, v) = 1 + 3\rho_1 \ (2u-1)(2v-1), \qquad u, v \in [0, 1].$$

The correlation coefficient of (X, Y) is

$$Cor(X, Y) = \frac{9}{\pi^2}\rho_1.$$

Hence the bound for $Cor(X, Y)$ is $\frac{3}{\pi^2} \simeq 0.30396$, lower than the bound for the FGM copula (which is $\frac{1}{3}$).

Example 7.4.2 *Figures 7.7 and 7.8 show these densities when $\rho_1 = \frac{1}{3}$.*

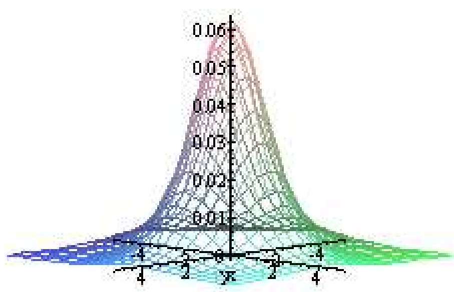

Figure 7.7: FGM density: $\rho_1 = \frac{1}{3}$

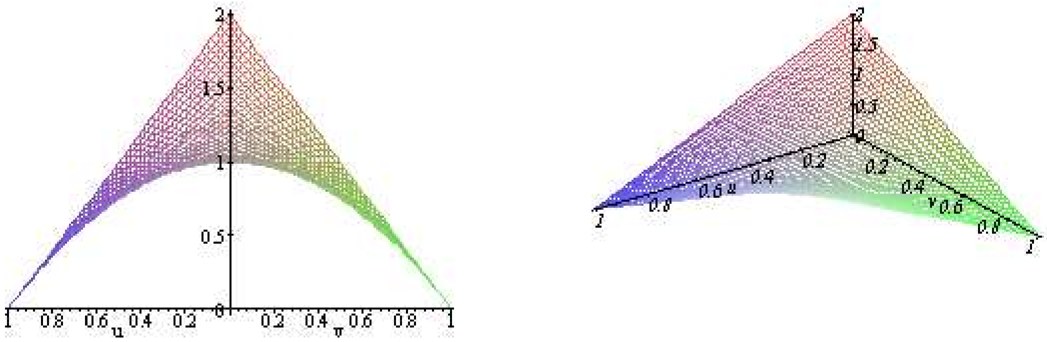

Figure 7.8: Density of the FGM copula

Two-dimensional

If only the first and second canonical correlations are positive then we obtain the family

$$\begin{aligned} h(x,y) &= f(x)\, f(y) \left[1 + 3\rho_1 (2F(x) - 1)(2F(y) - 1) \right. \\ &\quad \left. + 5\rho_2 \left(6F(x)^2 - 6F(x) + 1\right)\left(6F(y)^2 - 6F(y) + 1\right)\right], \end{aligned}$$

$x, y \in \mathbb{R}$, where (ρ_1, ρ_2) must satisfy the necessary condition (7.15) if the marginals are standard logistic. Then $-\sqrt{3}\rho_1 L_1(u) - \sqrt{5}\rho_2 L_2(u) \le 1$, equivalent to

$$30\rho_2 u^2 + (6\rho_1 - 30\rho_2)\, u + (1 - 3\rho_1 + 5\rho_2) \ge 0.$$

This second degree polynomial in u takes only positive values iff the discriminant of the equation is non-positive, i.e., iff

$$(6\rho_1 - 30\rho_2)^2 - 4 \cdot 30\rho_2 (1 - 3\rho_1 + 5\rho_2) \le 0,$$

equivalent to

$$\frac{\rho_1^2}{1/3} + \frac{\left(\rho_2 - \frac{1}{5}\right)^2}{1/25} \le 1.$$

This condition is satisfied by the pairs of canonical correlations $\rho_1 > \rho_2$ lying in the area represented by Figure 7.9:

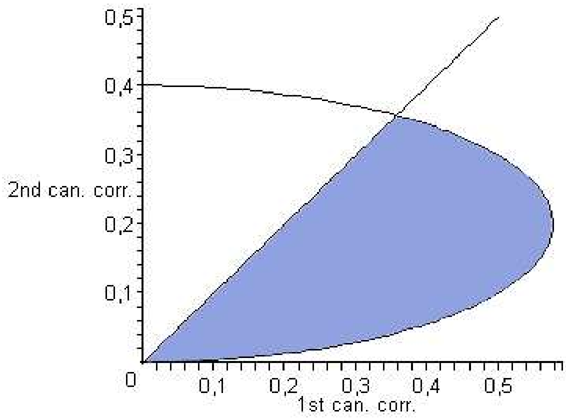

Figure 7.9: First and second canonical correlations for logistic margins

If $N = 2$ Pearson's correlation coefficient has the same expression as in the one-dimensional case, but with higher bound:

$$Cor\,(X,Y) = \frac{9}{\pi^2}\rho_1 \leq \frac{3\sqrt{3}}{\pi^2} \simeq 0.52648.$$

Example 7.4.3 *If* $\rho_1 = \frac{1}{\sqrt{3}}, \rho_2 = \frac{1}{5}$ *the function*

$$c\,(u,v) = 1 + \frac{3}{\sqrt{3}}\,(2u-1)\,(2v-1) + \left(6u^2 - 6u + 1\right)\left(6v^2 - 6v + 1\right),$$

$u, v \in [0,1]$, *is the density of a copula corresponding to a bivariate distribution with logistic marginals:*

$$\begin{aligned} h\,(x,y) &= f\,(x)\,f\,(y)\,[1 + \frac{3}{\sqrt{3}}\,(2F\,(x)-1)\,(2F\,(y)-1) + \\ &+ \left(6F\,(x)^2 - 6F\,(x) + 1\right)\left(6F\,(y)^2 - 6F\,(y) + 1\right)], \end{aligned}$$

$x, y \in \mathbb{R}$. *Figure 7.10 below is the density* $c\,(u,v)$.

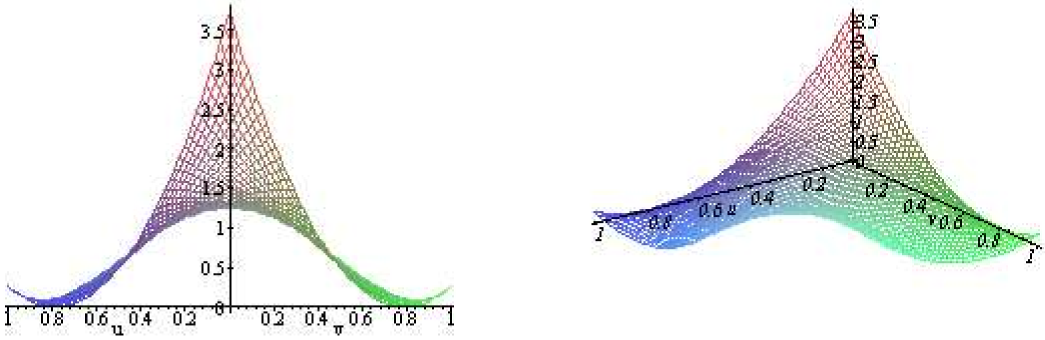

Figure 7.10: Density: $\rho_1 = \frac{1}{\sqrt{3}}, \rho_2 = \frac{1}{5}$

Remark 7.4.4 *It is evident that* $\rho_1 = Cor\left(L_1\left(F\left(X\right)\right), L_1\left(F\left(Y\right)\right)\right)$ *is Spearman's rho, since the first shifted Legendre polynomial is linear in* F. *This known result is in accordance with the expansion*

$$\rho_S\left(X,Y\right) = \sum_{i=1}^{\infty} \rho_i Cor\left(a_i\left(X\right), F\left(X\right)\right) Cor\left(b_i\left(Y\right), G\left(Y\right)\right), \qquad (7.17)$$

(see [18]). If (X,Y) *is a random vector with density belonging to the family defined by the expansion*

$$h\left(x,y\right) = f\left(x\right) f\left(y\right)\left[1 + \sum_{j=1}^{N} \rho_j L_j\left(F\left(x\right)\right) L_j\left(F\left(y\right)\right)\right], \qquad x,y \in \mathbb{R},$$

and we substitute the standardized principal components, $L_i\left(F\right)$, *for the canonical variables in* (7.17), *we obtain*

$$\begin{aligned}\rho_S\left(X,Y\right) &= \sum_{i=1}^{\infty} \rho_i Cor\left(L_i\left(X\right), F\left(X\right)\right) Cor\left(L_i\left(Y\right), F\left(Y\right)\right) \\ &= \sum_{i=1}^{\infty} \rho_i \delta_{i1} = \rho_1.\end{aligned}$$

7.5 Diagonal family with Pareto marginals

In this last section we construct a family with Pareto marginals.

Proposition 7.5.1 *Let* X, Y *be two r.v.'s with Pareto distribution with density given by* $f\left(x\right) = 3x^{-4}\,\mathbf{1}_{(1,\infty)}\left(x\right)$. *Thus* X *(and* Y*) follows the Pareto*(α,θ) *distribution with* *parameters* $\alpha = 3, \theta = 1$. *Let* η_j *be the* $j-th$ *positive root of the equation* $x = \tan(x)$. *Then:*

1. *The expansion*

$$h(x,y) = f(x) f(y) \left[1 + \sum_{j=1}^{N} \rho_j \frac{2}{3\sin^2(\eta_j)} \; x \sin(\eta_j/x) \;\; y \sin(\eta_j/y) \right], \quad (7.18)$$

$x, y > 1$, *with* $0 \le \rho_j \le 1$, *for* $j \ge 1$, *is a bivariate density with Pareto*$(3,1)$ *marginals iff*

$$\sup_{(u,v)\in[0,1]^2} \left[\frac{-2}{3} \sum_{j=1}^{N} \frac{\rho_j \sin\left(\eta_j \sqrt[3]{1-u}\right) \sin\left(\eta_j \sqrt[3]{1-v}\right)}{\sin^2(\eta_j) \sqrt[3]{1-u}\sqrt[3]{1-v}} \right] \le 1. \quad (7.19)$$

Each ρ_j *is the correlation coefficient between the* $j-th$ *standardized principal components* $F_j(X) = \sqrt{\frac{2}{3}} X \sin(\eta_j/X) / \sin(\eta_j)$, $F_j(Y)$,

$$\rho_j = Cor\left(X \sin(\eta_j/X), Y \sin(\eta_j/Y)\right).$$

2. *These coefficients must satisfy the condition*

$$\frac{-2}{3} \sum_{j=1}^{N} \frac{\rho_j \eta_j \sin\left(\eta_j \sqrt[3]{1-u}\right)}{\sin^2(\eta_j) \sqrt[3]{1-u}} \le 1, \quad (7.20)$$

uniformly in $u \in [0,1]$.

3. *The correlation coefficient is given by*

$$Cor(X,Y) = 8 \sum_{j=1}^{N} \rho_j \left(\frac{1-\cos(\eta_j)}{\eta_j \sin(\eta_j)} \right)^2. \quad (7.21)$$

PROOF. The eigenfunctions of the covariance kernel (6.3) when marginals are Pareto$(3,1)$ are

$$\varphi_j(t) = a_j \left(\sin(\eta_j/t) - \frac{1}{t} \cos(\eta_j/t) \right), \quad t > 1$$

where $a_j = 2\eta_j^{-\frac{1}{2}} (2\eta_j - \sin(2\eta_j))^{-\frac{1}{2}}$. Since $\eta_j = \tan(\eta_j)$, and $\sin(2\alpha) = 2\sin\alpha\cos\alpha$, we have $a_j = 2^{\frac{1}{2}} (\eta_j \sin(\eta_j))^{-1}$. Thus the $j-th$ principal component is

$$\begin{aligned} f_j(X) &= \int_1^X a_j \left(\sin(\eta_j/t) - \frac{1}{t}\cos(\eta_j/t) \right) dt \\ &= a_j \left(X \sin(\eta_j/X) - \sin(\eta_j) \right) \\ &= 2^{\frac{1}{2}} (\eta_j \sin(\eta_j))^{-1} \left(X \sin(\eta_j/X) - \sin(\eta_j) \right). \end{aligned}$$

Finally, as $\mu_j = -\sqrt{2}/\eta_j$, and $\lambda_j = 3/\eta_j^2$, the $j-th$ standardized principal component is

$$F_j(X) = \sqrt{\frac{2}{3}} \frac{X \sin(\eta_j/X)}{\sin(\eta_j)}.$$

As in the previous constructions, substitution of F_j in (6.4) gives (7.18). Then 1) and 2) follow from Theorem 7.1.1.
From $Var(X) = Var(Y) = \frac{3}{4}$, Pearson's correlation coefficient is

$$Cor(X,Y) = \frac{4}{3} \sum_{j=1}^{N} \rho_j I_j^2,$$

where I_j is given by (7.3):

$$\begin{aligned} I_j &= \int_0^1 \sqrt{\frac{2}{3}} \frac{(1-u)^{-\frac{2}{3}}}{\sin(\eta_j)} \sin\left(\eta_j (1-u)^{\frac{1}{3}}\right) du \\ &= \tfrac{\sqrt{6}}{\eta_j \sin(\eta_j)} (1 - \cos(\eta_j)), \end{aligned}$$

and (7.21) follows. □

7.5.1 The cumulative distribution function

The cdf $H(x,y) = \int_1^x \int_1^y h(s,t)\, ds dt$, is given by:

$$H(x,y) = F(x) F(y) + \sum_{j=1}^{N} \frac{6\rho_j}{\sin^2(\eta_j)} \int_1^x s^{-3} \sin(\eta_j/s)\, ds \int_1^y t^{-3} \sin(\eta_j/t)\, dt.$$

Integration by parts, taking $u = s^{-1}$, $dv = s^{-2} \sin(\eta_j/s)$ gives

$$\begin{aligned} \int_1^x s^{-3} \sin(\eta_j/s)\, ds &= \left(s^{-1} \frac{-\cos(\eta_j/s)}{\eta_j}\right]_1^x - \int_1^x \frac{\cos(\eta_j/s)}{\eta_j} s^{-2} ds \\ &= \frac{\cos(\eta_j)}{\eta_j} - \frac{\cos(\eta_j/x)}{x\eta_j} + \left(\frac{\sin(\eta_j/s)}{\eta_j^2}\right]_1^x \\ &= \frac{\cos(\eta_j)}{\eta_j} - \frac{\cos(\eta_j/x)}{x\eta_j} + \frac{\sin(\eta_j/x)}{\eta_j^2} - \frac{\sin(\eta_j)}{\eta_j^2}, \end{aligned}$$

where the first and the last terms of this sum are equal since $\eta_j = \tan(\eta_j)$. Thus

$$\begin{aligned} H(x,y) &= F(x) F(y) + \sum_{j=1}^{N} \frac{6\rho_j}{\eta_j^4 \sin^2(\eta_j)} \times \\ &\times \left(\frac{x \sin(\eta_j/x) - \eta_j \cos(\eta_j/x)}{x}\right) \left(\frac{y \sin(\eta_j/y) - \eta_j \cos(\eta_j/y)}{y}\right), \end{aligned}$$

$x, y > 1$. The change of variables $F(x) = u = 1 - x^{-3}$, $F(y) = v$ gives the corresponding copula

$$\begin{aligned} C(u,v) &= uv + \sum_{j=1}^{N} \frac{6\rho_j \sqrt[3]{(1-u)(1-v)}}{\eta_j^4 \sin^2(\eta_j)} \left(\frac{\sin\left(\eta_j \sqrt[3]{1-u}\right)}{\sqrt[3]{1-u}} - \eta_j \cos\left(\eta_j \sqrt[3]{1-u}\right) \right) \\ &\quad \times \left(\frac{\sin\left(\eta_j \sqrt[3]{1-v}\right)}{\sqrt[3]{1-v}} - \eta_j \cos\left(\eta_j \sqrt[3]{1-v}\right) \right), \end{aligned}$$

$u, v \in [0,1]$.

