

Chapter 1

Copulas, Sklar's Theorem, and Distributional Transform

In this chapter we introduce some useful tools in order to construct and analyse multivariate distributions. The distributional transform and its inverse the quantile transform allow to deal with general one-dimensional distributions in a similar way as with continuous distributions. A nice and simple application is a short proof of the general Sklar Theorem. We also introduce multivariate extensions, the multivariate distributional transform, and its inverse, the multivariate quantile transform. These extensions are a useful tool for the construction of a random vector with given distribution function respectively allow to build functional models of classes of processes. They are also a basic tool for the simulation of multivariate distributions. We also describe some applications to stochastic ordering, to goodness of fit tests and to a general version of the empirical copula process. We introduce to some common classes of copula models and explain the pair copula construction method as well as a construction method based on projections. In the final part extensions to generalized Fréchet classes with given overlapping multivariate marginals are discussed. The construction of dependence models by projections is extended to the generalized Fréchet class where some higher dimensional marginals are specified.

1.1 Sklar's Theorem and the Distributional Transform

The notion of copula was introduced in Sklar (1959) to decompose an n -dimensional distribution function F into two parts, the marginal distribution functions F_i and the copula C , describing the dependence part of the distribution.

Definition 1.1 (Copula). Let $X = (X_1, \dots, X_n)$ be a random vector with distribution function F and with marginal distribution functions F_i , $X_i \sim F_i$, $1 \leq i \leq n$. A distribution function C with uniform marginals on $[0, 1]$ is called a “copula” of X if

$$F = C(F_1, \dots, F_n). \quad (1.1)$$

By definition the set of copulas is identical to the Fréchet class $\mathcal{F}(\mathcal{U}, \dots, \mathcal{U})$ of all distribution functions C with uniform marginals distribution functions

$$\mathcal{U}(t) = t, \quad t \in [0, 1]. \quad (1.2)$$

For the case that the marginal distribution functions of F are continuous it is easy to describe a corresponding copula. Define C to be the distribution function of $(F_1(X_1), \dots, F_n(X_n))$. Since $F_i(X_i) \sim U(0, 1)$ have uniform distribution C is a copula and furthermore we obtain the representation

$$\begin{aligned} C(u_1, \dots, u_n) &= P(F_1(X_1) \leq u_1, \dots, F_n(X_n) \leq u_n) \\ &= P(X_1 \leq F_1^{-1}(u_1), \dots, X_n \leq F_n^{-1}(u_n)) \\ &= F_X(F_1^{-1}(u_1), \dots, F_n^{-1}(u_n)). \end{aligned} \quad (1.3)$$

Here F_i^{-1} denotes the generalized inverse of F_i , the “quantile transform”, defined by

$$F_i^{-1}(t) = \inf\{x \in \mathbb{R}^1; F_i(x) \geq t\}.$$

C is a copula of F since by definition of C

$$\begin{aligned} F(x_1, \dots, x_n) &= P(X_1 \leq x_1, \dots, X_n \leq x_n) \\ &= P(F_1(X_1) \leq F_1(x_1), \dots, F_n(X_n) \leq F_n(x_n)) \\ &= C(F_1(x_1), \dots, F_n(x_n)). \end{aligned} \quad (1.4)$$

The argument for the construction of the copula is based on the property that for continuous distribution functions F_i , $F_i(X_i)$ is uniformly distributed on $(0, 1)$: $F_i(X_i) \sim U(0, 1)$. There is a simple extension of this transformation which we call “distributional transform.”

Definition 1.2 (Distributional transform). Let Y be a real random variable with distribution function F and let V be a random variable independent of Y , such that $V \sim U(0, 1)$, i.e. V is uniformly distributed on $(0, 1)$. The modified distribution function $F(x, \lambda)$ is defined by

$$F(x, \lambda) := P(X < x) + \lambda P(Y = x). \quad (1.5)$$

We call

$$U := F(Y, V) \quad (1.6)$$

the (generalized) “distributional transform” of Y .

For continuous distribution functions F , $F(x, \lambda) = F(x)$ for all λ , and $U = F(Y) \stackrel{d}{=} U(0, 1)$, $\stackrel{d}{=}$ denoting equality in distribution. This property is easily extended to the generalized distributional transform (see e.g. Rü¹ (2005)).

Proposition 1.3 (Distributional transform). *Let $U = F(Y, V)$ be the distributional transform of Y as defined in (1.6). Then*

$$U \stackrel{d}{=} U(0, 1) \text{ and } Y = F^{-1}(U) \text{ a.s.} \quad (1.7)$$

An equivalent way to introduce the distributional transform is given by

$$U = F(Y-) + V(F(Y) - F(Y-)), \quad (1.8)$$

where $F(y-)$ denotes the left-hand limit. Thus at any jump point of the distribution function F one uses V to randomize the jump height.

The distributional transform is a useful tool which allows in many respects to deal with general (discontinuous) distributions similar as with continuous distributions. In particular it implies a simple proof of Sklar's Theorem in the general case (see Moore and Spruill (1975) and Rü (1981b, 2005)).

Theorem 1.4 (Sklar's Theorem). *Let $F \in \mathcal{F}(F_1, \dots, F_n)$ be an n -dimensional distribution function with marginals F_1, \dots, F_n . Then there exists a copula $C \in \mathcal{F}(\mathcal{U}, \dots, \mathcal{U})$ with uniform marginals such that*

$$F(x_1, \dots, x_n) = C(F_1(x_1), \dots, F_n(x_n)). \quad (1.9)$$

Proof. Let $X = (X_1, \dots, X_n)$ be a random vector on a probability space $(\Omega, \mathfrak{A}, P)$ with distribution function F and let $V \sim U(0, 1)$ be independent of X . Considering the distributional transforms $U_i := F_i(X_i, V)$, $1 \leq i \leq n$, we have by Proposition 1.3 $U_i \stackrel{d}{=} U(0, 1)$ and $X_i = F_i^{-1}(U_i)$ a.s., $1 \leq i \leq n$. Thus defining C to be the distribution function of $U = (U_1, \dots, U_n)$ we obtain

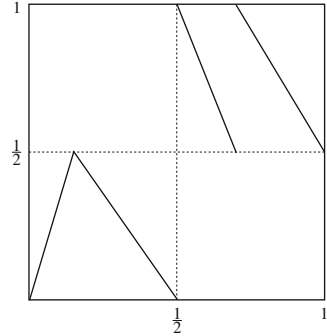
$$\begin{aligned} F(x) &= P(X \leq x) = P(F_i^{-1}(U_i) \leq x_i, 1 \leq i \leq n) \\ &= P(U_i \leq F_i(x_i), 1 \leq i \leq n) = C(F_1(x_1), \dots, F_n(x_n)), \end{aligned}$$

i.e. C is a copula of F . □

Remark 1.5. (a) Copula and dependence. From the construction of the distributional transform it is clear that the distributional transform is not unique in the case when the distribution has discrete parts. Different choices of the

¹Within the whole book the author's name is abbreviated to Rü.