7.5.2 Some finite dimensional densities

Several examples illustrate different cases which appear when considering a finite set of strictly positive canonical correlations. There is not a closed form for the solutions of $x = \tan(x)$. Approximate values are given below, where $\zeta_j = \left(\frac{1-\cos\eta_j}{\eta_j \sin\eta_j}\right)^2$.

j	$\eta_j = \tan(\eta_j)$	$\sin(\eta_j)$	ζ_j
1	4.49341	−0.9761	0.77017
2	7.72525	0.9917	0.12943
3	10.9041	−0.9958	0.10101
4	14.0662	0.9975	0.00438

Table 7.2: First roots of $x = \tan(x)$, and other related values

One-dimensional

If only the first canonical correlation is non-null then condition (7.19) is satisfied iff

$$\sup_{(u,v)\in[0,1]^2} \left[\frac{-2\rho_1}{3\sin^2\eta_1} \frac{\sin\left(\eta_1 \sqrt[3]{1-u}\right)\sin\left(\eta_1 \sqrt[3]{1-v}\right)}{\sqrt[3]{1-u}\sqrt[3]{1-v}} \right] \le 1.$$

Since

$$\min_{u\in[0,1]} \frac{\sin\left(\eta_1 \sqrt[3]{1-u}\right)}{\sqrt[3]{1-u}} = \sin\eta_1 < 0,$$

and the minimum is attained at $u = 0$, while

$$\max_{v\in[0,1]} \frac{\sin\left(\eta_1 \sqrt[3]{1-v}\right)}{\sqrt[3]{1-v}} = \eta_1 > 0,$$

and the maximum is attained at $v=1$, (7.19) is satisfied iff $\frac{-2}{3}\frac{\rho_1\eta_1}{\sin(\eta_1)} \leq 1$, equivalent to

$$\rho_1 \leq \frac{-3\sin(\eta_1)}{2\eta_1} \simeq 0.3258.$$

From Table 7.2 we can compute the correlation coefficient from (7.21):

$$\begin{aligned} Cor(X,Y) &= 8\rho_1 \left(\frac{1-\cos(\eta_1)}{\eta_1 \sin(\eta_1)}\right)^2 \\ &= 0.61614\rho_1. \end{aligned}$$

Hence, if $N=1$ the bound for the correlation is 0.20074.

Example 7.5.2 *Expansion (7.18) with* $\rho_1 = 0.3258 > \rho_2 = 0$, *gives*

$$\begin{aligned} h(x,y) &= f(x) f(y) \left[1 + \frac{2 \cdot 0.3258}{3\sin^2\eta_1}\, x\sin(\eta_1/x)\;\; y\sin(\eta_1/y)\right], \\ &= 9x^{-4}y^{-4}\left[1 + \frac{0.2172}{\sin^2\eta_1}\, x\sin(\eta_1/x)\;\; y\sin(\eta_1/y)\right], \qquad x,y>1, \end{aligned}$$

that is the density of a 1*-dimensional diagonal distribution with marginals Pareto*(3, 1)*. Figure 7.11 shows the density of the corresponding copula*

$$c(u,v) = 1 + \frac{2}{3}\frac{0.3258}{\sin^2(\eta_1)}\frac{\sin\left(\eta_1\sqrt[3]{1-u}\right)}{\sqrt[3]{1-u}}\frac{\sin\left(\eta_1\sqrt[3]{1-v}\right)}{\sqrt[3]{1-v}}.$$

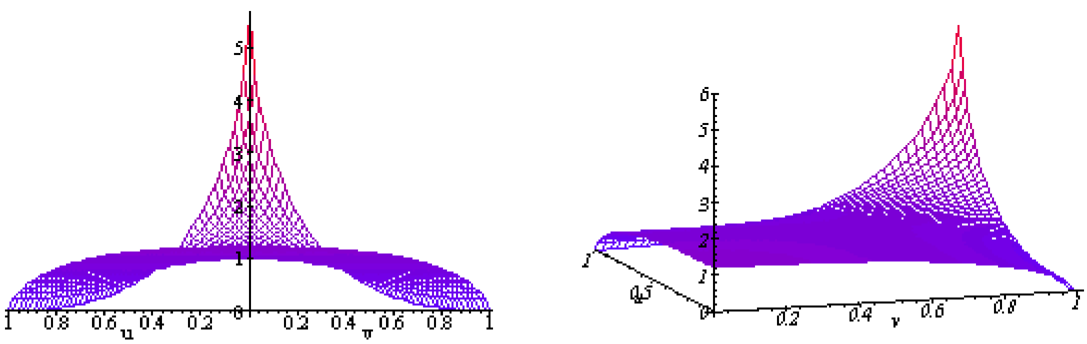

Figure 7.11: Density $c(u,v)$*(Pareto marginals):* $\rho_1 = 0.3258$

Two-dimensional

The complexity of this function, when two positive correlations are considered, has not so far allowed us to find the values ρ_1, ρ_2 which satisfy neither condition (7.19), nor condition (7.20). We give a less general necessary condition, obtained when the second term of the sum vanishes[2], i.e., taking $\sqrt[3]{1-u} = \frac{\arcsin 0}{\eta_2} = \frac{2\pi}{\eta_2}$. Then we have

$$\begin{aligned}
\frac{-2}{3}\sum_{j=1}^{2}\frac{\rho_j\eta_j \sin\left(\eta_j\sqrt[3]{1-u}\right)}{\sin^2(\eta_j)\sqrt[3]{1-u}} &\leq 1, \\
-\frac{2}{3}\frac{\rho_1\eta_1\eta_2}{\sin^2(\eta_1)}\frac{\sin(2\pi\eta_1/\eta_2)}{2\pi} &\leq 1, \\
\rho_1 \leq -\frac{3\pi\sin^2(\eta_1)}{\eta_1\eta_2\sin(2\pi\eta_1/\eta_2)} &\simeq 0.52706.
\end{aligned}$$

As in the previous case, from Table 7.2 we can compute the correlation coefficient from (7.21):

$$\begin{aligned}
Cor(X,Y) &= 8(\rho_1\zeta_1 + \rho_2\zeta_2) \\
&= 0.61614\rho_1 + 0.10355\rho_2.
\end{aligned}$$

We cannot compute the bound in this case but it is clearly higher than the correlation in the one-dimensional case, as the following example shows.

Example 7.5.3 *Let us consider* $\rho_1 = 0.5 > \rho_2 = 0.1 > \rho_3 = 0$. *Expansion* (*7.18*) *gives a* $2-$*dimensional bivariate diagonal density with marginals Pareto* $(3,1)$. *The graph below shows the density of the corresponding copula:*

$$\begin{aligned}
c(u,v) = 1 + \frac{0.5\cdot 2}{3\sin^2(\eta_1)}\frac{\sin\left(\eta_1\sqrt[3]{1-u}\right)}{\sqrt[3]{1-u}}\frac{\sin\left(\eta_1\sqrt[3]{1-v}\right)}{\sqrt[3]{1-v}} + \\
+\frac{0.1\cdot 2}{3\sin^2(\eta_2)}\frac{\sin\left(\eta_2\sqrt[3]{1-u}\right)}{\sqrt[3]{1-u}}\frac{\sin\left(\eta_2\sqrt[3]{1-v}\right)}{\sqrt[3]{1-v}},
\end{aligned}$$

$(u,v) \in [0,1]^2$.

[2]Taking $\sqrt[3]{1-u} = \frac{\pi}{\eta_2}$ we obtain a negative value, so the condition is trivially satisfied.

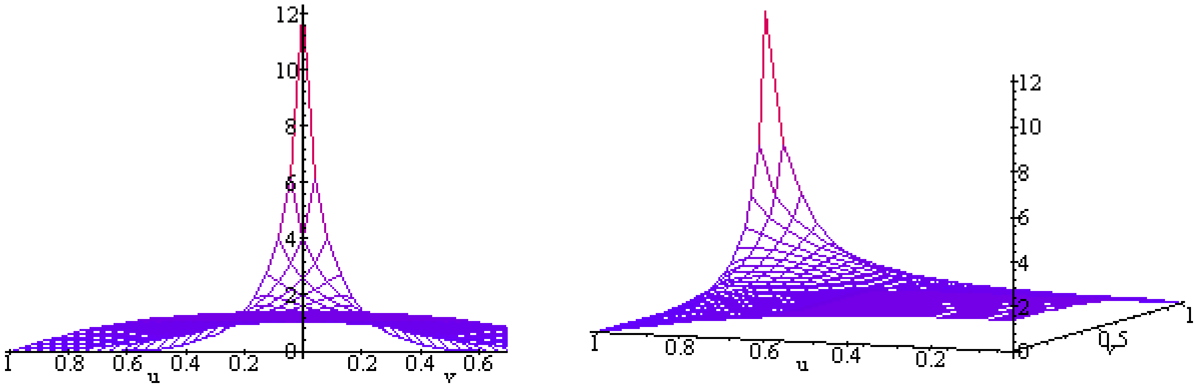

Figure 7.12: Density $c(u,v)$ (Pareto marginals): $\rho_1 = 0.5$, $\rho_2 = 0.1$

Here $Cor(X,Y) = 0.31843$

7.6 Summary

Table 7.3 summarizes the main properties of the constructed joint bivariate densities $h(x,y)$: for every marginal variable we give the correlation of the family, the necessary and sufficient condition for h to be a density if $N=1$, and a necessary condition if $N=2$.

Marginal distribution	$Cor(X,Y)$	$N=1$	$N=2$
$Uniform\ (0,1)$	$\frac{96}{\pi^4}\sum_{k=1}^{N}\frac{\rho_{2k-1}}{(2k-1)^4}$	$\rho_1 \le \frac{1}{2}$	$\rho_1^2 + \frac{\left(\rho_2-\frac{1}{4}\right)^2}{1/8} \le \frac{1}{2}$
$Exponential\ (1)$	$\sum_{j=1}^{N}\rho_j\left(\frac{4(1-J_0(\xi_j))}{\xi_j^2 J_0(\xi_j)}\right)^2$	$\rho_1 \le 0.40276$	$\rho_1 \le 0.61198$
$Logistic\ (0)$	$\frac{3}{\pi^2}\sum_{k=1}^{N}\frac{\rho_{2k-1}(4k-1)}{k^2(2k-1)^2}$	$\rho_1 \le \frac{1}{3}$	$\frac{\rho_1^2}{1/3} + \frac{\left(\rho_2-\frac{1}{5}\right)^2}{1/25} \le 1$
$Pareto\ (3,1)$	$8\sum_{j=1}^{N}\rho_j\left(\frac{1-\cos(\eta_j)}{\eta_j\sin(\eta_j)}\right)^2$	$\rho_1 \le 0.3258$	$\rho_1 \le 0.52706$

Table 7.3: Some properties of the constructed bivariate densities

7.7 Testing independence

Suppose that h is the density of a cdf H, of the form

$$h(x,y) = c(\theta)\exp\left\{\sum_{j=1}^{k}\theta_j L_j(F(x))\,L_j(G(y))\right\}, \tag{7.22}$$

where $\theta = (\theta_1, ..., \theta_k)'$, $c(\theta)$ is a normalizing constant and $L_j(t)$ are the shifted Legendre polynomials on $[0,1]$. If the marginal cdf's F, G are known, Kallenberg and Ledwina (1999, [62]) showed that the null hypothesis of independence ($H_0 : \theta = \mathbf{0}$ vs. $H_1 : \theta \neq \mathbf{0}$) is rejected by the score test for large values of

$$\sum_{j=1}^{k}\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n} L_j(F(X_i))\,L_j(G(Y_i))\right\}^2, \tag{7.23}$$

where (X_i, Y_i), $i = 1, ..., n$ are iid as (X, Y). Thus this test is based on the correlations $Cor(L_j(F(X)), L_j(G(Y)))$, $j = 1, ..., k$. If $k = 1$, this is a test on the significance of Spearman's correlation coefficient ρ_S. Notice that variables $L_j(F(X))$ are uncorrelated and standardized, but they are only principal components if X is standard logistic. Cuadras (2002, [16]) suggests that the score test for independence should be based on $L_j(F(X))$ only if marginals are logistic. Otherwise, the principal components F_j, G_j should replace $L_j(F(X))$. For instance, the statistic

$$\sum_{j=1}^{k}\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{1}{J_0(\xi_j)^2}J_0(\xi_j\exp(-X_i/2))\,J_0(\xi_j\exp(-Y_i/2))\right\}^2 \tag{7.24}$$

is proposed as a score test when marginals are known to be exponential with unit mean. For example, if the true bivariate family is given by (7.22) with exponential marginals with unit mean, $n = 25$, $k = 1$, significance level of 0.05, the test for independence proves to be uniformly more powerful using (7.24) than using (7.23).

Concluding remarks

We have developed the theoretical framework required to extend the Related Metric Scaling techniques to the continuous approach. Next step must be to formulate this extension explicitely, using the formulae obtained in Chapter 3 for the joint and the intersection kernels, from the Distance Based methodology.

The covariance between functions of two random variables, fixing one of the marginal variables, has been used in the definition of the intersection between kernels. We have obtained a variety of properties of the covariance between functions of two random variables, joined by a symmetric and PQD bivariate distribution. Most of these properties are obvious. Our purpose has been to present most of the expected properties of this inner product, and to relate them with the joint distribution function. We hope that other researchers find these presentation useful. As an application of the properties of the covariance, we have proved the addequacy of a new affinity measure between functions, which depends on H. We have defined and studied the main properties of the $H-$affinity. A number of functions and joint parametric families of distributions have been studied. We expect to find further results involving other sets of functions and distributions, relevant in Probability and Statistics.

The dimension of a joint distribution $H \in \mathcal{F}(F, G)$ has been defined by means of the $H-$affinity. The cardinality of the set of eigenvalues of the covariance kernel $H(x, y) - F(x) G(y)$ may determine this dimension. If H can be represented via a diagonal expansion, the number of non-null canonical correlations gives the dimension of H. It has been obtained for some parametric families and we expect to obtain the dimension of other distributions. The dimension of a joint distribution is null if there is stochastic independence, and infinite countable if H is the Fréchet upper bound. Hence, the dimension may be viewed as another "measure" of dependence between two random variables.

Finally, we have applied the diagonal expansion method to construct some specific diagonal families via principal components. We have given necessary and sufficient conditions for the densities in the one-dimensional case. Our research is now focused in the case of two non-null canonical correlations. When H can be represented by means of a diagonal expansion, we have found that the maximum correlation increases if the dimension (the number of non-null correlations) is higher. We think that this method may be useful to simulate bivariate samples with fixed correlations and given marginals. This method also suggests that the standardized principal components may improve some independence tests and goodness-of-fit tests as well. We expect to confirm this via simulation studies. The non-symmetric case has not been studied yet.

Bibliography

[1] Ali, M.M., Mikhail, N.N, Haq, M.S. (1978), A class of bivariate distributions including the bivariate logistic. *Journal of Multivariate Analysis*, **8**, 405-412.

[2] Arenas, C., Escudero, T., Mestres, F., Coll, M.D., Cuadras, C.M. (2000), Cacromos: A computer program to reconstruct the position of chromosomas on the metaphase plate. *Hereditas*, **132**, 157-159.

[3] Arenas, C., Cuadras, C.M., Turbón, D. (2001), Joint representation of multivariate data with applications to human evolution. In: *VIII Spanish Biometric Conference*, Pamplona, March 21-30.

[4] Arenas, C., Cuadras, C.M. (2004), Comparing two methods for joint representation of multivariate data. *Communications in Statistics-Simul. and Comp.*, **33**, 415-430.

[5] Barrett, J.F., Lampard, D.G. (1955), An expansion of some second-order probability distributions and its applications to noise problems. *IEEE (Institute of Electrical and Electronics Engineers) Transactions on Information Theory,* IT-**1**, 10-15.

[6] Capéràa, P., Genest, C. (1993), Spearman's ρ is larger than Kendall's τ for positively dependent random variables. *J. Nonparametr. Statist.*, **2**, 183-194.

[7] Cerdà, J. (1996), *Análisis Real.* Edicions de la Universitat de Barcelona, (in Spanish).

[8] Cerdà, J. (2001), *Análisis Funcional.* Unpublished lecture notes on Functional Analysis, Facultat de Matemàtiques, Universitat de Barcelona, (in Spanish).

[9] Consonni, G., Scarsini, M. (1982), Lo studio de la concordanza nel contesto della teoria della variabilità superficiale. *Statistica*, **42**, 69-77.

[10] Conway, D.A. (1979), Farlie-Gumbel-Morgenstern distributions. In: *Encyclopedia of Statistical Sciences*, **3**, S.Kotz and N.L. Johnson, editors (John Wiley & Sons, New York), 28-31.

[11] Courant, R., Hilbert, D. (1953), *Methods of Mathematical Physics.* 1st English ed. vol I. John Wiley & Sons.

[12] Cuadras, C.M. (1992), Probability Distributions with given multivariate marginals and given dependence structure. *Journal of Multivariate Analysis,* **42**, 51-66.

[13] Cuadras, C.M. (1996), A distribution with given marginals and given regression curve. In: *Distributions with fixed marginals and related topics* (L. Ruschendorf, B. Schweizer and M. Taylor, eds.), IMS Lecture Notes - Monograph Series, **28**, Instit. of Mathem. Stat. Hayward, CA, 76-83.

[14] Cuadras, C.M. (1998), *Anàlisi de Dades Multivariants.* 3rd ed. Universitat de Barcelona, (in Catalan).

[15] Cuadras, C.M. (1998), Multidimensional Dependencies in classification and ordination. In *Analyses Multidimensionelles des Doneés*; K. Fernández and A. Morineau, Eds.; CISIA-CERESTA, Saint-Mandé France, 15-25.

[16] Cuadras, C.M. (2002), Diagonal Distributions via orthogonal expansions and test of independence. In: *Distributions with given marginals and Statistical Modelling.* C.M. Cuadras, J. Fortiana and J.A. Rodríguez-Lallena, eds. Kluwer Ac. Pub., Dordrecht, 35-42..

[17] Cuadras, C.M. (2002), Correspondence analysis and diagonal expansions in terms of distribution functions. *Journal of Statistical Planning and Inference*, **103**, 137-150.

[18] Cuadras, C.M. (2002), On the Covariance between Functions. *Journal of Multivariate Analysis*, **81**, 19-27.

[19] Cuadras, C.M. (2003), First principal component characterization of a continuous random variable. IMUB, Math. Preprint Series 327.

[20] Cuadras, C.M. (2004), The importance of being the bivariate upper bound. In: *Proceedings of the International Conference on Distribution Theory, Order Statistics and Inference*, University of Cantabria, Spain, 1-14.

[21] Cuadras, C.M., Continuous canonical correlation analysis on a bivariate distribution. Working paper.

[22] Cuadras, C.M., Augé, J. (1981), A continuous general multivariate distribution and its properties. *Communications in Statistics -A Theory and Methods*, **10**(4), 339-353.

[23] Cuadras, C.M, Cuadras, D., Lahlou, Y. (2004), Principal directions for the general Pareto distribution. *Journal of Statistical Planning and Inference*, in press.

[24] Cuadras, C.M., Fortiana, J. (1993), Continuous metric scaling and prediction. In: *Multivariate Analysis: Future Directions 2.* C.M. Cuadras and C.R. Rao, eds. Elsevier, North-Holland, Amsterdam, 47-66.

[25] Cuadras, C.M., Fortiana, J. (1995), A continuous metric scaling solution for a random variable. *Journal of Multivariate Analysis*, **52**, 1-14.

[26] Cuadras, C.M., Fortiana, J. (1996), Weighted continuous metric scaling. In: *Multidimensional Statistical Analysis and Theory of Random Matrices.* A.K. Gupta and V.L. Girko, eds. VSP, The Netherlands, 27-40.

[27] Cuadras, C.M., Fortiana, J. (1997), Continuous scaling on a bivariate copula. *Distributions with given marginals and moment problems.* V. Benes and J. Stepan, eds. Kluwer Ac. Press, 137-142.

[28] Cuadras, C.M., Fortiana, J. (1998), Visualizing categorical data with related metric scaling. In *Visualization of categorical data;* J. Blasius and M. Greenacre, Eds.; Academic Press, London, 365-376.

[29] Cuadras, C.M., Fortiana, J., Greenacre, M.J. (1999), Continuous Extensions of Matrix Formulations in Correspondence Analysis, with Applications to the FGM Family of Distributions. In: *Innovations in Multivariate Statistical Analysis,* R.D.H. Heijmans, D.S.G. Pollock and A. Satorra eds., Kluwer Ac. Publ., 101-116.

[30] Cuadras, C.M, Lahlou, Y. (2000), Some orthogonal expansions for the logistic distribution. *Communications in Statistics - Theory and Methods*, **29**, 2643-2663.

[31] Cuadras, C.M., Lahlou, Y. (2002), Principal Components of the Pareto Distribution. In: *Distributions with given marginals and Statistical Modelling.* C.M.

Cuadras, J. Fortiana and J.A. Rodríguez-Lallena, eds. Kluwer Ac. Pub., Dordrecht, 35-42.

[32] Darsow, W.F., Nguyen, B., Olsen, E.T. (1992), Copulas and Markov processes. *Illinois Journal of Mathematics*, **36**, 600-642.

[33] Devroye, L. (1986), *Non-Uniform Random Variate Generation.* Springer, New York.

[34] Eagleson, G.K. (1964), Polynomial expansions of bivariate distributions. *Annals of Mathematical Statistics*, **35**, 1208-1215.

[35] Eagleson, G.K. (1979), Orthogonal Expansions and U-Statistics. *Australian Journal of Statistics*, **21**(3), 221-237.

[36] Eaton, M.L. (1988), *Group Invariance Applications in Statistics.* Regional Conference Series in Probability and Statistics, **1**.

[37] Esteve, A. (2003), *Distancias estadísticas y relaciones de dependencia entre conjuntos de variables.* Ph.D. Thesis, Universitat de Barcelona (in Spanish).

[38] Eyraud, H. (1936), Les principes de la mesure des corrélations. *Annales de l'Université de Lyon*. Series A **1**, 30-47.

[39] Farlie, D.J.G. (1960), The performance of some correlation coefficients for a general bivariate distribution. *Biometrika* **47**, 307-323.

[40] Feller, W. (1971), *An Introduction to Probability Theory and Its Applications*, vol. II, 2nd ed. John Wiley & Sons, New York.

[41] Fisher, N.I. (1997), Copulas. In: *Encyclopedia of Statistical Sciences*, Update vol. **1**, S.Kotz, C.B. Read, and D.L. Banks, editors (John Wiley & Sons, New York), 159-163.

[42] Fréchet, M. (1951), Sur les tableaux de corrélation dont les marges son données. *Annales de l'Université de Lyon*. Sect. A **9**, 53-77.

[43] Gebelein, H. (1941), Das statistische Problem der Korrelation als Variations und Eigenwertproblem und sein Zusammenhang mit der Ausgleichsrechnung. *Zeitschrift für Angewandte Mathematik und Mechanik*, **21**, 364-379.

[44] Genest, C., MacKay, J. (1986), The joy of copulas: Bivariate distributions with uniform marginals. *The American Statistician*, **40**, 280-285.

[45] Giri, N.C. (1996), *Group Invariance in Statistical Inference.* World Scientific.

[46] Gower, J.C. (1966), Some distance properties of latent root and vector methods in multivariate analysis. *Biometrika*, **53,** 325-338.

[47] Graybill, F.A. (1969), *Matrices with applications in Statistics*, 2nd ed. The Wadsworth Statistics/Probability Series.

[48] Gumbel, E.J. (1960), Bivariate exponential distributions. *Journal of the American Statistical Association*, **55**, 698-707.

[49] Gumbel. E.J. (1960), Multivariate distributions with given margins and analytical examples. *Bulletin de l'Institut International de Statistique*, **37**, Book 3, 363-373.

[50] Gumbel, E.J. (1961), Bivariate logistic distributions. *Journal of the American Statistical Association*, **56**, 335-349.

[51] Harville, D.A. (1997), *Matrix Algebra from a Statistician's perspective.* Springer-Verlag, New York.

[52] Hoeffding, W. (1940), Masstabinvariante Korrelationtheorie. *Schriften Math. Inst. Univ. Berlin*, **5**, 181-223. English translation (1999): Scale-Invariant Correlation Theory in *The Collected Works of Wassily Hoeffding*, N.I. Fisher and P.K. Sen, Eds., 57-107, Springer-Verlag, New York.

[53] Horn, R.A., Johnson, C.R. (1985), *Matrix Analysis*, Cambridge University Press.

[54] Hutchinson, T.P., Lai, C.D. (1990), *Continuous Bivariate Distributions, Emphasising Applications.* Rumsby Scientific Publishing, Adelaide, South Australia.

[55] Joe, H. (1993), Parametric family of multivariate distributions with given margins. *Journal of Multivariate Analysis*, **46**, 262-282.

[56] Joe, H. (1997), *Multivariate Models and Dependence Concepts.* Chapman and Hall, London.

[57] Jogdeo, K. (1982), Concepts of dependence. In *Encyclopedia of Statistical Sciences*, **1**, S.Kotz and N.L. Johnson, editors (John Wiley & Sons, New York), 324-334.

[58] Johnsonbaugh, R., Pfaffenberger, W.E. (1981), *Foundations of Mathematical Analysis*. Marcel Dekker, Inc., New York.

[59] Johnson, M.E. (1987), *Multivariate Statistical Simulation*. Wiley, New York.

[60] Johnson, M.E., Wang, C., Ramberg, J. (1984), Generation of continuous multivariate distributions for statistical applications. *American Journal of Mathematical and Management Sciences*, **4**, 225-248.

[61] Kac, M, Siegert, A.J.F. (1947), An explicit representation of a stationary Gaussian process. *Annals of Mathematical Statistics*, **18**, 438-442.

[62] Kallenberg, W.C.M., Ledwina, T. (1999), Data-driven tests for independence. *Journal of the American Statistical Association* **94**, 285 - 301.

[63] Kanwall, R. (1971), *Linear Integral Equations Theory and Technique.* Academic Press, New York.

[64] Kendall, M.G. (1938), A new measure of rank correlation. *Biometrika*, **30**, 81-93.

[65] Kimeldorf, G., Sampson, A. (1975), One-parameter families of bivariate distributions with fixed marginals. *Communications in Statistics A-Theory Methods*, **4**(3), 293-301.

[66] Kimeldorf, G., Sampson, A. (1975), Uniform representations of bivariate distributions. *Communications in Statistics A-Theory Methods*, **4**(7), 617-627.

[67] Kimeldorf, G., Sampson, A. (1978), Monotone dependence. *The Annals of Statistics*, **6**(4), 895-903.

[68] Kotz, S. (1975), Multivariate distributions at a cross-road. In: *A modern course on Distributions in Scientific Work*. Vol. **1**- *Models and Structures*, G.P. Patil, S. Kotz, and J.K. Ord (editors), Dordrecht: Reidel, 247-270.

[69] Kowalski, C.J. (1973), Non-Normal Bivariate Distributions with Normal Marginals. *The American Statistician,* **27**(3), 103-106.

[70] Lancaster, H.O. (1957), Some properties of the bivariate normal distribution considered in the form of a contingency table. *Biometrika*, **44**, 289-292.

[71] Lancaster, H.O. (1958), The Structure of Bivariate Distributions. *Annals of Mathematical Statistics*, **29**, 719-736.

[72] Lancaster, H.O. (1963), Correlations and Canonical Forms of Bivariate Distributions. *Annals of Mathematical Statistics*, **34**, 532-538.

[73] Lancaster, H.O. (1963), Correlation and Complete Dependence of Random Variables. *Annals of Mathematical Statistics*, **34**, 1315-1321.

[74] Lancaster, H.O. (1969), *The Chi-squared Distribution.* Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Inc.

[75] Lancaster, H.O. (1982), Dependence, measures and indices of. In *Encyclopedia of Statistical Sciences*, **2**, S.Kotz and N.L. Johnson, editors (John Wiley & Sons, New York), 334-339.

[76] Lee, M.-L. T. (1996), Properties and Applications of the Sarmanov Family of Bivariate Distributions. *Communications in Statistics -Theory Methods*, **25**(6), 1207-1222.

[77] Lehmann, E. L. (1966), Some concepts of dependence. *Ann. Math. Statist.* **37**, 1137-1153.

[78] Mardia, K.V. (1967), *Families of bivariate distributions*, Charles Griffin and Co., London.

[79] Mardia, K.V. (1970), A translation family of bivariate distributions and Fréchet's bounds. *Sankhyā Ser. A*, **32,** 119-22.

[80] Mardia, K.V., Thompson, J.W. (1972), Unified treatment of moment-formulae, *Sankhyā Ser. A*, **34**, 121-132.

[81] Mardia, K.V., Kent, J.T., Bibby, J.M. (1979), *Multivariate Analysis.* Probability and Mathematical Statistics, Academic Press, London.