Figure 1.1 Copula of uniform distribution on line segments (published in: *Journal of Stat. Planning Inf.*, 139: 3921–3927, 2009)



randomizations V at the jumps or in the components, i.e. choosing $U_i = F_i(X_i, V_i)$, may introduce *artificial* local dependence between the components of a random vector on the level of the copula. From the copula alone one does not see whether some local positive or negative dependence is a real one or just comes from the choice of the copula. For dimension $n = 2$ the copula in Figure 1.1 could mean a real switch of local positive and negative dependence for the original distribution, but it might also be an artefact resulting from the randomization in case the marginals are e.g. both two-point distributions while the joint distribution in this case could be even comonotone. Thus the copula information alone is not sufficient to decide all dependence properties.

- (b) **Conditional value at risk.** A more recent application of the distributional transform is to risk measures. It is well known that the conditional tail expectation

$$\text{TCE}_\alpha(X) := -E(X \mid X \leq q_\alpha), \quad (1.10)$$

where q_α is the lower α -quantile of the risk X , does not define a coherent risk measure except when restricted to continuous distributions. This defect can be overcome by using the distributional transform $U = F(X, V)$ and defining the modified version, which we call conditional value at risk (CVR_α)

$$\text{CVR}_\alpha(X) = -E(X \mid U \leq \alpha). \quad (1.11)$$

By some simple calculations (see Burgert and Rü (2006b)) one sees that

$$\text{CVR}_\alpha(X) = -\frac{1}{\alpha} [EX1(X < q_\alpha) + q_\alpha(\alpha - P(X < q_\alpha))] = ES_\alpha(X). \quad (1.12)$$

Thus the more natural definition of CVR_α coincides with the well-established “expected shortfall risk measure” $ES_\alpha(X)$ which is a coherent risk measure.

As a consequence the expected shortfall is represented as conditional expectation and our definition in (1.11) of the conditional value at risk seems to be appropriate for this purpose.

- (c) **Stochastic ordering.** The construction of copulas based on the distributional transform as in the proof of Sklar's Theorem above has been used in early papers on stochastic ordering. The following typical example of this type of result is from Rü (1981b, Proposition 7).

Proposition 1.6. *Let F_i, G_i be one-dimensional distribution functions with $F_i \leq G_i$ (or equivalently $G_i \leq_{\text{st}} F_i$), $1 \leq i \leq n$. Then to any $F \in \mathcal{F}(F_1, \dots, F_n)$ there exists an element $G \in \mathcal{F}(G_1, \dots, G_n)$ with $G \leq_{\text{st}} F$. Here \leq_{st} denotes the multivariate stochastic ordering w.r.t. increasing functions.*

Proof. Let $X = (X_1, \dots, X_n) \sim F$ and let $U_i = F_i(X_i, V)$ denote the distributional transforms of the components X_i . Then $U = (U_1, \dots, U_n)$ is a copula vector of F . Define $Y = (Y_1, \dots, Y_n)$ as vector of the quantile transforms of the components of U , $Y_i = G_i^{-1}(U_i)$. Then $Y \sim G \in \mathcal{F}(G_1, \dots, G_n)$ and from the assumption $F_i \leq G_i$ we obtain that $Y \leq X$ pointwise. In consequence $G \leq_{\text{st}} F$. \square

In particular the above argument shows that $G \leq_{\text{st}} F$ if F and G have the same copula. \diamond

1.2 Copula Models and Copula Constructions

In order to construct a distributional model for a random vector X it is by Sklar's representation theorem sufficient to specify a copula model fitting to the normalized data. There are a large number of suggestions and principles for the construction of copula models. Classical books on these models for $n = 2$ are Mardia (1962) and Hutchinson and Lai (1990). More recent books for $n \geq 2$ are Joe (1997), Nelsen (2006), Mari and Kotz (2001), Denuit et al. (2005), and McNeil et al. (2005b). A lot of material on copulas can also be found in the conference volumes of the "Probability with given marginals" conferences – i.e. the series of conferences in Rome (1990), Seattle (1993), Prague (1996), Barcelona (2000), Québec (2004), Tartu (2007) and Sao Paulo (2010) as well as in the recent proceedings edited by Jaworski, Durante, Härdle, and Rychlik (2010). A huge basket of models has been developed and some of its properties have been investigated, concerning in particular the following questions:

- What is the "range of dependence" covered by these models and measured with some dependence index. Is the range of dependence wide and flexible enough?
- What models exhibit tail dependence and thus are able to model situations with strong positive dependence in tails?
- Is some parameter available describing the degree of dependence?

- Is there a natural probabilistic representation respectively interpretation of these models describing situations when to apply them?
- Is there a closed form of the copula or a simple simulation algorithm so that goodness of fit test can be applied to evaluate whether they fit the data?

Several questions of this type have been discussed in the nice survey papers of Schweizer (1991) and Durante and Sempi (2010).

1.2.1 Some Classes of Copulas

Copulas, their properties and applications are discussed in detail in the above-mentioned literature. Copulas are not for any class of distributions the suitable standardizations as for example for elliptical distributions. In extreme value theory a standardization by the exponential distribution may be better adapted. For some applications it may be more natural and simpler to model the distributions or their densities directly without referring to copulas. Conceptually and for some investigation of dependence properties like tail-dependence the notion of copula is however a useful tool.

Some classical classes of copulas are the following.

- The “Farlie–Gumbel–Morgenstern (FGM) copula”:

$$C_\alpha(u, v) = uv(1 + \alpha(1 - u)(1 - v)), \quad |\alpha| \leq 1 \quad (1.13)$$

as well as its generalized versions, “EFGM copulas” are extensions of FGM to the n -dimensional case and given by

$$C(u) = \prod_{i=1}^n u_i \left(1 + \sum_{T \subset \{1, \dots, n\}} \alpha_T \prod_{j \in T} (1 - u_j) \right) \quad (1.14)$$

for suitable $\alpha_T \in \mathbb{R}^1$.

- The “Archimedean copulas”

$$C_\Phi(u_1, \dots, u_n) = \Phi^{-1}(\Phi(u_1) + \dots + \Phi(u_n)) \quad (1.15)$$

for some generator $\Phi : (0, 1] \rightarrow \mathbb{R}_+$ with $\Phi(1) = 0$ such that Φ is completely monotone. In fact C_Φ is an n -dimensional copula if and only if Φ is “ n -monotone”, i.e. Φ is $(n - 2)$ times differentiable and

$$(-1)^k \Phi^{(k)}(t) \geq 0, \quad 0 \leq k \leq n - 2, \forall t > 0 \quad (1.16)$$

and $(-1)^{n-2} \Phi^{(n-2)}$ is decreasing and convex (see McNeil and Nešlehová (2010)).