[82] Marshall, A.W., Olkin, I. (1967), A multivariate exponential distribution. *Journal of the American Statistical Association*, **62**, 30-44.

[83] Marshall, A.W., Olkin, I. (1967), A generalized bivariate exponential distribution. *Journal of Applied Probability*, **4**, 291-302.

[84] Morgenstern, D. (1956), Einfache Beispiele Zweidimensionaler Verteilungen. *Mitteilingsblatt für Mathematische Statistik* **8**, 234-235.

[85] Nelsen, R.B. (1986), Properties of a one-parameter family of bivariate distributions with specified marginals, *Communications in Statistics - Theory Methods*, **15**(11), 3277-3285.

[86] Nelsen, R.B. (1987), Discrete bivariate distributions with given marginals and correlation. *Communications in Statistics - Simulation*, **16**(1), 199-208.

[87] Nelsen, R.B. (1989), Some two-parameter families of bivariate distributions with arbitrary margins. In: *Abstracts of papers presented to the American Mathematical Society*, **10**, 115.

[88] Nelsen, R.B., Quesada Molina, J.J., Rodríguez Lallena, J.A., Ubeda Flores, M. (2004), Best-possible bounds on sets of bivariate distributions. *Journal of Multivariate Analysis*, **90**, 348-358.

[89] Nelsen, R.B. (1999), *An Introduction to Copulas*. Lecture Notes in Statistics 139, Springer-Verlag, New York.

[90] Nualart, D., Sanz, M. (1990), *Curs de Probabilitats*. Promociones y Publicaciones Universitarias, (in Catalan).

[91] Quesada Molina, J.J., Rodríguez Lallena, J.A., (1995), Bivariate copulas with quadratic sections. *J. Nonparamtr. Statist.*, **5**, 323-337.

[92] Rao, C.R. (1964), The use and interpretation of principal component analysis in applied research. *Sankhya. The Indian Journal of Statistics, Series A*, **26**, 329-358.

[93] Rényi, A. (1959), On measures of dependence. *Acta Mathematica Academiae Scientiarum Hungaricae*, **10**, 441-451.

[94] Sarmanov, I.O. (1974), New forms of correlation relationships between positive quantities applied in hydrology. In: *Mathematical Models in Hydrology Symposium,* 104-109. *IAHS Publication No. 100, International Association of Hydrological Sciences.*

[95] Scarsini, M. (1984), On measures of concordance. *Stochastica*, **8**(3), 201-218.

[96] Schoenberg, I.J. (1935), Remarks to Maurice Fréchet's article "Sur la définition axiomatique d'une classe d'espaces distanciés vectoriellement applicables sur l'espace de Hilbert". *The Annals of Mathematics*, **36**, 724-732.

[97] Schucany, W.R., Parr, W.C., Boyer, J.E. (1978), Correlation structure in Farlie-Gumbel-Morgenstern distributions. *Biometrika*, **65**, 650-653.

[98] Schweizer, B., Wolf, E.F. (1981), On Nonparametric Measures of Dependence for Random Variables. *The Annals of Statistics*, **9(**4), 879-885.

[99] Shorack, G.R., Wellner, J.A. (1986), *Empirical processes with applications to statistics.* John Wiley & Sons.

[100] Sklar, A. (1959), Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris* **8**, 229-231.

[101] Spearman, C. (1904), The proof and measurement of association between two things. *Amer. J. Psych.*, **15**, 72-101.

[102] Strang, G. (1988), *Linear Algebra and its applications*, 3rd ed. San Diego, Harcourt Brace Jovanovich.

[103] Thomas, J.B. (1978), Representation of bivariate distribution with applications. In: *Communication Systems and Random Processes Theory. Alphen aan der Rijn: Sijthoff and Noordhoff.* J.K. Skwirzynski (Editor), 541-560.

[104] Tiit, E.-M. (1986), Random vectors with given arbitrary marginals and given correlation matrix. *Acta et Commentationes Universitatis Tartuensis* **733,** 14-39.

[105] Torgerson, W.S. (1952), Multidimensional Scaling: 1. Theory and Methods. *Psychometrika*, **17**, 401-419.

Resum/Summary

Introducció

L'interés per comprendre les lleis que regeixen els fenomens de l'univers podríem dir que és una característica de la naturalesa racional de l'home. Partint de la simple observació del fenomen, la raó procura esbrinar el cóm, el quan i el perquè es produeix. El coneixement de les lleis que regeixen la Natura permet a l'home exercir un cert domini sobre aquésta. L'home pot volar, per exemple, gràcies a haver conegut les lleis de la gravitació universal, de la mecànica, de la hidrodinàmica, etc. Aquestes lleis són molt complexes perquè la realitat ho és: la caiguda lliure d'un cos que es troba a una certa alçada ve determinada per la força d'atracció que exerceix la terra sobre aquést; l'energia, lligada a la força d'un cos que es desplaça, la podem calcular mitjançant el producte de la massa d'aquest cos, l'alçada a la qual es troba en la seva caiguda i l'acceleració gravitatòria (els físics ens han dit que val 9.8 m/s^2). Si tenim en compte el fregament de l'aire que s'oposa a la caiguda lliure del cos, els càlculs es compliquen, però s'ajusten més a la realitat. La caiguda lliure d'un cos es regeix per lleis majoritàriament "deterministes". Sabem amb total certeza que si deixem anar la poma que portem a la ma, aquésta es desplaçarà fins a terra amb un moviment uniformement accelerat; i això passarà sempre que repetim l'experiència de deixar anar la poma.

D'altres fenomens, en canvi, no observen aquest comportament. No totes les pomes d'un arbre assoleixen el mateix pes malgrat créixer a la mateixa branca, haver rebut el mateix sol, etc. La variabilitat del pes d'una poma ens adverteix que som davant de fenomens de naturalesa diferent, no determinista. Abans de veure la poma i posar-la sobre una bàscula no podem saber quan pesa; en canvi, estem segurs que caurà si la deixem anar a l'aire. Diem que el Pes és una *variable aleatòria* (en el sentit de l'Estadística). Si hem estat pesant la collita sencera d'un camperol i hem pogut saber el pes de totes i cadascuna de les pomes, i hem averiguat que el pes mitjà ha estat de 200 grams, allò que és raonable esperar és que una poma qualsevol pesi uns 200 grams. La freqüència amb la qual es presenten uns valors de pes concret ens fa intuïr allò que és més "probable". Cal estudiar la *llei* (de probabilitat) que regeix cada fenomen, deixant de banda d'altres variables que complicarien excessivament el model. Una vegada caracteritzada la llei de probabilitat d'una sola variable, podrem calcular la probabilitat que la variable prengui un valor concret, o valors menors o igual que un valor donat. La funció que assigna a qualsevol nombre real, diguem-li x, la probabilitat que una variable (per exemple, el Pes) prengui valors menors o iguals que x s'anomena *funció de distribució* de la variable. El comportament en probabilitat de les variables aleatòries ve perfectament explicat per llur funció de distribució. Si assignem el valor 0 al resultat de sortir cara en llençar una moneda, i 1 si surt creu, la funció de distribució val 0 per tota $x < 0$, val $\frac{1}{2}$ si $0 \leq x < 1$, i val 1 si $x \geq 1$. Si assignem el valor 0 al resultat de sortir parell en llençar un dau, i 1 si surt senar, la funció de distribució val exactament el mateix que en l'exemple anterior. S'han trobat models generals als quals s'ajusten moltes variables des d'un punt de vista experimental. Així, s'ha descrit la funció de distribució de probabilitat de variables que prenen valors en un conjunt numerable o finit (com les dels darrers exemples) i també de variables que s'anomenen contínues perquè poden prendre infinits valors dins d'un conjunt no numerable (i llur funció de distribució també és contínua). La més coneguda és la llei Normal, i va ser la distribució més utilitzada fins aproximadament l'any 1930. Després, en créixer la recerca en àmbits on predominaven dades amb un comportament que difícilment podia ajustar-se al model Normal, es van obtenir les funcions de distribució d'altres lleis univariants (uniforme, Cauchy, Gamma, Laplace, Pareto, Weibull, etc.). Les variables amb un comportament en probabilitat segons algun d'aquests models tenen totes la mateixa forma en la seva distribució, i només es diferencien per la localització d'aquestes dades, o la seva dispersió en torn a la mitjana.

Una vegada l'estudi de les distribucions univariants ha assolit un cert nivell, la necessitat de comprendre millor les interrelacions entre les variables porta a que s'estudïi el model de probabilitat d'un conjunt de variables. Si es tracta de només dues variables aleatòries X i Y, es parla de la *distribució bivariant del vector aleatori* (X, Y). La funció de distribució bivariant modela d'alguna manera la relació de dependència entre dues variables. La mesura de dependència més utilitzada és la *covariança* entre les variables. Aquésta es defineix com l'esperança del producte de les diferències entre cada variable i la seva esperança. Aquesta mesura de dependència depén de les unitats de cada variable, i s'estandarditza dividint per l'arrel quadrada del producte de les covariances de cadascuna de les variables amb ella mateixa (la variança de cada variable). Aquést és l'anomenat *coeficient de correlació de Pearson*. L'índex obtingut és adimensional, pren valors entre -1 i 1, assoleix aquests màxims quan hi ha una relació lineal entre les variables i per això es considera una mesura de dependència lineal. D'altres tipus de dependència tenen molt d'interés: la dependència que poden presentar variables que només prendran dos possibles valors, i el comportament en probabilitat de les quals s'estudia en una taula de freqüències (o de contingència) 2×2; el cas d'*independència estocàstica*, en el qual la probabilitat d'una variable no s'afecta en absolut per la de l'altra; o el concepte de *dependència positiva* (o *negativa*), és a dir, que valors "grans" d'una variable tendeixen a anar units (probabilísticament) a valors grans (o petits) de l'altra variable.

Tal i com va passar amb les distribucions univariants, durant molt de temps el model bivariant que ha predominat ha estat el Normal Bivariant. Les característiques d'aquest model el fan molt fàcil d'utilitzar, i bona part de les tècniques estadístiques d'anàlisi assumeixen normalitat (uni o bivariant) en les dades. Malgrat això, aquesta assumpció és molt difícil de verificar i, fins i tot, de sostenir en moltes situacions experimentals en les quals apareixen dades clarament no normals. Actualment s'han descrit molts models multivariants, els quals ajusten bé dades experimentals, tal com han pogut comprovar diversos investigadors. Els primers treballs en els quals es construeixen distribucions bivariants, fixant les distribucions de cada variable per separat (és a dir, les marginals) són de Hoeffding (1940, [52]) i Fréchet (1951, [42]); aquest darrer troba les fites superior i inferior per la classe (el conjunt) de totes les distribucions bivariants amb les mateixes marginals. Aquesta classe rep el nom de classe de Fréchet. Kimeldorf i Sampson (1975, [65]) proposen cinc condicions que hauria de satisfer qualsevol família uniparamètrica de distribucions. Hi ha moltes famílies, a més de la Normal Bivariant, satisfent aquestes condicions, i s'utilitzen en molts camps d'experimentació: la família Farlie-Gumbel-Morgenstern (FGM), Ali-Mikhail-Haq, Frank, Cuadras-Augé, etc. Donada una família paramètrica de distribucions, cal conéixer quin és el rang de dependència que cobreix. També interessa saber si la dependència entre les variables augmenta quan ho fa el valor del paràmetre.

L'estudi de la distribució conjunta de dues variables, per tant, pertany a l'àmbit de l'estudi de les relacions multivariants. L'objecte principal de l'Estadística és establir conclusions sobre fenòmens poblacionals a partir de mostres de dades, amb una certa confiança. Aleshores, plantejarem models poblacionals. Amb aquesta perspectiva, ens hem plantejat l'estudi de la dependència entre dues variables aleatòries a partir de la covariança als diferents nivells en els quals es treballa en Estadística. En primer lloc, una mostra de cada variable aleatòria ens permet obtenir una matriu de dades amb p columnes (una per cada variable), i on cada fila conté les dades d'un sol individu; la matriu té un nombre finit de files, tantes com individuus formen la mostra. Si estudiem tota la població i la variable pot prendre valors en un conjunt no numerable (continu), tindrem una matriu infinita o, millor, una funció de dues variables o nucli. En el cas que tinguem dues matrius obtingudes del mateix conjunt d'individuus, o bé, dos nuclis definits sobre el mateix espai en el cas continu, podem estudiar la relació entre aquestes dues matrius, o entre aquests dos nuclis, mitjançant alguna operació que ens permeti obtenir la seva covariança. A un nivell diferent, podem estudiar la covariança entre dues variables aleatòries i, també, entre qualsevol parella de funcions d'aquestes variables. Naturalment, la dependència entre aquests objectes (matrius, nuclis, funcions de variables aleatòries) està inserida en la funció de distribució conjunta de les dues variables. Si hem estat capaços de construïr una família de distribucions bivariants, amb un o més paràmetres multivariants, haurem d'estudiar si aquest paràmetre és veritablement un paràmetre de dependència, els seus possibles valors, etc.

Aquésta ha estat la motivació d'aquesta tesi. A continuació presentem un resum dels principals resultats d'aquest treball, seguint l'ordre que acabem d'introduïr. Hem treballat amb matrius, nuclis i funcions de variables aleatòries lligades per una distribució conjunta que té la propietat adicional de ser simètrica i amb dependència quadrant positiva (PQD); aleshores, les matrius estudiades són simètriques i semidefinides positives, els nuclis també ho són, i les propietats que s'obtenen de l'estudi de la covariança entre funcions i també la mesura d'afinitat que hem definit a partir de la covariança, són propietats vàlides en el cas que la distribució conjunta sigui simètrica i PQD. Finalment, hem construït distribucions bivariants simètriques a partir de les expansions diagonals, i hem assumit que les correlacions canòniques també son no negatives. En finalitzar la presentació dels Resultats donarem les conclusions principals que se n'han derivat.

Resultats

S'estudia, en primer lloc, la dependència entre una classe de matrius, mitjançant la definició d'unes operacions *unió* i *intersecció*. Extendrem al cas continu aquests mateixos conceptes, definint un producte entre funcions de dues variables (nuclis) i les operacions *unió* i *intersecció*. Veurem que el producte de dos nuclis és, en particular, la covariança entre dues funcions, i estudiarem la covariança entre funcions de variables aleatòries i una mesura d'afinitat. Finalment, aplicarem el métode de les expansions diagonals de Lancaster per construïr distribucions bivariants simètriques utilitzant com a variables canòniques les components principals de les variables (aquéstes s'han pogut obtenir en quatre casos: uniforme, exponencial, logística i Pareto). Estudiem el rang de dependència que cobreix cadascuna de les famílies construïdes. Per tal d'agilitzar la lectura d'aquest resum, ens remetem al text original per les proves completes als resultats presentats.

Capítol 2. Dependència entre una classe de matrius

Considerem el conjunt $\mathcal{M}_n = \{\mathbf{A} \mid \mathbf{A}$ matriu simètrica $n \times n$ (s.)d.p.$\}$. És a dir, $\mathbf{A} = \mathbf{A}'$ (la matriu trasposada d'$\mathbf{A}$) i tots els valors propis són estrictament positius (definida positiva, d.p.) o al menys n'hi ha un de nul (semidefinida positiva, s.d.p.). Siguin $\mathbf{X}$ i $\mathbf{Y}$ matrius $n \times p$ amb $p \leq n$, i definim $\mathbf{A} = \mathbf{XX}'$, $\mathbf{B} = \mathbf{YY}'$. Aleshores $\mathbf{A}, \mathbf{B} \in \mathcal{M}_n$. Per la simetria d'$\mathbf{A}$ i de $\mathbf{B}$, existeixen (teorema de descomposició espectral) matrius ortogonals $\mathbf{U}, \mathbf{V}$ i matrius diagonals $\mathbf{\Lambda} = diag\,(\lambda_1, ..., \lambda_n)$, $\mathbf{M} = diag\,(\mu_1, ..., \mu_n)$, tals que

$$\mathbf{A} = \mathbf{U\Lambda U}' = \sum_{i=1}^{n} \lambda_i \mathbf{u}_i \mathbf{u}_i', \qquad \mathbf{B} = \mathbf{VMV}' = \sum_{i=1}^{n} \mu_i \mathbf{v}_i \mathbf{v}_i',$$

on $\mathbf{u}_i$ i $\mathbf{v}_i$, $i = 1, ..., n$ són les columnes ortonormals d'$\mathbf{U}$ i de $\mathbf{V}$, respectivament. Per ser s.d.p., existeixen les arrels racionals de la matriu, diguem-li $\mathbf{A}^{\frac{r}{s}}$, $r < s$, per qualssevol enters positius r, s. En particular, l'arrel quadrada es pot expressar com

$$\mathbf{A}^{\frac{1}{2}} = \mathbf{U\Lambda}^{\frac{1}{2}}\mathbf{U}' = \sum_{i=1}^{n} \lambda_i^{\frac{1}{2}} \mathbf{u}_i \mathbf{u}_i'.$$

Les operacions unió i intersecció de matrius

La intersecció $\mathbf{A} \wedge \mathbf{B}$ i la unió $\mathbf{A} \vee \mathbf{B}$ d'$\mathbf{A}, \mathbf{B} \in \mathcal{M}_n$ es defineixen com

$$\begin{aligned} \mathbf{A} \wedge \mathbf{B} &= \frac{1}{2}\left(\mathbf{A}^{\frac{1}{2}}\mathbf{B}^{\frac{1}{2}} + \mathbf{B}^{\frac{1}{2}}\mathbf{A}^{\frac{1}{2}}\right), \\ \mathbf{A} \vee \mathbf{B} &= \mathbf{A} + \mathbf{B} - \mathbf{A} \wedge \mathbf{B}. \end{aligned}$$

La següent proposició resumeix les propietats d'aquestes operacions.

Proposició 1 *Siguin $\mathbf{A}, \mathbf{B} \in M_n$ i sigui $\mathbf{0}$ la matriu de zeros, $n \times n$. La intersecció i la unió d'$\mathbf{A}$ i $\mathbf{B}$ són també matrius $n \times n$, simètriques, i la unió és, a més, (semi)definida positiva. Les propietats són:*

1. *Commutativitat:*

$$\mathbf{A} \wedge \mathbf{B} = \mathbf{B} \wedge \mathbf{A}, \qquad \mathbf{A} \vee \mathbf{B} = \mathbf{B} \vee \mathbf{A}.$$

2. *Ortogonalitat:* $\mathbf{AB} = \mathbf{0}$ *si, i només si (sii)*

$$\mathbf{A} \wedge \mathbf{B} = \mathbf{0}, \qquad \mathbf{A} \vee \mathbf{B} = \mathbf{A} + \mathbf{B}.$$

3. *Igualtat: si* $\mathbf{A} = \mathbf{B}$ *aleshores*

$$\mathbf{A} \wedge \mathbf{A} = \mathbf{A}, \qquad \mathbf{A} \vee \mathbf{A} = \mathbf{A}.$$

4. *Element nul: la matriu de zeros és l'element nul per a la intersecció i el neutre per a la unió:*

$$\mathbf{A} \wedge \mathbf{0} = \mathbf{0}, \qquad \mathbf{A} \vee \mathbf{0} = \mathbf{A}.$$

Una relació d'equivalència

Siguin $\mathbf{A}, \mathbf{B} \in \mathcal{M}_n$ i $\mathbf{u} \in \mathbb{R}^n$, un vector tal que

$$\mathbf{Au} = \lambda \mathbf{u}, \qquad \mathbf{Bu} = \mu \mathbf{u},$$

per alguns $\lambda, \mu \in \mathbb{R}$. Direm que $\mathbf{u}$ es un vector propi comú a $\mathbf{A}$ i a $\mathbf{B}$, i es comprova que $\mathbf{A} \wedge \mathbf{B}$ i $\mathbf{A} \vee \mathbf{B}$ tenen el mateix vector propi $\mathbf{u}$, i que $\lambda^{\frac{1}{2}} \mu^{\frac{1}{2}}$ i $\lambda + \mu - \lambda^{\frac{1}{2}} \mu^{\frac{1}{2}}$ són els corresponents valors propis d'$\mathbf{A} \wedge \mathbf{B}$ i $\mathbf{A} \vee \mathbf{B}$.

Suposem que $\mathbf{A}, \mathbf{B}$ comparteixen tots els vectors propis, és a dir existeixen $\mathbf{\Lambda}, \mathbf{M}$ matrius diagonals dels valors propis tals que $\mathbf{A} = \mathbf{U \Lambda U}'$, i $\mathbf{B} = \mathbf{UMU}'$. Podem definir les següents relacions d'equivalència. Siguin $\mathbf{A}, \mathbf{B} \in \mathcal{M}_n$, amb vectors propis normalitzats $\{\mathbf{u}_i\}_{i=1}^n$i $\{\mathbf{v}_i\}_{i=1}^n$, respectivament. Diem que $\mathbf{A}$ és equivalent a $\mathbf{B}$ sii $\{\mathbf{u}_i\}_{i=1}^n = \{\mathbf{v}_i\}_{i=1}^n$. Escriurem $\mathbf{A} \sim \mathbf{B}$. Indicarem la classe per

$$[\mathbf{U}] = \{\mathbf{M} \in \mathcal{M}_n \mid \mathbf{M} = \mathbf{U \Lambda U}', \text{ per alguna matriu diagonal } \mathbf{\Lambda} \geq 0\}.$$

El resultat obtingut per un sol vector propi, aplicat a tots els vectors propis $\mathbf{u}_i$, de dues matrius $\mathbf{A}, \mathbf{B} \in [\mathbf{U}]$, ens permet caracteritzar els valors propis de les matrius unió i intersecció. En efecte, siguin $\mathbf{A} = \mathbf{U \Lambda U}'$ i $\mathbf{B} = \mathbf{UMU}'$ matrius de $\mathcal{M}_n$ amb

$\mathbf{A} \sim \mathbf{B}$, i $\mathbf{U}$, la matriu ortogonal (comuna) dels vectors propis, amb $\mathbf{\Lambda}$ i $\mathbf{M}$ diagonals. Aleshores

$$\mathbf{A} \wedge \mathbf{B} = \mathbf{U\Gamma U}', \quad \mathbf{A} \vee \mathbf{B} = \mathbf{U\Sigma U}',$$

on $\mathbf{\Gamma} = \mathbf{\Lambda}^{\frac{1}{2}}\mathbf{M}^{\frac{1}{2}}$ i $\mathbf{\Sigma} = \mathbf{\Lambda} + \mathbf{M} - \mathbf{\Lambda}^{\frac{1}{2}}\mathbf{M}^{\frac{1}{2}}$ són diagonals.

Una relació d'ordre parcial

Siguin $\mathbf{A}, \mathbf{B} \in \mathcal{M}_n$, $\mathbf{A} \sim \mathbf{B}$, amb valors propis $\{\lambda_i\}_{i=1}^n$ i $\{\mu_i\}_{i=1}^n$, respectivament, i amb matriu comuna de vectors propis $\mathbf{U}$. Direm que $\mathbf{A}$ és més petita que $\mathbf{B}$, i notarem aquest ordre parcial per $\mathbf{A} \lesssim \mathbf{B}$, si $\lambda_i \leq \mu_i$, per tota $i = 1, ..., n$, on l'índex i es refereix a l'$i-$èssim vector propi.

Com a conseqüència immediata es té el següent resultat: siguin $\mathbf{A}, \mathbf{B} \in \mathcal{M}_n$, amb $\mathbf{A} \sim \mathbf{B}$ i $\mathbf{A} \lesssim \mathbf{B}$. Aleshores
(i) $\mathbf{A} \wedge \mathbf{B} \sim \mathbf{A}$ $(\sim \mathbf{B})$ i també $\mathbf{A} \vee \mathbf{B} \sim \mathbf{A}$ $(\sim \mathbf{B})$,
(ii) $\mathbf{A} \lesssim \mathbf{A} \wedge \mathbf{B} \lesssim \mathbf{A} \vee \mathbf{B} \lesssim \mathbf{B}$.

La prova es deriva de les desigualtats $\lambda_i \leq \lambda_i^{\frac{1}{2}}\mu_i^{\frac{1}{2}} \leq \lambda_i + \mu_i - \lambda_i^{\frac{1}{2}}\mu_i^{\frac{1}{2}} \leq \mu_i$, que són els valors propis de $\mathbf{A}$, $\mathbf{A} \wedge \mathbf{B}$, $\mathbf{A} \vee \mathbf{B}$ i $\mathbf{B}$, respectivament.

Finalment, s'obtenen alguns resultats sobre la traça de les matrius ordenades. La traça d'una matriu quadrada (amb el mateix nombre de files que de columnes) és la suma dels elements de la seva diagonal (si $\mathbf{A} = (a_{ij})_{i,j=1,...,n}$, $tr(\mathbf{A}) = \sum_{i=1}^n a_{ii}$). En una matriu simètrica, es comprova que $tr(\mathbf{A}) = \sum_{i=1}^n \lambda_i$, éssent $\{\lambda_i\}$ el conjunt de valors propis d'$\mathbf{A}$. Cal observar que tots els valors propis són no negatius, i d'aquest fet es deriven fàcilment els resultats presentats:

Proposició 2 *Siguin* $\mathbf{A}, \mathbf{B} \in \mathcal{M}_n$, $\mathbf{A} \sim \mathbf{B}$, *tals que* $\mathbf{A} = \mathbf{U\Lambda U}'$, $\mathbf{B} = \mathbf{UMU}'$, *on* $\mathbf{\Lambda} = diag(\lambda_i)$, $\mathbf{M} = diag(\mu_i)$. *Aleshores:*
(a) $tr(\mathbf{A} \wedge \mathbf{B}) = \sum_{i=1}^n \lambda_i^{\frac{1}{2}}\mu_i^{\frac{1}{2}}$,
(b) $tr(\mathbf{A} \vee \mathbf{B}) = \sum_{i=1}^n \left(\lambda_i + \mu_i - \lambda_i^{\frac{1}{2}}\mu_i^{\frac{1}{2}}\right)$.

La prova s'obté dels valors propis de les matrius simètriques $\mathbf{A} \wedge \mathbf{B}$, $\mathbf{A} \vee \mathbf{B}$. Observem que si $\mathbf{A} \lesssim \mathbf{B}$ són matrius de $\mathcal{M}_n$, aleshores, $tr(\mathbf{A}) \leq tr(\mathbf{B})$. Per tant es prova que

$$tr(\mathbf{A}) \leq tr(\mathbf{A} \wedge \mathbf{B}) \leq tr(\mathbf{A} \vee \mathbf{B}) \leq tr(\mathbf{B}).$$

Aquests resultats per matrius simètriques, semidefinides positives en general coincideixen amb els obtinguts a [28] per les matrius de productes creuats, d'aquestes mateixes característiques, associades a dues matrius de distàncies obtingudes del mateix conjunt de n objectes, mitjançant la definició d'una distància conjunta. Aquesta tècnica és el RMS (Related Metric Scaling). Així, hem establert una base teòrica que podria ser útil en d'altres tècniques d'Anàlisi Multivariant basada en Distàncies, així com en la interpretació geomètrica del RMS.

Capítol 3. Extensions al continu

Presentem les operacions unió i intersecció entre nuclis simètris semidefinits positius, establint la base per l'extensió al continu del Related Metric Scaling.

Calen algunes nocions preliminars. Siguin $(\Omega_1, \mathcal{F}_1, \mathcal{M}_1)$ i $(\Omega_2, \mathcal{F}_2, \mathcal{M}_2)$ dos espais de mesura. Una funció de dues variables a valors reals, $K(s,t)$, mesurable a l'espai producte $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \times \mathcal{F}_2)$ s'anomena nucli. Es diu que K és simètric si per a cada parella $(x,y) \in \Omega_1 \times \Omega_2$, $K(x,y) = K(y,x)$.

Si K és un nucli (simètric) de Hilbert-Schmidt, i.e.,

$$\int_a^b \int_a^b K^2(s,t)\, dF(s)\, dF(t) < \infty,$$

existeix una expansió en funcions pròpies de K, convergent en mitjana quadràtica respecte de la mesura producte $dF(s)\, dF(t)$. Això és, existeix una base $\{\xi_i\}$, completa i ortonormal a $L^2([a,b], F)$, tal que

$$K(x_1, x_2) = \sum_{i=0}^{\infty} \lambda_i \xi_i(x_1)\, \xi_i(x_2), \qquad (i)$$

on els conjunts numerables $\{\lambda_i\}, \{\xi_i\}$ de valors i funcions pròpies de K sobre F satisfan

$$\int_a^b \xi_i(x_1)\, K(x_1, x_2)\, dF(x_1) = E[\xi_i(X_1)\, K(X_1, x_2)] = \lambda_i \xi_i(x_2).$$

Direm que un nucli K es s.d.p. sii tots els seus valors propis no nuls són estrictament positius. Direm que és d.p. si tots els seus valors propis són estrictament més grans que 0 i els corresponents vectors propis formen un sistema ortonormal complet (vegi's [63]). Denotarem el conjunt dels nuclis continuus, simètrics i (s.)d.p. a $\mathcal{X} \times \mathcal{X}$ per $\mathcal{K}_{\mathcal{X}\times\mathcal{X}}$. El Teorema de Mercer (vegi's [11]) estableix que els nuclis d'aquest conjunt es poden expandir com a (i) i que la sèrie convergeix absoluta i uniformement.

Introduïm una classe de nuclis que apareix en el context de l'estudi assimptòtic dels U-estadístics. Un nucli K, simètric, es diu degenerat si l'esperança matemàtica

$$E_Y(K(x,Y)) = \int_a^b K(x,y)\, dF(y)$$

és constant gairebé per tota x (g.p.t. x).

Com que treballem amb nuclis simètrics, també es pot definir per $E_X(K(X,y)) = c$, amb c constant, g.p.t. y.