In case $\Phi(t) = \frac{1}{\alpha}(t^{-\alpha} - 1)$ one gets the “Clayton copula”

$$C_\alpha(u) = \left(\sum u_i^{-\alpha} - n + 1 \right)^{-1/\alpha}, \quad \alpha > 0. \quad (1.17)$$

In case $\Phi_\delta(t) = -\frac{1}{\delta} \log(1 - (1 - e^{-\delta})e^{-t})$ one gets the “Frank copula”

$$C_\delta(u) = -\frac{1}{\delta} \log \left(1 + \frac{\prod_{i=1}^n (e^{-\delta u_i} - 1)}{(e^{-\delta} - 1)^{n-1}} \right). \quad (1.18)$$

Φ is by Bernstein’s theorem completely monotone if and only if it is n -monotone for all $n \in \mathbb{N}$.

Archimedean copulas are connected with mixing models (“frailty models”). Let

$$F_X(x) = \int \prod_{i=1}^n (G_i(x_i))^\delta dF_\Psi(\delta),$$

where Ψ is a positive real mixing variable. Then with L_Ψ , the Laplace transform of Ψ , we obtain that the i -th marginal is given by

$$F_i(x_i) = \int \exp(\delta \ln G_i(x_i)) dF_\Psi(\delta) = L_\Psi(-\ln G_i(x_i)).$$

As a consequence one gets

$$F_X(x) = L_\Psi \left(\sum_{i=1}^n L_\Psi^{-1}(F_i(x_i)) \right), \quad (1.19)$$

i.e. a representation as in (1.15) holds with the generator $\Phi = L_\Psi$ which by Bernstein’s theorem is completely monotone.

Archimedean copulas arise from multivariate distributions that have a stochastic representation of the form

$$X \stackrel{d}{=} R \cdot U \quad (1.20)$$

where U is uniformly distributed on the unit simplex (1-sphere) in \mathbb{R}_+^n , $\{x \in \mathbb{R}_+^n; \sum_{i=1}^n x_i = 1\}$ respectively the simplex in \mathbb{R}^n , and R is a scaling random variable with values in \mathbb{R}_+ independent of U . Formulae for the relation between the distribution function F_R and the generator Φ are available (see McNeil and Nešlehová (2010)).

- “Elliptical distributions” arise from stochastic representations of the form

$$X = \mu + RAU, \quad (1.21)$$

where U is uniformly distributed on the unit 2-sphere $S_{n-1} = \{x \in \mathbb{R}^n; \sum_{i=1}^n u_i^2 = 1\}$ in \mathbb{R}^n . R is a real positive random variable independent from U and A is an $n \times n$ -matrix. X has a density of the form

$$f(x) = |\Sigma|^{-1} \Phi((x - \mu)^\top \Sigma^{-1} (x - \mu)) \text{ with } \Sigma = AA^\top.$$

The distribution is denoted as $\mathcal{E}(\mu, \Sigma, F_R)$. Multivariate normal distributions are elliptical (or elliptically contoured). The copulas of elliptical distributions are in general not available in closed form.

- “Extreme value copulas” are defined as possible limits of max sequences. Let $X_i = (X_{i,1}, \dots, X_{i,d})$, $1 \leq i \leq n$ (note that here d = dimension but n = sample index) be a sequence of iid random vectors with copula C_F . Let $M_n = \max\{X_1, \dots, X_n\} = (M_{n,1}, \dots, M_{n,d})$ be the componentwise max of the X_i , then the copula C_n of M_n is given by

$$C_n(u) = C_F^n(u_1^{1/n}, \dots, u_d^{1/n}).$$

C is called an “extreme value copula” if for any $n \in \mathbb{N}$ there exist some copula C_F such that

$$C = C_F^n(u_1^{1/n}, \dots, u_d^{1/n}). \quad (1.22)$$

A well-known fact is:

C is an extreme value copula if and only if C is “max stable”, i.e. $C(u) = C^m(u_1^{1/m}, \dots, u_d^{1/m})$ for all $m \in \mathbb{N}$.

De Haan and Resnick (1977) and Pickands (1981) gave the following representation of extreme value copulas:

C is an extreme value copula if and only if it has a representation of the form

$$C(u) = \exp(-l(-\log u_1, \dots, -\log u_d)). \quad (1.23)$$

Here the “tail dependence function” $l : (0, \infty)^d \rightarrow [0, \infty]$ is given by

$$l(x) = \int_{S_{d-1}} \max_{j \leq d} (w_j x_j) dH(w) \quad (1.24)$$

where H is a measure on the simple S_{d-1} satisfying $\int_{S_{d-1}} w_j dH(w) = 1$, $1 \leq j \leq d$. H is called the “spectral measure” of C .

An example of copulas which are of interest in extreme value theory are “Gumbel copulas” given by

$$C_\delta(u) = \exp\left(-\left(\sum_{i=1}^n (-\log u_i)^\delta\right)^{1/\delta}\right) \quad (1.25)$$

for $\delta \in [1, \infty)$. Gumbel copulas are extreme value copulas. The parameter δ is a dependence parameter. If $1 \leq \delta_1 < \delta_2$, then G_{δ_2} is more strongly dependent (in the sense of supermodular ordering) than G_{δ_1} (see Wei and Hu (2002)).

1.2.2 Copulas and L^2 -Projections

A general principle for the construction of dependence models (copulas) was introduced in Rü (1985). It is based on the following representation result. Let $\mathcal{C}_n(\mathbb{X}^n)$ denote the set of all copulas on $[0, 1]^n$ which are Lebesgue-continuous and thus have Lebesgue-densities f . Let $\mathcal{C}_n^s(\mathbb{X}^n)$ be the set of all signed measures on $[0, 1]^n$ with uniform marginals and which have Lebesgue-densities f .

For any integrable $f \in L^1([0, 1]^n, \mathbb{X}^n)$ and $T \subset \{1, \dots, n\}$ we define

$$f_T = \int f \prod_{i \in T} dy_i, \quad (1.26)$$

i.e. we integrate over the components in T . We consider f_T as a real function on $[0, 1]^n$ being constant in the T components. The following linear operator $S : L^1(\mathbb{X}^n) \rightarrow L^1(\mathbb{X}^n)$ is of interest,

$$Sf = f - \sum_{\substack{T \subset \{1, \dots, n\} \\ |T| = n-1}} f_T + (n-1) f_{\{1, \dots, n\}} \quad (1.27)$$

since it leads to the following representation result.