Un nucli $G \in \mathcal{K}_{\mathcal{X}\times\mathcal{X}}$ és doblement centrat respecte d'una distribució F si per tota parella $x, x' \in \mathcal{X}$:

$$\begin{aligned} E_X\left(G\left(X,x'\right)\right) &= \int_{\mathcal{X}} G\left(x,x'\right) dF\left(x\right) = 0, \\ E_{X'}\left(G\left(x,X'\right)\right) &= \int_{\mathcal{X}} G\left(x,x'\right) dF\left(x'\right) = 0. \end{aligned}$$

Sigui (X,Y) un vector aleatori definit a $\mathcal{X}\times\mathcal{Y}$, amb funció de distribució conjunta H i marginals F, G, respectivament. Així, H pertany a la classe de Fréchet $\mathcal{F}(F,G)$. Notarem $(X,Y) \sim H$. Definim a continuació el producte dels nuclis $K_1 \in \mathcal{K}_{\mathcal{X}\times\mathcal{X}}$ i $K_2 \in \mathcal{K}_{\mathcal{Y}\times\mathcal{Y}}$, que, com es veurà, és la base de les operacions unió i intersecció.

Definició 3 *Sigui $H \in \mathcal{F}(F,G)$ una funció de distribució bivariant absolutament contínua respecte de la mesura producte FG. Siguin $K_1 \in \mathcal{K}_{\mathcal{X}\times\mathcal{X}}$, $K_2 \in \mathcal{K}_{\mathcal{Y}\times\mathcal{Y}}$. L' $H-$producte $K_1 \star K_2$ és un nucli definit a $\mathcal{X}\times\mathcal{Y}$ com*

$$\left(K_1 \star K_2\right)_H(x,y) = Cov\left(K_1(x,X), K_2(Y,y)\right), \qquad (ii)$$

per tota parella $(x,y) \in \mathcal{X}\times\mathcal{Y}$, amb $(X,Y) \sim H$.

Teorema 4 *Siguin $K_1 \in \mathcal{K}_{\mathcal{X}\times\mathcal{X}}$, $K_2 \in \mathcal{K}_{\mathcal{Y}\times\mathcal{Y}}$ dos nuclis que satisfans les condicions de la definició del producte (ii), i considerem llur expansions en funcions i valors propis*

$$K_1 = \sum_{i=0}^{\infty} \lambda_i \xi_i \otimes \xi_i, \qquad K_2 = \sum_{i=0}^{\infty} \mu_i \eta_i \otimes \eta_i,$$

on $\xi_i \otimes \xi_i$ expressa $\xi_i(x)\,\xi_i(x')$ per cada parella $(x,x') \in \mathcal{X}\times\mathcal{X}$ i anàlogament, $\eta_i \otimes \eta_i$. Aleshores, donat un vector aleatori $(X,Y) \sim H$, definit a $\mathcal{X}\times\mathcal{Y}$,

$$K_1 \star K_2 = \sum_{i,j=0}^{\infty} \lambda_i \mu_j Cov\left(\xi_i(X), \eta_j(Y)\right) \xi_i \otimes \eta_j.$$

La prova d'aquest teorema es deriva de l'ortonormalitat dels conjunts $\{\xi_i\}$, $\{\eta_i\}$, el teorema de la Convergència Dominada, que ens permet intercanviar la integral amb les sumes infinites i, el teorema de Fubini, en el càlcul de la covariança (ii).

Les operacions unió i intersecció de nuclis

Sigui $K \in \mathcal{K}_{\mathcal{X}\times\mathcal{X}}$ un nucli doblement centrat, amb $K = \sum_{i=1}^{\infty} \lambda_i \xi_i \otimes \xi_i$ per tota parella $(x, x') \in \mathcal{X} \times \mathcal{X}$. L'arrel quadrada de K, que notarem $K^{\frac{1}{2}}$, és el nucli de $\mathcal{K}_{\mathcal{X}\times\mathcal{X}}$ definit per $K^{\frac{1}{2}} = \sum_{i=1}^{\infty} \lambda_i^{\frac{1}{2}} \xi_i \otimes \xi_i$, amb $\sum_{i=1}^{\infty} \lambda_i < \infty$. Qualsevol arrel racional $K^{\frac{m}{n}}$, $m < n$ es defineix anàlogament.

Les definicions de l'$H-$producte i l'arrel quadrada ens permeten definir les operacions unió i intersecció entre nuclis.

Siguin $K_1, K_2 \in \mathcal{K}_{\mathcal{X}\times\mathcal{X}}$. Sigui $H \in \digamma(F, F)$ una funció de distribució simètrica, amb rang de les variables marginals igual a $\mathcal{X}$. La intersecció $K_1 \wedge K_2$ i la unió $K_1 \vee K_2$ de $K_1, K_2 \in \mathcal{K}_{\mathcal{X}\times\mathcal{X}}$ es defineixen com

$$
\begin{aligned}
K_1 \wedge K_2 &= \frac{1}{2}\left(K_1^{\frac{1}{2}} \star K_2^{\frac{1}{2}} + K_2^{\frac{1}{2}} \star K_1^{\frac{1}{2}}\right), \\
K_1 \vee K_2 &= K_1 + K_2 - K_1 \wedge K_2,
\end{aligned}
$$

on $\star$ és l'$H-$producte (ii).

És obvi que $K_1 \wedge K_2$, $K_1 \vee K_2$ també són nuclis simètrics a $\mathcal{X} \times \mathcal{X}$. La següent proposició resumeix les propietats d'aquestes operacions.

Proposició 5 *Siguin $K_1, K_2 \in \mathcal{K}_{\mathcal{X}\times\mathcal{X}}$ i sigui K_0 el nucli que val 0 g.p.t. $(x, y) \in \mathcal{X} \times \mathcal{X}$, i.e., $K_0(x, y) = 0$. L'unió $K_1 \vee K_2$ és un nucli simètric semidefinit positiu. En canvi, $K_1 \wedge K_2$ és simètric, però no és s.d.p., en general. Les operacions intersecció i unió satisfan les propietats següents:*

1. *Commutativitat: per tota $H \in \digamma(F, F)$*

$$K_1 \wedge K_2 = K_2 \wedge K_1, \quad K_1 \vee K_2 = K_2 \vee K_1.$$

2. *Element zero: per tota $H \in \digamma(F, F)$, K_0 és el neutre de la unió,*

$$K_1 \wedge K_0 = K_0, \quad K_1 \vee K_0 = K_1.$$

3. *Igualtat: si H assoleix la fita superior de Fréchet, H^+, i $K_1 = K_2$, i.e., g.p.t. $(x, y) \in \mathcal{X} \times \mathcal{X}$, $K_1(x, y) = K_2(x, y)$, aleshores*

$$K_1 \wedge K_2 = K_1 = K_2, \quad K_1 \vee K_2 = K_1 = K_2.$$

4. *Ortogonalitat: si X, Y son estocàsticament independents ($H(x,y) = F(x)F(y)$ g.p.t. $(x,y) \in \mathcal{X} \times \mathcal{X}$), aleshores*

$$K_1 \wedge K_2 = K_0, \quad K_1 \vee K_2 = K_1 + K_2.$$

La simetria dels nuclis intersecció i unió es deriva directament de la simetria dels nuclis implicats. Fent servir raonaments anàlegs als utilitzats en el cas de matrius, s'obté la no negativitat dels valors propis de la unió. La commutativitat se deriva directament de la commutativitat de la suma de funcions mesurables a l'espai producte i 2) també és directe. Es comprova que l'arrel quadrada només està ben definida si la covariança entre les funcions pròpies dels nuclis és δ_{ij}, la delta de Kronecker. En particular, si l'$H-$producte és respecte d'H^+, $K^{\frac{1}{2}} \star K^{\frac{1}{2}} = K$. Aleshores 3) és evident. El càlcul de la covariança (del producte) de les funcions $K_1(X, \cdot)$, $K_2(\cdot, Y)$ dóna 0 si X i Y són independents.

Una relació d'equivalència

Suposem que els nuclis simètrics s.d.p. K_1, K_2 comparteixen un conjunt numerable (o finit) de funcions ortonormals $\{\xi_i\}$ per alguns conjunts $\{\lambda_i\}, \{\mu_i\}$ de valors propis, amb dimensions $N_1 \leq N_2 \leq \infty$, respectivament, en el sentit que

$$K_1 = \sum_{i=0}^{N_1} \lambda_i \xi_i \otimes \xi_i, \quad K_2 = \sum_{i=0}^{N_2} \mu_i \xi_i \otimes \xi_i,$$

i assumint que $\lambda_0 = \mu_0 = 0$ podria no ser un valor propi d'aquests nuclis. Aleshores direm que K_1 i K_2 són equivalents, i notarem $K_1 \sim K_2$.

Òbviament, si $N_1 < N_2$ podem escriure la primera expansió com

$$K_1 = \sum_{i=0}^{N_2} \lambda_i f_i \otimes f_i,$$

on $\lambda_{N_1+1} = \lambda_{N_1+2} = ... = \lambda_{N_2} = 0$.

Denotem la classe d'equivalència dels nuclis de $\mathcal{X} \times \mathcal{X}$ que comparteixen les funcions pròpies $\{\xi_i\}$, $i = 0, ..., N$, per

$$[\xi_N] = \left\{ K \in \mathcal{K}_{\mathcal{X}\times\mathcal{X}} \mid K = \sum_{i=0}^{N} \lambda_i \xi_i \otimes \xi_i \geq 0, \ \sum_{i=0}^{N} \lambda_i^2 < \infty \right\}.$$

Així, $K_1 \sim K_2$ sii $K_1, K_2 \in [\xi_N]$. Siguin $K_1, K_2 \in [\xi_N]$ amb valors propis $\{\lambda_i\}, \{\mu_i\}$, respectivament, amb $\sum_{i=0}^{\infty} \lambda_i < \infty$, $\sum_{i=0}^{\infty} \mu_i < \infty$. Aleshores

$$K_1 \wedge K_2 \in [\xi_N] \quad \text{amb valors propis } \left\{ \lambda_i^{\frac{1}{2}} \mu_i^{\frac{1}{2}} \right\},$$

i

$$K_1 \vee K_2 \in [\xi_N] \quad \text{amb valors propis } \left\{ \lambda_i + \mu_i - \lambda_i^{\frac{1}{2}} \mu_i^{\frac{1}{2}} \right\}.$$

Es comprova fàcilment que l'$H-$producte és commutatiu si els nuclis comparteixen les funcions pròpies. El càlcul dels nuclis intersecció i unió prova directament el resultat, utilitzant l'ortonormalitat de les funcions pròpies.

La dimensió dels nuclis

Es tenen les següents propietats (vegi's [11] per la prova):

Teorema 6 *La dimensió dels nuclis satisfà les següents propietats:*

1. *Un nucli continu, simètric no nul posseix valors i funcions pròpies amb cardinalitat $\varkappa_0$ sii el nucli no es pot escriure com a suma finita del producte de funcions de cada variable per separat.*

2. *Si el nucli K només té un nombre finit de valors propis $\lambda_1, \lambda_2, ..., \lambda_n$ ha de tenir dimensió finita i s'ha de poder representar com*

$$K(s,t) = \sum_{i=1}^{n} \lambda_i \xi_i(s) \xi_i(t).$$

   Un nucli de dimensió finita només té un nombre finit de valors propis.

3. *Tots els valors propis d'un nucli real simètric són reals. El nombre de valors propis no nuls d'un nucli és la seva dimensió. Si un nucli és definit positiu (tots els valors propis són estrictament positius) la seva dimensió és infinita. Si K és semidefinit positiu, la seva dimensió és finita.*

Una relació d'ordre parcial

Considerem $K_1 \sim K_2$, tals que $\dim(K_1) = N_1 \leq \dim(K_2) = N_2$. Siguin $\{\lambda_i\}, \{\mu_i\}$, els conjunts de valors propis respectius. Diem que K_1 és més petit que K_2, i denotem aquest ordre parcial per $K_1 \lesssim K_2$, sii $\lambda_i \leq \mu_i, \ i = 0, 1, ..., N_1$.

Proposició 7 *Siguin K_1, K_2 dos nuclis de $[\xi_N]$, amb $K_1 \lesssim K_2$. Aleshores*

$$K_1 \lesssim K_1 \wedge K_2 \lesssim K_1 \vee K_2 \lesssim K_2.$$

La prova es deriva de la relació entre els valors propis d'aquests nuclis. $K_1 \lesssim K_2$ equival a $\lambda_i \leq \mu_i$, $\forall i$. Com a conseqüència, s'obté que

$$\lambda_i \leq \lambda_i^{\frac{1}{2}} \mu_i^{\frac{1}{2}} \leq \lambda_i + \mu_i - \lambda_i^{\frac{1}{2}} \mu_i^{\frac{1}{2}} \leq \mu_i.$$

La traça d'un operador lineal $\mathcal{K}$ de nucli K, mesurable respecta a una mesura μ es defineix com

$$tr\left(K_\mu\right) = \int_{\mathcal{X}} K\left(s,s\right) d\mu\left(s\right).$$

Denotem, simplement, $tr\left(K_\mu\right) \equiv tr\left(K\right)$. La traça d'un nucli K respecte d'F és la suma dels valors propis del nucli respecte de la mesura F, $tr\left(K\right) = \sum_{i=0}^{\infty} \lambda_i$, on K pot expandir-se com $K = \sum_{i=0}^{\infty} \lambda_i \xi_i \otimes \xi_i$, per un conjunt apropiat de funcions pròpies $\left\{\xi_i\right\}$, ortonormals respecte d'F.

Proposició 8 *Siguin* $\left\{\lambda_i\right\}, \left\{\mu_i\right\}$ *conjunts de valors propis de nuclis* $K_1 \sim K_2$*. Les traces són:*

$$\begin{aligned} tr\left(K_1 \wedge K_2\right) &= \sum_{i=0}^{\infty} \lambda_i^{\frac{1}{2}} \mu_i^{\frac{1}{2}}, \\ tr\left(K_1 \vee K_2\right) &= tr\left(K_1\right) + tr\left(K_2\right) - tr\left(K_1 \wedge K_2\right). \end{aligned}$$

Es comprova fàcilment, per la positivitat dels valors propis, que si $K_1 \lesssim K_2$, aleshores $tr\left(K_1\right) \leq tr\left(K_2\right)$ i, per tant

$$tr\left(K_1\right) \leq tr\left(K_1 \wedge K_2\right) \leq tr\left(K_1 \vee K_2\right) \leq tr\left(K_2\right).$$

Els resultats que hem obtingut per nuclis simètrics són els anàlegs als obtinguts per matrius simètriques. Per tant, aquést és el marc teòric per extendre al continu el Related Metric Scaling.

Hem vist que el producte entre nuclis que ens permet definir les operacions intersecció i unió entre nuclis és la covariança entre una funció de la primera variable i una altra funció de la segona variable. Per tant, en el següent capítol hem volgut aprofondir en l'estudi de la covariança entre funcions.