Theorem 1.7 (Copulas and L^2 -projections). *All probability distributions on $[0, 1]^n$ with Lebesgue-density and uniform marginals are of the form $(1 + Sf)\mathbb{X}^n$ where $f \in L^1(\mathbb{X}^n)$, more precisely*

$$\mathcal{C}_n^s(\mathbb{X}^n) = \{Q = (1 + Sf)\mathbb{X}^n; \ f \in L^1(\mathbb{X}^n)\} \quad (1.28)$$

and

$$\mathcal{C}_n(\mathbb{X}^n) = \{P = (1 + Sf)\mathbb{X}^n; \ f \in L^1(\mathbb{X}), \ Sf \geq -1\}. \quad (1.29)$$

Proof. Any $P \in \mathcal{C}_n^s(\mathbb{X}^n)$ has by the Radon–Nikodým theorem a density of the form $P = (1 + f)\mathbb{X}^n$ where $f_T = 0$ for $|T| \geq n-1$. This implies that $Sf = f$, i.e. $P = (1 + Sf)\mathbb{X}^n$. Conversely, let $f \in L^1(\mathbb{X}^n)$ and $P = (1 + Sf)\mathbb{X}^n$. Then for $T_0 \subset \{1, \dots, n\}$, $|T_0| = n-1$ and any further $T \subset \{1, \dots, n\}$ with $|T| = n-1$, $T \neq T_0$ we have $(f_T)_{T_0} = f_{\{1, \dots, n\}}$ and, therefore,

$$(Sf)_{T_0} = \left(f - f_{T_0} - \sum_{\substack{|T|=n-1 \\ T \neq T_0}} f_T + (n-1) f_{\{1, \dots, n\}} \right)_{T_0} \quad (1.30)$$

$$= f_{T_0} - f_{T_0} - \sum_{\substack{|T|=n-1 \\ T \neq T_0}} (f_T)_{T_0} + (n-1) f_{\{1, \dots, n\}} = 0. \quad (1.31)$$

Therefore $Q = (1 + Sf)\lambda^n \in \mathcal{C}_n^s(\lambda^n)$. This implies (1.28). Equation (1.29) is immediate from (1.28). \square

The operator S in (1.27) can be considered as an L^2 -projection of a (signed) measure $Q = (1 + f)\lambda^n$, $f \in L^1(\lambda^n)$ to a copula $P = (1 + Sf)\lambda^n$ if $Sf \geq -1$. The basic idea of the construction method for dependence (copula) models in Rü (1985) is to describe in a first step the relevant dependence by the density function $1 + f$. In a second step this model $Q = (1 + f)\lambda^n$ is projected to the class of copulas in the form $P = (1 + Sf)\lambda^n$.

The resulting proposed model construction method is the following:

Construction by “projection method”: Let f_ϑ , $\vartheta \in \Theta$ be a parametric family of functions in $L^1(\lambda^n)$ (describing the intended dependence structure) such that $Sf_\vartheta \geq -1$, $\vartheta \in \Theta$. Then the proposed model obtained by the projection method is given by

$$\mathcal{P} = \{P_\vartheta = (1 + Sf_\vartheta)\lambda^n; \vartheta \in \Theta\}. \quad (1.32)$$

The inherent idea in this construction is the hope that the projection on the correct marginal structure does not change the intended dependence structure too much. Obviously we can use the same procedure as above also for the construction of elements in the Fréchet class $M(P_1, \dots, P_n)$ which are continuous w.r.t. a product measure $\mu_1 \otimes \dots \otimes \mu_n$.

This projection idea is also underlying Sklar's representation theorem. Let $G \in \mathcal{F}_n(G_1, \dots, G_n)$ be a multivariate distribution function with marginals G_1, \dots, G_n and with copula C . Assume that G is a good description of the dependence structure for a model. At the same time assume that the correct marginals should be F_1, \dots, F_n instead of G_1, \dots, G_n . Then by Sklar's Theorem the following “nonmetric projection” would seem natural:

$$G \rightarrow C \rightarrow F = C(F_1, \dots, F_n). \quad (1.33)$$

By this transition the dependence structure is retained in the copula C . The difficulty of calculation may however prevent use of this transition in examples. Typically it will be easier to calculate (only) densities.

The following examples illustrate the projection method, see Rü (1985) and Mari and Kotz (2001, pp. 73–78).

Example 1.8. Let $f \in L^1(\lambda^n)$, let $\alpha_0 = \inf \{(Sf)(x); x \in [0, 1]^n\} > -\infty$ and consider $f_\vartheta(x) = \vartheta f(x)$, $\vartheta \in [0, \frac{1}{|\alpha_0|}]$.

- (1) **Generalized FGM-family:** If $f(x) = \prod_{i=1}^n v_i(x_i)$ where $\int v_i(x_i) dx_i = 0$, $1 \leq i \leq n$, then $Sf = f$ and \mathcal{P} gives a “generalized FGM-family” (see Johnson and Kotz (1975) and Kimeldorf and Sampson (1975a,b), and (1.13)).
- (2) **Polynomial copula models:** If $f(x) = \prod_{i=1}^n x_i^{m_i}$ is a polynomial, then

$$Sf(x) = \prod_{i=1}^n x_i^{m_i} - \sum_{i=1}^n \left(\prod_{j \neq i} \frac{1}{m_j + 1} \right) x_i^{m_i} + (n-1) \prod_{i=1}^n \frac{1}{m_i + 1} \quad (1.34)$$

gives a “polynomial copula model” which is the same as the FGM-model generated by $v_i(x_i) = x_i^{m_i} - \frac{1}{m_i+1}$. Linear combinations then yield the general form of polynomial copula models. The restrictions to the parameters are given in Mari and Kotz (2001, p. 74).

This polynomial model has also been discussed in Wei, Fang, and Fang (1998). Several families of polynomial copulas have been suggested as local approximations of other copula families. So the Ali–Mikhail–Haq family is a polynomial approximate in the neighbourhood of the independence copula, the FGM is an approximation of the Plackett family (see Mari and Kotz (2001, p. 75)).

- (3) **A copula concentrated strongly near the diagonal:** Let $n = 2$ and consider $f(x, y) = \frac{1}{\sqrt{|x-y|}}$. The function f has singularities at the diagonal, i.e. we expect a large density in the neighbourhood of the diagonal. Then by simple calculation one gets

$$Sf(x, y) = \frac{1}{\sqrt{|x-y|}} - 2\left(\sqrt{x} + \sqrt{1-x} + \sqrt{y} + \sqrt{1-y}\right) + \frac{8}{3}. \quad (1.35)$$

\mathcal{P} describes a family of copulas which is concentrated near the diagonal. ϑ is a positive dependence parameter; higher values of ϑ yield higher order positive dependence.