Capítol 4. Covariança entre funcions i H-afinitat

Cuadras (2002, [18]) generalitzà el lema de Hoeffding (1940), que estableix cóm es pot expressar la covariança en termes de funcions de distribució, a la covariança entre funcions de variació afitada. En aquest capítol ampliem les propietats d'aquest producte quasi-escalar i definim una afinitat entre funcions, depenent d'H, la funció de distribució conjunta.

El resultat de partida, provat a [18], és el següent ($VA\left([a,b]\right)$ representa el conjunt de funcions de variació afitada en $[a,b]$).

Teorema 9 *Si $\alpha(x)$ i $\beta(y)$ són dues funcions definides sobre $[a,b]$, $[c,d]$, respectivament, tals que:*

1. *Les dues són de variació afitada, $\alpha \in VA\left([a,b]\right), \beta \in VA\left([c,d]\right),$*
2. *$E\left(|\alpha(X)\beta(Y)|\right), E\left(|\alpha(X)|\right), E\left(|\beta(Y)|\right) < \infty,$*

aleshores

$$Cov\left(\alpha(X), \beta(Y)\right) = \int_a^b \int_c^d \left(H(x,y) - F(x)G(y)\right) d\alpha(x) d\beta(y).$$

Al conjunt de funcions de variació afitada sobre un interval $[a,b] \subseteq \mathbb{R}$, definim el producte escalar

$$\langle \alpha, \beta \rangle_H = Cov\left(\alpha(X), \beta(Y)\right).$$

$H(x,y)$ és una distribució conjunta simètrica i amb dependència quadrant positiva (PQD). Les propietats són les d'un producte quasi-escalar.

Proposició 10 *Siguin $\alpha, \beta \in BV\left([a,b]\right), [a,b] \subseteq \mathbb{R}$. Notarem $\langle \cdot, \cdot \rangle_H \equiv \langle \cdot, \cdot \rangle$. Es tenen les següents propietats:*

1. *Per tota $\alpha \in BV\left([a,b]\right)$, $\langle \cdot, \alpha \rangle$ i $\langle \alpha, \cdot \rangle$ són lineals:*

   $$\langle r\alpha_1 + s\alpha_2, l\beta_1 + m\beta_2 \rangle = rl\langle \alpha_1, \beta_1 \rangle + rm\langle \alpha_1, \beta_2 \rangle + sl\langle \alpha_2, \beta_1 \rangle + sm\langle \alpha_2, \beta_2 \rangle,$$

   on $r, s, l, m \in \mathbb{R}$. Si $\mathbf{0}$ és la funció constant igual a 0, aleshores $\langle \alpha, \mathbf{0} \rangle = \langle \mathbf{0}, \alpha \rangle = 0$.
2. *$\langle \alpha, \beta \rangle = \langle \beta, \alpha \rangle$ (simetria).*
3. *$\langle \alpha, \alpha \rangle \geq 0$ (no negativitat).*

4. $\langle \alpha, \beta \rangle_H$ *satisfà*

$$\langle \alpha, \beta \rangle_H^2 \leq \langle \alpha, \alpha \rangle_H \langle \beta, \beta \rangle_H \tag{iii}$$

amb igualtat sii

$$\langle \alpha + t\beta, \alpha + t\beta \rangle_H = 0$$

per alguna constant t.

Aquestes propietats es comproven aplicant la definició del producte i les propietats de la integral de Riemann-Stieltjes.

Una relació d'equivalència

Sigui $\alpha \in BV([a,b])$. Diem que $\gamma \in BV([a,b])$ és equivalent a α i escrivim $\gamma \sim \alpha$ si existeix un nombre real g tal que $\gamma(x) = \alpha(x) + g$ (g.p.t. $x \in [a,b]$). La classe d'equivalència que conté a α es nota $[\alpha]$. La covariança entre membres de dues classes d'equivalència $[\alpha]$ i $[\beta]$ és la mateixa i qualsevol resultat s'ha de considerar vàlid gairebé per tot membre de la classe. A més, els càlculs se simplifiquen centrant les variables aleatòries, i.e., considerat $\alpha_0 \in [\alpha]$, amb $E(\alpha_0(X)) = 0$ i, anàlogament β_0.

Aquesta relació d'equivalència és un cas particular de la següent: sigui $\alpha \in BV([a,b])$ una funció de variació afitada. Diem que ξ és equivalent a α ($\xi \dot{\sim} \alpha$) si existeixen nombres reals h, g tals que $\xi(x) = h\alpha(x) + g$, g.p.t. x. Notarem la classe d'equivalència que conté α per $[\alpha]^*$.

Una afinitat entre funcions depenent d'H

Cuadras [18] exten el coeficient de correlació entre funcions per

$$Cor(\alpha(X), \beta(Y)) = \frac{Cov(\alpha(X), \beta(Y))}{(Var(\alpha(X))Var(\beta(Y)))^{\frac{1}{2}}} \tag{iv}$$

on

$$Var(\alpha(X) = \int_a^b \int_a^b (\min\{F(x), F(y)\} - F(x)F(y))\, d\alpha(x)\, d\alpha(y).$$

Aquesta correlació pot existir fins i tot si la variança és 0. Observi's que (iv) és el coeficient de correlació de Pearson si α i β són la funció identitat.

Gràcies a la desigualtat (iii) a la Proposició 10 podem definir un altre coeficient $A_H(\alpha, \beta)$, tal que $A_H^2(\alpha, \beta)$ prengui valors entre 0 i 1. Aquest coeficient hauria de ser una mesura de l'afinitat entre α i β respecte d'H.

Definició 11 *Sigui $H(x,y)$ una funció de distribució conjunta, simètrica en x,y, i PQD. Siguin $\alpha,\beta \in BV([a,b])$. Aleshores $A_H(\alpha,\beta)$, definida per*

$$A_H^2(\alpha,\beta) = \frac{\langle\alpha,\beta\rangle_H^2}{\langle\alpha,\alpha\rangle_H \langle\beta,\beta\rangle_H}$$

és l'$H-$afinitat entre les funcions α i β o afinitat respecte d'H.

La següent proposició resumeix les propietats de $A_H^2(\alpha,\beta)$.

Proposició 12 *La mesura d'afinitat respecte d'H, A_H satisfà:*

1. $0 \leq A_H^2 \leq 1$.
2. *Si les variables aleatòries X i Y són estocàsticament independents, aleshores $A_H^2 = 0$.*
3. *Si α,β són ortogonals respecte de $\langle\cdot,\cdot\rangle_H$, aleshores $A_H^2 = 0$.*
4. *Si $\alpha = h\beta + g$, per algunes constants $h,g \in \mathbb{R}$, $h \neq 0$, i α,β no són ortogonals respecte a $\langle\cdot,\cdot\rangle_H$, aleshores $A_H^2 = 1$ si X i Y no són estocàsticament independents; altrament, $A_H^2 = 0$.*
5. *L'$H-$afinitat entre funcions de variables aleatòries és invariant sota el grup de transformacions lineals.*

La propietat 1 es deriva de la desigualtat (*iii*) i la resta s'obtenen de les propietats anàlogues de la covariança entre variables aleatòries.

També es comprova fàcilment que $A_H^2(\alpha,\alpha) = 1$ si $H \neq F \otimes F$, que és un corol·lari directe de la propietat 4.

Es tenen les fites següents per $\langle\alpha,\alpha\rangle_H$, quan H és simètrica i PQD:

$$0 = \langle\alpha,\alpha\rangle_{F\otimes F} \leq \langle\alpha,\alpha\rangle_H \leq \langle\alpha,\alpha\rangle_{H^+}.$$

Si $H = F \otimes F$ hi ha independència. Considerem la funció positiva, contínua, i de variació afitada K

$$K \equiv H^+ - F \otimes F - (H - F \otimes F) \geq 0$$

Se segueix que $K \in RS\left(\alpha \times \alpha\right)$. Definim

$$\left\langle \alpha, \alpha \right\rangle_K = \int_a^b \int_a^b K(x,y) d\alpha(x) d\alpha(y),$$

que és no negatiu, com hem dit. Només cal aplicar la linealitat de la integral de Riemann-Stieltjes i s'obté

$$0 \leq \left\langle \alpha, \alpha \right\rangle_K = \left\langle \alpha, \alpha \right\rangle_{H^+} - \left\langle \alpha, \alpha \right\rangle_H .$$

Per últim, sigui $Cor\left(\alpha\left(X\right), \beta\left(Y\right)\right)$ el coeficient de correlació entre dues funcions $\alpha\left(X\right), \beta\left(Y\right)$, i sigui $A_H^2(\alpha,\beta)$ la mesura d'$H-$afinitat. Aleshores $A_{H^+}(\alpha,\beta) = Cor\left(\alpha\left(X\right), \beta\left(Y\right)\right)$ i $A^2_{H^+}(\alpha,\beta) \leq A_H^2\left(\alpha,\beta\right)$.

Capítol 5. H-afinitat i dimensió

Definim la dimensió d'H en termes del producte escalar i l'afinitat.

Definició 13 *Sigui* $\Phi_H = (\varphi_\imath, \imath \in \Im)$ *un conjunt de funcions reals de* $L^2([a,b],F)$, *on* F *és la mesura de probabilitat induïda per la funció de distribució marginal* F. *La dimensió d'*H *és* $\#(\Phi_H)$ *si aquest conjunt de funcions satisfà:*

1. $A^2_H(\varphi_\imath,\varphi_\jmath) = 0$, $\imath \neq \jmath$, *i* $A^2_H(\varphi_\imath,\varphi_\imath) = 1$, $\varphi_\imath,\varphi_\jmath \in \Phi_H$.
2. $A^2_H\left(\alpha,\beta\right) = 0$ *si* $\alpha,\beta \in \Phi^\perp_H$ *on l'ortogonalitat és respecte de* $\left\langle \cdot,\cdot \right\rangle_H$.

Clarament, si la dimensió és finita o numerable, aleshores $\#(\Phi_H) \leq \#(\Phi_{H^+})$, donat que $\left\langle \alpha, \alpha \right\rangle_H \leq \left\langle \alpha, \alpha \right\rangle_{H^+}$ éssent H^+ la fita superior de Fréchet. Llavors podríem tenir $\left\langle \alpha, \alpha \right\rangle_H = 0$ però $\left\langle \alpha, \alpha \right\rangle_{H^+} \neq 0$ per alguna $\alpha \neq 0$.

Exemples de dimensions són:

1. Dimensió 0 si hi ha independència estocàstica.
2. Dimensió finita, $n > 0$ si H és la familia FGM generalitzada.
3. Dimensió numerable $\varkappa_0$ si H és la fita superior de Fréchet, la familia de Regressió, o bé, la família Ali-Mikhail-Haq.
4. Dimensió contínua $\varkappa_1$ si H és la família Cuadras-Augé.

Conjunts ortonormals complets de funcions que apareixen en algunes expansions de distribucions bivariants podrien satisfer les condicions per la Definició 13. Per exemple, suposem que H és una funció de distribució bivariant general, amb marginals F, G (possiblement $F \neq G$), la mesura $dH(x,y)$ és absolutament contínua respecte a $dF(x)dG(y)$ i que el coeficient de contingència de Pearson ϕ^2, definit per

$$\phi^2 + 1 = \int_a^b \int_c^d (dH(x,y))^2 / (dF(x)dG(y))$$

és finit. Aleshores es té la següent expansió

$$dH(x,y) - dF(x)dG(y) = \sum_{n \geq 1} \rho_n a_n(x) b_n(y) dF(x) dG(y), \qquad (v)$$

on ρ_n són les correlacions canòniques, ordenades en ordre decreixent, i $a_n(x), b_n(y)$ són les funcions canòniques. Si l'expansió diagonal (v) existeix, la dimensió d'H ve determinada pel nombre de variables canòniques amb correlacions canòniques positives. En general, quan una distribució H pot expandir-se mitjançant un sistema ortogonal complet de funcions, el nombre d'aquestes funcions amb covariança positiva donarà la dimensió d'H.

Finalitzem aquesta tesi aplicant el mètode de l'expansió diagonal de Lancaster per construïr distribucions bivariants, utilitzant les components principals de les variables, estandarditzades, que són, per construcció, variables canòniques.

Capítols 6 i 7. Construcció de distribucions diagonals via components principals

S'ha generalitzat el mètode de les expansions diagonals de distribucions bivariants (Lancaster, 1958, [71]) utilitzant les dimensions principals de cada variable marginal. Aquesta generalització la proposà Cuadras al 2002. En un primer capítol donem la base teòrica del mètode de les expansions diagonals (que es pot veure a [71], [54]), i donem la base teòrica de les components principals de les variables aleatòries. Es comprova que les components principals estandarditzades de les variables X, Y, diguem-ne F_j, G_j , farien el paper de funcions canòniques a_j, b_j a l'expansió (v). Així, en termes de densitats conjunta $h(x,y)$ i marginals $f(x), g(y)$ aquesta expansió equival a

$$h(x,y) = f(x)\, g(y) \left[1 + \sum_{j=1}^{\infty} \rho_j F_j(x)\, G_j(y) \right], \qquad (vi)$$

que és la representació d'una distribució bivariant amb densitats marginals f, g ([16], [17]). El vector (de dimensió, pot ser, infinita) de coeficients de correlació $(\rho_1, \rho_2, ...)$

es pot escollir per a construïr diverses distribucions amb aquesta representació. Per exemple, prenent $\rho_N > \rho_{N+1} = 0$, tenim que (vi) representa una família diagonal anidada $N-$paramètrica de distribucions. Per tant, aquesta representació proporciona un métode per la construcció de distribucions bivariants amb marginals donades, via components principals.

Per construcció, es tenen les següents propietats:

1. $\{F_j(X)\}, \{G_j(Y)\}$ són successions de variables aleatòries centrades i incorrelacionades entre sí.
2. $Cor(F_i(x), G_j(y)) = \rho_j \delta_{ij}$.
3. $\{\rho_j\}_j$ és la successió de correlacions canòniques.
4. $\{F_j(X), G_j(Y)\}$ és la successió de variables canòniques.

L'expansió (vi) ens permet d'obtenir la densitat $c(u,v)$ de la còpula $C(u,v)$ corresponent, de forma senzilla. Com

$$h(x,y) = f(x)\, g(y)\, c(F(x), G(y)),$$

n'hi ha prou amb aplicar el canvi de variables $F(x) = u$, $G(y) = v$, així,

$$c(u,v) = 1 + \sum_{j=1}^{\infty} \rho_j F_j\left(F^{-1}(u)\right) G_j\left(G^{-1}(v)\right).$$

Cuadras i Fortiana ([26]) provaren que aquesta expansió pot veure's com un cas particular de l'escalament ponderat continu (continuous weighted scaling). Cuadras ([17]) expressà aquesta expansió diagonal en termes de funcions de distribució, i va introduïr una altra extensió d'aquest métode que consisteix a utilitzar les dimensions principals de cada variable marginal com a conjunts de funcions ortogonals ([16]). Les components principals (o direccions) s'han obtingut per X, uniforme a l'interval $[0,1]$, X exponencial, X logística i, finalment, X amb distribució de Pareto (α, θ), $\alpha > 2$, $\theta \geq 0$. Vegi's Cuadras i Fortiana, 1995 [25], Cuadras i Lahlou, 2000 [30], i Cuadras i Lahlou, 2002 [31].