- (4) **Diagonal weakly concentrated copula:** For $\Theta = [0, 1]$, $n = 2$ define

$$f_{\vartheta}(x, y) = -1_{\{|x-y| > \vartheta\}},$$

so $(1 + f_{\vartheta})\lambda^2$ is concentrated in a neighbourhood of the diagonal but it is uniformly distributed on $\{(x, y) \in [0, 1]^2 : |x - y| \leq \vartheta\}$ in contrast to the singular concentration in Example 1.8, (1.8).

By the projection to the correct marginals we obtain

$$Sf_{\vartheta}(x, y) = f_{\vartheta}(x, y) + g_{\vartheta}(x) + g_{\vartheta}(y) - (1 - \vartheta)^2, \quad (1.36)$$

where

$$g_{\vartheta}(x) = (2 - 2\vartheta)1_{\{\vartheta \leq x \leq 1-\vartheta\}} + (1 - x - y)1_{\{0 \leq x < \vartheta\}} - (x - y)1_{\{x > 1-\vartheta\}}$$

for $0 \leq \vartheta \leq \frac{1}{2}$

and $g_{\vartheta}(x) = (x - \vartheta)1_{\{x > \vartheta\}} + (\vartheta - x)1_{\{x < 1-\vartheta\}}$, $\frac{1}{2} < \vartheta \leq 1$.

The projected probability model $P_{\vartheta} = (1 + Sf_{\vartheta})\lambda^2$ approaches for $\vartheta \rightarrow 1$ the product measure while for ϑ small (neglecting *small and large x-values*)

$$1 + (Sf_{\vartheta})(x, y) \approx \begin{cases} 2 - 2\vartheta & \text{for } |x - y| < \vartheta, \\ 1 - 2\vartheta & \text{for } |x - y| > \vartheta. \end{cases} \quad (1.37)$$

In order to introduce a stronger dependence effect one can start with $f_{\vartheta,a}(x, y) = af_{\vartheta}(x, y)$ which for ϑ small and a large centres the distribution near the diagonal. \diamond

1.3 Multivariate Distributional and Quantile Transform

The distributional transform $F(X, V)$ as well as the inverse quantile transform $F^{-1}(V)$ have been generalized to the multivariate case.

Definition 1.9 (Multivariate quantile transform). Let F be a d -dimensional distribution function and let V_1, \dots, V_n be iid $U(0, 1)$ -distributed random variables. Then the “multivariate quantile transform” $Y := \tau_F^{-1}(V)$ is defined recursively as

$$\begin{aligned} Y_1 &:= F_1^{-1}(V_1) \\ Y_k &:= F_{k|1,\dots,k-1}^{-1}(V_k | Y_1, \dots, Y_{k-1}), \quad 2 \leq k \leq n, \end{aligned} \tag{1.38}$$

where $F_{k|1,\dots,k-1}$ denote the conditional distribution functions of the k -th component π_k given the first $k - 1$ components π_1, \dots, π_{k-1} .

The multivariate quantile transform is a basic method to construct a random vector Y with specified distribution function F from a given vector $V = (V_1, \dots, V_n)$ of iid uniformly on $[0, 1]$ distributed random variables. By construction the multivariate quantile transform uses only one-dimensional conditional distribution functions.

Theorem 1.10 (Multivariate quantile transform). *The multivariate quantile transform $Y = \tau_F^{-1}(V)$ is a random vector with distribution function F .*

Proof. In case $n = 2$ and for any $x_1, x_2 \in \mathbb{R}$ we have using the independence of (V_i)

$$\begin{aligned} P(Y_1 \leq x_1, Y_2 \leq x_2) &= \int_{-\infty}^{x_1} P(Y_2 \leq x_2 | Y_1 = y_1) dF_1(y_1) \\ &= \int_{-\infty}^{x_1} P(F_{2|1}^{-1}(V_2 | y_1) \leq x_2 | Y_1 = y_1) dF_1(y_1) \\ &= \int_{-\infty}^{x_1} P(F_{2|1}^{-1}(V_2 | y_1) \leq x_2) dF_1(y_1) \\ &= \int_{-\infty}^{x_1} F_{2|1}(x_2 | y_1) dF_1(y_1) = F_{12}(x_1, x_2). \end{aligned}$$

By induction this argument extends to yield $P(Y_1 \leq x_1, \dots, Y_n \leq x_n) = F(x_1, \dots, x_n)$. \square

The multivariate quantile transform was introduced in O'Brien (1975), Arjas and Lehtonen (1978), and Rü (1981b). The construction Y is an inductive construction of a random vector with specified distribution function F and is called “regression representation” since Y is of the form

$$\begin{aligned} Y_1 &= h_1(V_1), \\ Y_2 &= h_2(Y_1, V_2), \\ Y_3 &= h_2(Y_1, Y_2, V_3), \\ &\vdots \\ Y_n &= h_n(Y_1, \dots, Y_{n-1}, V_n) \end{aligned} \tag{1.39}$$

representing Y_k as function of the past Y_1, \dots, Y_{k-1} and some innovation V_k . As a consequence this implies the “standard representation”

$$Y_k = f_k(V_1, \dots, V_k), \quad 1 \leq k \leq n. \tag{1.40}$$

The multivariate quantile transform and particularly Theorem 1.10 is a basic method for the “simulation of multivariate distributions”. In order to construct a random vector with distribution function F one needs to simulate iid $U(0, 1)$ -random variables and to determine the inverses of the one-dimensional conditional distribution functions $F_{i|x_1, \dots, x_{i-1}}(x_i)$, which can be done in many examples either analytically or numerically. There is a large literature on (Monte Carlo) simulation of multivariate distributions, which uses essentially the multivariate quantile transform. This transform was introduced first in the above-mentioned papers dealing with representation of stochastic sequences and with stochastic orderings.