De cadascuna de les variables de les quals es coneixen les components principals donem les components principals estandarditzades i, a partir d'aqui, obtenim la forma general de la densitat bivariant en funció d'aquestes, la corresponent distribució, la còpula, i obtenim condicions necessaries i suficients per tal que la representació diagonal correspongui veritablement a una densitat en el cas d'una sola correlacio canònica estrictament positiva, i condicions només necessàries en el cas de dues correlacions. Obtenim també la correlació de cada família. En alguns casos, hem pogut identificar les densitats obtingudes com a pertanyents a una família coneguda.

Si φ_j són les funcions pròpies del nucli de la covariança associat al vector aleatori (X,Y),

$$K(s,t) = \min(F(s),F(t)) - F(s)F(t), \quad s,t \in I = [a,b],$$

i $f_j(x) = \int_a^x \varphi_j(t)dt$, aleshores les components principals $f_j(X)$, llurs esperances μ_j i variances λ_j, ens permeten obtenir les components principals estandarditzades $F_j(X) = (f_j(X) - \mu_j)/\lambda_j$. Així, substituïnt aquestes funcions a l'expansió diagonal (vi) s'obtenen les famílies diagonals.

Aplicarem el següent resultat general.

Teorema 14 *Siguin X,Y dues variables aleatòries, idènticament distribuïdes (i.d.) i absolutament contínues amb funció de distribució marginal F. Amb el canvi de variables $F(x) = u$, $F(y) = v$, escrivim la familia diagonal com la densitat de la còpula corresponent*

$$c(u,v) = 1 + \sum_{j=1}^{N} \rho_j F_j^*(u) F_j^*(v),$$

on $F_j^*(\cdot)$ és $F_j \circ F^{-1}(\cdot)$. Aleshores:

1. *$c(u,v)$, així com la corresponent $h(x,y)$, és una densitat si i només si*

$$\sup_{(u,v)\in[0,1]^2} \left[-\sum_{j=1}^{N} \rho_j F_j^*(u) F_j^*(v) \right] \leq 1.$$

2. *Les correlacions canòniques han de satisfer la condició necessària*

$$-\sum_{j=1}^{N} \rho_j F_j^*(u) F_j^*(1) \leq 1,$$

uniformement en $u \in [0,1]$.

3. *El coeficient de correlació de Pearson és*

$$Cor(X,Y) = \left(\sum_{j=1}^{N} \rho_j I_j^2 \right) / Var(X),$$

on

$$I_j = \int_0^1 F^{-1}(u) F_j^*(u) du.$$

La prova de 1) es basa en imposar la positivitat a la funció $c(u,v)$. Prenent $v=1$, 2) és conseqüència de 1). Finalment, com X i Y són i.d., les variances són iguals, i el càlcul de la covariança de (X,Y) amb el canvi de variables proposat permet obtenir 3).

A continuació, aplicarem aquest resultat general a cadascuna de les components principals que es coneixen, per tal d'obtenir les corresponents famílies diagonals.

Família diagonal amb marginals uniformes

Siguin U,V variables aleatòries amb distribució uniforme a $[0,1]$, i.e., amb densitat $f_U(u) = \mathbf{1}_{[0,1]}(u)$. Sigui $\{\rho_j\}$ una succesió decreixent de reals no negatius tals que $1 \geq \rho_1 > \rho_2 > ... > 0$. Aleshores

1. La funció

$$c(u,v) = 1 + \sum_{j=1}^{N} \rho_j \; 2\cos(j\pi u)\cos(j\pi v), \qquad u,v \in [0,1]$$

   és la densitat d'una família diagonal amb marginals uniformes a $[0,1]$ sii

$$\sup_{(u,v)\in[0,1]^2} \left[-2\sum_{j=1}^{N} \rho_j \cos(j\pi u)\cos(j\pi v) \right] \leq 1.$$

   Cada ρ_j és el coeficient de correlació entre les $j-$èssimes components principals estandarditzades $F_j(U) = -\sqrt{2}\cos(j\pi U)$, i $F_j(V)$,

$$\rho_j = Cor(\cos(j\pi U), \cos(j\pi V)).$$

2. Aquests coeficients han de satisfer la condició

$$-2\sum_{j=1}^{N} \rho_j \cos(j\pi u)(-1)^j \leq 1,$$

   uniformement en $u \in [0,1]$.

3. El coeficient de correlació de (U,V) ve donat per

$$Cor(U,V) = \frac{96}{\pi^4}\sum_{k=1}^{N} \frac{\rho_{2k-1}}{(2k-1)^4}.$$

Família diagonal amb marginals exponencials

Siguin X, Y exponencials amb mitjana 1, i.e., amb la mateixa densitat donada per $f(x) = \exp(-x)\,\mathbf{1}_{[0,\infty]}(x)$. Siguin ξ_j les $j-$èssimes arrels positives de J_1, éssent J_0, J_1 les funcions de Bessel de primera classe, amb $0 \leq \rho_j \leq 1$. Aleshores:

1. La funció

$$h(x,y) = f(x)\, f(y) \left[1 + \sum_{j=1}^{N} \frac{\rho_j}{J_0(\xi_j)^2}\; J_0\left(\xi_j \exp\left(\frac{-x}{2}\right)\right) J_0\left(\xi_j \exp\left(\frac{-y}{2}\right)\right)\right],$$

   $x, y \geq 0$, és la densitat d'una família diagonal $h \in (f, f)$ sii

$$\sup_{(u,v)\in[0,1]^2} \left[-\sum_{j=1}^{N} \frac{\rho_j}{J_0(\xi_j)^2} J_0\left(\xi_j\sqrt{1-u}\right) J_0\left(\xi_j\sqrt{1-v}\right)\right] \leq 1.$$

   Cada ρ_j és la correlació entre les $j-$èssimes components principals estandarditzades $F_j(X) = J_0(\xi_j \exp(-X/2))\,/J_0(\xi_j)$, i $F_j(Y)$,

$$\rho_j = Cor\left(J_0(\xi_j \exp(-X/2)), J_0(\xi_j \exp(-Y/2))\right).$$

2. Aquests coeficients han de satisfer la condició

$$-\sum_{j=1}^{N} \rho_j \frac{J_0\left(\xi_j\sqrt{1-u}\right)}{J_0(\xi_j)^2} \leq 1,$$

   uniformement en $0 \leq u \leq 1$.

3. El coeficient de correlació de (X, Y) ve donat per

$$Cor(X,Y) = \sum_{j=1}^{N} \rho_j \left(\frac{4(1 - J_0(\xi_j))}{\xi_j^2 J_0(\xi_j)}\right)^2$$

Família diagonal amb marginals logístiques

Siguin X, Y variables aleatòries amb distribució logística estàndard, i.e., amb densitat donada per

$$f(x) = \frac{\exp(-x)}{(1 + \exp(-x))^2}, \quad x \in \mathbb{R}.$$

Sigui $L_j(t)$, $t \in [0,1]$ el $j-$èssim polinomi de Legendre traslladat al $[0,1]$, i $F(x) = 1/(1 + \exp(-x))$ la funció de distribució. Sigui $\{\rho_j\}$ una succesió decreixent de reals no negatius tals que $1 \geq \rho_1 > \rho_2 > ... > 0$. Aleshores

1. La funció

$$h(x,y) = f(x) f(y) \left[1 + \sum_{j=1}^{N} \rho_j \ L_j(F(x)) L_j(F(y)) \right], \qquad x,y \in \mathbb{R}$$

és la densitat d'una família diagonal amb marginals logístiques sii

$$\sup_{(u,v)\in[0,1]^2} \left[-\sum_{j=1}^{N} \rho_j L_j(u) L_j(v) \right] \leq 1.$$

Cada ρ_j és la correlació entre les $j-$èssimes components principals estandarditzades $F_j(X) = L_j(F(X))$, i $F_j(Y)$,

$$\rho_j = Cor\left(L_j(F(X)), L_j(F(Y))\right).$$

2. Aquests coeficients han de satisfer la condició

$$-\sum_{j=1}^{N} \rho_j \sqrt{2j+1} L_j(u) \leq 1,$$

uniformement en $u \in [0,1]$.

3. El coeficient de correlació de (X,Y) ve donat per

$$Cor(X,Y) = \frac{3}{\pi^2} \sum_{k=1}^{N} \frac{4k-1}{k^2(2k-1)^2} \rho_{2k-1}.$$

Família diagonal amb marginals Pareto

Siguin X, Y dues variables aleatòries amb distribució de Pareto amb densitat donada per $f(x) = 3x^{-4}\ \mathbf{1}_{(1,\infty)}(x)$. Així X (i Y) segueix la distribució de Pareto(α,θ) amb paràmetres $\alpha = 3, \theta = 1$. Sigui η_j la $j-$èssima arrel positiva de l'equació $x = \tan(x)$. Aleshores

1. L'expansió

$$h(x,y) = f(x) f(y) \left[1 + \sum_{j=1}^{N} \rho_j \ \frac{2}{3\sin^2(\eta_j)} \ x\sin(\eta_j/x) \ \ y\sin(\eta_j/y) \right],$$

$x, y > 1$, amb $0 \leq \rho_j \leq 1$, per $j \geq 1$, és una densitat bivariant amb marginals Pareto$(3,1)$ sii

$$\sup_{(u,v)\in[0,1]^2} \left[\frac{-2}{3} \sum_{j=1}^{N} \frac{\rho_j \sin\left(\eta_j \sqrt[3]{1-u}\right) \sin\left(\eta_j \sqrt[3]{1-v}\right)}{\sin^2(\eta_j) \sqrt[3]{1-u}\sqrt[3]{1-v}} \right] \leq 1.$$

Cada ρ_j és el coeficient de correlació entre les $j-$èssimes components principals estandarditzades $F_j(X) = \sqrt{\frac{2}{3}} X \sin(\eta_j/X) / \sin(\eta_j)$, i $F_j(Y)$,

$$\rho_j = Cor\left(X \sin(\eta_j/X), Y \sin(\eta_j/Y)\right).$$

2. Aquests coeficients han de satisfer la condició

$$\frac{-2}{3} \sum_{j=1}^{N} \frac{\rho_j \eta_j \sin\left(\eta_j \sqrt[3]{1-u}\right)}{\sin^2(\eta_j) \sqrt[3]{1-u}} \leq 1,$$

uniformement en $u \in [0,1]$.

3. El coeficient de correlació ve donat per

$$Cor(X,Y) = 8 \sum_{j=1}^{N} \rho_j \left(\frac{1-\cos(\eta_j)}{\eta_j \sin(\eta_j)}\right)^2.$$

La següent taula resumeix les propietats principals de les densitats bivariants $h(x,y)$ que hem construït. Per cada variable marginal donem la correlació de la família, la condició necessària i suficient per tal que h sigui una densitat si $N=1$, i una condició necessària si $N=2$.

Marginal	$Cor(X,Y)$	$N=1$	$N=2$
Uniforme $(0,1)$	$\frac{96}{\pi^4} \sum_{k=1}^{N} \frac{\rho_{2k-1}}{(2k-1)^4}$	$\rho_1 \leq \frac{1}{2}$	$\rho_1^2 + \frac{\left(\rho_2 - \frac{1}{4}\right)^2}{1/8} \leq \frac{1}{2}$
Exponencial (1)	$\sum_{j=1}^{N} \rho_j \left(\frac{4(1-J_0(\xi_j))}{\xi_j^2 J_0(\xi_j)}\right)^2$	$\rho_1 \leq 0.40276$	$\rho_1 \leq 0.61198$
Logística (0)	$\frac{3}{\pi^2} \sum_{k=1}^{N} \frac{\rho_{2k-1}(4k-1)}{k^2(2k-1)^2}$	$\rho_1 \leq \frac{1}{3}$	$\frac{\rho_1^2}{1/3} + \frac{\left(\rho_2 - \frac{1}{5}\right)^2}{1/25} \leq 1$
Pareto $(3,1)$	$8 \sum_{j=1}^{N} \rho_j \left(\frac{1-\cos(\eta_j)}{\eta_j \sin(\eta_j)}\right)^2$	$\rho_1 \leq 0.3258$	$\rho_1 \leq 0.52706$

Taula 1. Distribucions diagonals amb marginals específiques

Conclusions

La definició de les operacions intersecció i unió entre matrius simètriques proporciona un marc teòric que pot ser d'utilitat en el *Related Metric Scaling*. Aquestes operacions tenen algunes propietats similars a la intersecció i la unió de subespais vectorials (per exemple, els generats per les columnes d'una matriu). Els aspectes teòrics desenvolupats en aquest treball podrien facilitar la interpretació geomètrica del RMS i podrien obrir vies de recerca en un futur a tècniques multivariants semblants. Les operacions aconsegueixen extraure l'estructura de dependència subjacent a la informació que es presenta en un context multivariant.

Hem definit les operacions unió i intersecció entre nuclis a partir de la covariança de funcions de les variables marginals, obtenint els fonaments teòrics que poden facilitar l'extensió al continu del Related Metric Scaling i qüestions relacionades.

S'ha definit la dimensió d'una distribució conjunta $H \in \mathcal{F}(F, G)$ mitjançant una mesura d'afinitat que depén d'H. El cardinal del conjunt de valors propis del nucli de la covariança $H(x, y) - F(x) G(y)$ podria determinar aquesta dimensió. Si H pot representar-se mitjançant una expansió diagonal, el nombre de correlacions canòniques no nul.les dóna la dimensió d'H. Aquésta s'ha obtingut per diverses famílies paramètriques. La dimensió d'una distribució conjunta es pot veure com una certa "mesura" de dependència entre dues variables aleatòries: és nul.la quan hi ha independència estocàstica, i és infinita numerable si H és la fita superior de Fréchet. Si H es pot representar mitjançant una expansió diagonal, hem trobat que la màxima correlació augmenta quan la dimensió (el nombre de correlacions no nul.les) és més gran.

Hem construït algunes famílies diagonals via components principals i hem donat condicions necessàries i suficients per a les densitats en el cas de dimensió 1. La nostra recerca actual se centra en el cas de dues correlacions canòniques. Aquest métode suggereix que les components principals estandarditzades podrien millorar alguns tests d'independència i, també, de bondat d'ajustament.

ERRATA
(corrections are in bold)

page 80, line 23: (Proposition 4.3.2) ... *Let* $\alpha \in BV([a,b])$ *and* $\beta \in BV([\mathbf{a},\mathbf{b}])$.

page 83, lines 8-9: (Proposition 4.3.10) ... *Then* $A_{H^+}(\alpha,\beta) = Cor(\alpha(X), \beta(\mathbf{X}))$ *and*

$$\boldsymbol{\rho}^2(\alpha,\beta) \leq A_H^2(\alpha,\beta).$$

page 100, line 13: ... and H can be expanded as

$$dH(x,y) = dF(x)\,dG(y)\sum_{i=0}^{\infty}\rho_i a_i(x)\,b_{\mathbf{i}}(y), \qquad (6.1)$$