Definition 1.11 (Multivariate distributional transform). Let X be an n -dimensional random vector and let V_1, \dots, V_n be iid $U(0, 1)$ -distributed random variables, $V = (V_1, \dots, V_n)$. For $\lambda = (\lambda_1, \dots, \lambda_n) \in [0, 1]^n$ define

$$\tau_F(x, \lambda) := (F_1(x_1, \lambda_1), F_2(x_2, \lambda_2 \mid x_1), \dots, F_d(x_d, \lambda_d \mid x_1, \dots, x_{d-1})), \tag{1.41}$$

where

$$\begin{aligned} F_1(x_1, \lambda_1) &= P(X_1 < x_1) + \lambda_1 P(X_1 = x_1), \\ F_k(x_k, \lambda_k \mid x_1, \dots, x_{k-1}) &= P(X_k < x_k \mid X_j = x_j, j \leq k-1) \\ &\quad + \lambda_k P(X_k = x_k \mid X_j = x_j, j \leq k-1) \end{aligned}$$

are the distributional transforms of the one-dimensional conditional distributions. Finally the “multivariate distributional transform” of X is defined as

$$U := \tau_F(X, V). \tag{1.42}$$

Rosenblatt (1952) introduced this transformation in the special case of absolutely continuous conditional distributions allowing the application of the transformation formula. Therefore this transformation is also called the “Rosenblatt transformation” in this case. For general distributions the multivariate distributional transform (generalized Rosenblatt transform) was introduced in Rü (1981b). The basic property is stated in the following theorem in Rü (1981b).

Theorem 1.12 (Multivariate distributional transform). *Let X be a random vector and let $U = \tau_F(X, V)$ denote its multivariate distributional transform. Then*

$$(a) \quad U \sim U((0, 1)^d), \quad (1.43)$$

i.e. the components U_i of U are iid $U(0, 1)$ distributed.

(b) *The multivariate quantile transform τ_F^{-1} is inverse to the multivariate distributional transform, i.e.*

$$X = \tau_F^{-1}(U) = \tau_F^{-1}(\tau_F(X, V)) \text{ a.s.} \quad (1.44)$$

Proof. $U_1 = F_1(X_1, V_1) \sim U(0, 1)$ by Proposition 1.3. Consider next $U_2 = F_2(X_2, V_2 \mid X_1)$. Conditionally given $X_1 = x_1$ we have again from Proposition 1.3 that

$$U_2 \stackrel{d}{=} F_2(X_2, V_2 \mid x_1) \sim U(0, 1).$$

Furthermore, $P^{U_2 \mid X_1 = x_1} = U(0, 1)$ is independent of x_1 and thus U_2, X_1 are independent. Since $U_1 = F_1(X_1, V_1)$ we get from iterated conditional expectation that also U_2, U_1 are independent. The general case then follows by induction. \square

Remark 1.13 (Regression and standard representation).

(a) By combining the multivariate quantile and the multivariate distributional transform one gets for any given stochastic sequence (X_k) a pointwise “standard (innovation) representation”

$$X_k = f_k(U_1, \dots, U_k) \text{ a.s.,} \quad (1.45)$$

respectively a pointwise “regression representation”

$$X_k = f_k(X_1, \dots, X_{k-1}, U_k) \text{ a.s.,} \quad 1 \leq k \leq n \quad (1.46)$$

with some iid sequences (U_i) , $U_i \sim U(0, 1)$. This result was proved first for $n = 2$ in Skorohod (1976). Related functional representations of classes of stochastic models are given in Rü and de Valk (1993). For example in the case of Markov chains the regression representation reduces to the following functional representation of (any) Markov chain:

Corollary 1.14 (Regression representation of Markov chains). *Any Markov chain (X_n) has a representation as a nonlinear regression model*

$$X_k = f_k(X_{k-1}, U_k) \text{ a.s.}, \quad (1.47)$$

where (U_k) is an iid sequence of $U(0, 1)$ -distributed random variables, U_k is independent of X_1, \dots, X_{k-1} .

(b) The copula transformation

$$X = (X_1, \dots, X_d) \rightarrow U = (U_1, \dots, U_d), U_i = F_i(X_i, V_i)$$

which transforms a vector X to a copula vector U , where U corresponds to the copula of X , forgets about the marginals but retains essential information on the dependence structure of X . On the contrary the multivariate distributional transform forgets also about the dependence structure. This is an interesting property, when one wants to identify a distribution. These two different properties of the copula transformation and the multivariate distributional transform lead to different kinds of applications. Some of them are described in Section 1.5. \diamond

1.4 Pair Copula Construction of Copula Models

Besides the multivariate quantile transform in Theorem 1.10, which is based on the one-dimensional conditional distributions $F_{i|x_1, \dots, x_{i-1}}$, several further methods to represent a distribution function F in terms of conditional distribution functions have been proposed. Particular interest in the recent literature and in various applications arose from the “pair copula construction (PCC)” method which is based on a series of certain (organized) pairs of variables.

In the original example Joe (1997) used the following pairwise construction of an m -dimensional distribution function F . For $m = 3$, $F = F_{123}$ can be represented as

$$F_{123}(x) = \int_{-\infty}^{x_2} F_{13|z_2}(x_1, x_3) dF_2(z_2) \quad (1.48)$$

where $F_{13|z_2}$ is the conditional distribution function of the pair X_1, X_3 given $X_2 = z_2$. By Sklar’s Theorem this can also be written in terms of the conditional copula $C_{13|z_2}$ in the form

$$F_{123}(x) = \int_{-\infty}^{x_2} C_{13|z_2}(F_{1|z_2}(x_1), F_{3|z_2}(x_3)) dF_2(z_2). \quad (1.49)$$

Similarly for general m one obtains recursively the representation

$$F_{1\dots m}(x) = \int_{-\infty}^{x_2} \dots \int_{-\infty}^{x_{m-1}} C_{1m|z_2, \dots, z_{m-1}}(F_{1|z_2, \dots, z_{m-1}}(x_1), F_{m|z_2, \dots, z_{m-1}}(x_m)) dF_{2\dots m-1}(z_2, \dots, z_{m-1}) \quad (1.50)$$

which is given in terms of pairwise conditional copulas.

In Bedford and Cooke (2001, 2002) and Kurowicka and Cooke (2006) some general classes of graphical organization principles representing multivariate distributions were developed. See also the survey of Czado (2010). Two basic examples of these classes of constructions are *C*-vines and *D*-vines (*C* = canonical, *D* = drawable).

- (a) ***D*-vines:** The construction of *D*-vines is based on densities and uses the representation

$$f(x_1, \dots, x_n) = \prod_{i=2}^n f_{i|x_1, \dots, x_{i-1}}(x_i) f_1(x_1). \quad (1.51)$$

By Sklar's Theorem we have

$$f_{12}(x_1, x_2) = c_{12}(F_1(x_1), F_2(x_2)) f_1(x_1) f_2(x_2), \quad (1.52)$$

where c_{12} is a bivariate copula density. This implies for the conditional density

$$f_{1|x_2}(x_1) = c_{12}(F_1(x_1), F_2(x_2)) f_1(x_1). \quad (1.53)$$

Using (1.53) for the conditional density of (X_1, X_i) given X_2, \dots, X_{i-1} we obtain by recursion

$$\begin{aligned} f_{i|x_1, \dots, x_{i-1}}(x_i) &= c_{1,i|2, \dots, i-1} \cdot f_{i|x_2, \dots, x_{i-1}}(x_i) \\ &= \left(\prod_{j=1}^{i-2} c_{j,i|j+1, \dots, i-1} \right) c_{i-1,i} \cdot f_i(x_i) \end{aligned} \quad (1.54)$$

using the conditional copula densities

$$c_{j,i|j+1, \dots, i-1} = c_{j,i|x_{j+1}, \dots, x_{i-1}}(F_{j|x_{j+1}, \dots, x_{i-1}}(x_j), F_{i|x_{j+1}, \dots, x_{i-1}}(x_i)).$$

As a result we obtain the “*D*-vine” density decomposition

$$\begin{aligned} f(x_1, \dots, x_n) &= \left(\prod_{i=2}^n \prod_{j=1}^{i-2} c_{j,i|j+1, \dots, i-1} \right) \prod_{i=2}^n c_{i-1,i} \prod_{l=1}^n f_l(x_l) \\ &= \left(\prod_{i=1}^{n-1} \prod_{j=1}^{n-i} c_{j,i+j|(j+1, \dots, j+i-1)} \right) \prod_{l=1}^n f_l(x_l). \end{aligned} \quad (1.55)$$

The conditional copula densities in (1.55) are evaluated at the conditional distribution functions $F_{j|x_{j+1}, \dots, x_{j+i-1}}, F_{i+j|x_{j+1}, \dots, x_{j+i-1}}$.

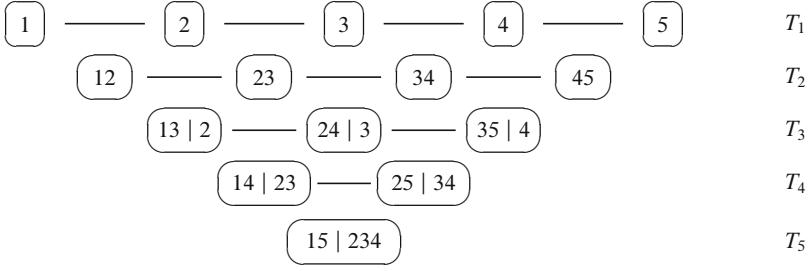


Figure 1.2 *D*-vine tree for $n = 5$

The *D*-vine decomposition in (1.55) can be organized iteratively by pairwise copulas using iteratively levels T_1, \dots, T_{n-1} . For $n = 5$ we obtain the representation

$$\begin{aligned}
 f(x_1, \dots, x_5) \\
 &= \prod_{l=1}^5 f_l(x_l) c_{12} \cdot c_{23} \cdot c_{34} \cdot c_{45} \cdot c_{13|2} \cdot c_{24|3} \cdot c_{35|4} \cdot c_{14|23} \cdot c_{25|34} \cdot c_{15|234}.
 \end{aligned}$$

This is described in the following graphical organization. Each transition step from level T_i to level T_{i+1} involves a (conditional) pair copula (Figure 1.2).

- (b) **C-vines:** *C*-vine decompositions are obtained when applying the representation of the conditional density successively to the conditional distribution of X_{i-1}, X_i given X_1, \dots, X_{i-2} . This gives

$$f_{i|X_1, \dots, X_{i-1}}(x_i) = c_{i-1, i|X_1, \dots, X_{i-2}} f_{i|X_1, \dots, X_{i-1}}(x_i). \quad (1.56)$$

Using (1.56) instead of (1.54) in (1.51) we obtain the *C*-vine decomposition

$$\begin{aligned}
 f(x_1, \dots, x_n) &= f_1(x_1) \prod_{i=2}^n \prod_{k=1}^{i-1} c_{i-k, i|1, \dots, i-k-1} f_k(x_i) \\
 &= \left(\prod_{i=2}^n \prod_{k=1}^{i-1} c_{i-k, i|1, \dots, i-k-1} \right) \prod_{k=1}^n f_k(x_k) \\
 &= \left(\prod_{j=1}^{n-1} \prod_{i=1}^{n-j} c_{j, j+i|1, \dots, j-1} \right) \prod_{k=1}^n f_k(x_k). \quad (1.57)
 \end{aligned}$$

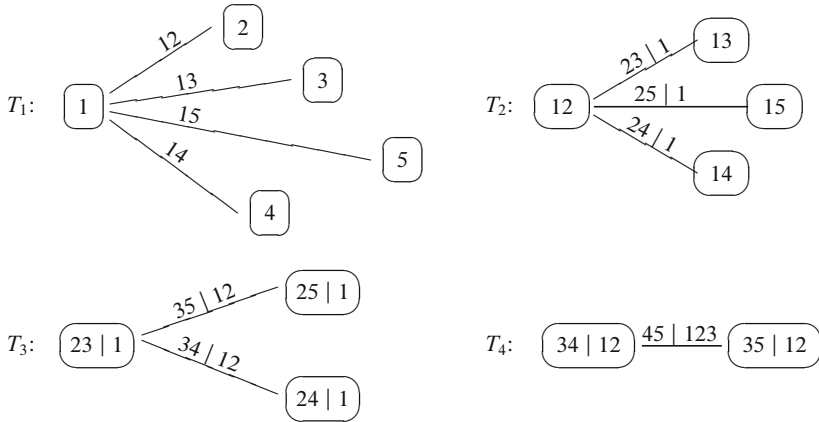


Figure 1.3 C-vine for $n = 5$

For $n = 5$ (1.57) gives a decomposition of the form

$$f(x_1, \dots, x_5) = \prod_{k=1}^5 f_k(x_k) c_{12} c_{13} c_{14} c_{15} c_{23|1} c_{24|1} c_{25|1} c_{34|12} c_{35|12} c_{45|123},$$

which is represented by the following graph in Figure 1.3 with levels T_1, \dots, T_4 . Again each transition of level T_i to T_{i+1} involves (conditional) pair copulas.

Remark 1.15. (a) PCC and Markov random fields:

More general systems to organize pairwise conditional copulas in order to represent uniquely an n -dimensional distribution function respectively density are described in Bedford and Cooke (2001, 2002) and Kurowicka and Cooke (2006) under the denomination “regular vine”. These decompositions are represented graphically as a nested tree. Edges of the tree denote the indices used in the conditional copula densities. The representation of the density in terms of these conditional distributions is an analogue of the Hammersley–Clifford theorem for Markov random field specifications by certain conditional distributions (of the components given their neighbourhoods). Depending on the choice of the neighbourhoods there is some strong similarity between both ways of constructing multivariate (spatial) models.

(b) Reduction of vine representation:

Haff, Aas, and Frigessi (2010) consider a reduction of the (great) complexity of regular vine models by assuming that the conditional copula densities $c_{j,i+j|j+1,\dots,j+i-1}$ respectively copulas $C_{1j|2,\dots,j-1}$ do not depend on the variables x_k in the condition but they depend only through the conditional distribution functions $F_{j|x_{j+1},\dots,x_{j+i-1}}$ etc. This reduction seems to yield good approximations in several applications. \diamond

1.5 Applications of the Distributional Transform

1.5.1 Application to Stochastic Ordering

Let \leq_{st} denote the usual stochastic ordering on \mathbb{R}^n , i.e. the integral induced ordering w.r.t. the class \mathcal{F}_i of increasing functions which is defined by

$$X \leq_{\text{st}} Y \text{ if } Ef(X) \leq Ef(Y) \text{ for all } f \in \mathcal{F}_i \quad (1.58)$$

such that the expectations exist.

The following sufficient condition for the multivariate stochastic order is a direct consequence of Theorem 1.10 on the multivariate quantile transform.

Proposition 1.16. *Let X, Y be n -dimensional vectors with distribution functions $F, G \in \mathcal{F}_n$ and let $(V_i)_{1 \leq i \leq n}$ be iid, $U(0, 1)$ -distributed, then*

$$\tau_F^{-1}(V) \leq \tau_G^{-1}(V) \text{ implies that } X \leq_{\text{st}} Y. \quad (1.59)$$

Condition (1.59) is stated in Rü (1981b). It implies various sufficient conditions for stochastic ordering going back to classical results of Veinott (1965), Kalmykov (1962), and Stoyan (1972). The comparison result of Veinott (1965) states

Corollary 1.17 (Comparison w.r.t. stochastic order \leq_{st}). *Let X, Y be n -dimensional random vectors such that $X_1 \leq_{\text{st}} Y_1$ and for $2 \leq i \leq n$*

$$P^{X_i | X_1=x_1, \dots, X_{i-1}=x_{i-1}} \leq_{\text{st}} P^{Y_i | Y_1=y_1, \dots, Y_{i-1}=y_{i-1}} \quad (1.60)$$

for all $x_j \leq y_j, 1 \leq j \leq i-1$, then

$$X \leq_{\text{st}} Y.$$

Proof. Condition (1.60) implies (by induction) that $\tau_F^{-1}(V) \leq \tau_G^{-1}(V)$ where $F = F_X, G = F_Y$. Since $\tau_F^{-1}(V) \stackrel{d}{=} X$ and $\tau_G^{-1}(V) \stackrel{d}{=} Y$ this implies $X \leq_{\text{st}} Y$. \square

The standard construction in (1.45) respectively the regression representation in (1.46) however are not applicable in general when $P \leq_{\text{st}} Q$, i.e. they do not in general produce pointwise a.s. constructions $X \sim P, Y \sim Q$ such that $X \leq Y$ a.s. The existence of such a coupling is however true under general conditions as follows from the following theorem due to Strassen (1965).

Let (S, \leq) be a Polish space supplied with a closed semiorde “ \leq ”, i.e., the set $\{(x, y) \in S \times S; x \leq y\}$ is closed w.r.t. the product topology. The closed order \leq induces the “stochastic order” \leq_{st} on the set $M^1(S)$ of probability measures on S defined by

$$P \leq_{\text{st}} Q \text{ if } \int f dP \leq \int f dQ \quad (1.61)$$

for all integrable increasing real functions $f \in \mathcal{F}_i = \mathcal{F}_i(S, \leq)$.

Theorem 1.18 (Strassen's ordering theorem, Strassen (1965)).

Let (S, \leq) be a Polish space supplied with a closed partial order. Let $P, Q \in M^1(S)$, then

$$P \leq_{\text{st}} Q$$

\Leftrightarrow There exist random variables $X \sim P, Y \sim Q$ such that $X \leq Y$ a.s.

In Section 3.4 we will discuss extensions of this a.s. ordering theorem to the class of “integral induced orderings $\leq_{\mathcal{F}}$ ” for some function class \mathcal{F} . These are defined via

$$P \leq_{\mathcal{F}} Q \text{ if } \int f dP \leq \int f dQ \text{ for all integrable } f \in \mathcal{F}.$$

Early sources for integral induced orderings are Marshall and Olkin (1979), Rü (1979), and Whitt (1986).

The regression and standard constructions are used essentially in various papers and textbooks on stochastic ordering and are closely connected with some notions of stochastic ordering respectively dependence orderings. We state as one example the notion of conditional increasing in sequence (CIS).

Definition 1.19 (Conditional increasing in sequence (CIS)).

A random vector $X = (X_1, \dots, X_n)$ is called “conditional increasing in sequence” (CIS) if for $2 \leq i \leq n$

$$X_i \uparrow_{\text{st}} (X_1, \dots, X_{i-1}), \quad (1.62)$$

i.e. the conditional distribution $P_{i|x_1, \dots, x_{i-1}} = P^{X_i|X_1=x_1, \dots, X_{i-1}=x_{i-1}}$ is stochastically increasing in (x_1, \dots, x_{i-1}) .

The CIS-property of a random vector X is a positive dependence property of X . This property is equivalent to the condition that the standard representation based on the multivariate quantile transform $Y = \tau_F^{-1}(V)$ is an increasing function in V , i.e.

$$Y_k = f_k(V_1, \dots, V_k) \text{ are increasing in } V, \quad 2 \leq k \leq n. \quad (1.63)$$

Proposition 1.20 (CIS and multivariate quantile transform).

Let X be a random vector with distribution function F , then X is CIS if and only if the construction by the multivariate quantile transformation $Y = \tau_F^{-1}(V) = (f_1(V_1), f_2(V_1, V_2), \dots, f_n(V_1, \dots, V_n))$ is monotonically increasing in V .

1.5.2 Optimal Couplings

The multivariate quantile transform is by Theorems 1.10 and 1.12 a basic construction method for random vectors. An extension of this construction leads to interesting connections with the construction of (optimal) couplings of distributions.

Let $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$ be a measurable function and let $P \in M^1(\mathbb{R}^n, \mathfrak{B}^n)$ have distribution function F . Let furthermore S, V be random variables on $(\Omega, \mathfrak{A}, R)$ such that the “distributional equation”

$$P^h = R^S,$$

holds, i.e. h and S have the same distributions w.r.t. P respectively R .

Theorem 1.21 (Solution of stochastic equations, Rü (1985) and Rachev and Rü (1991)). *Let S be a random variable on a probability space $(\Omega, \mathfrak{A}, R)$ that satisfies the distributional equation*

$$P^h = R^S.$$

Assume that $V = (V_i)_{1 \leq i \leq n}$ are further iid $U(0, 1)$ -distributed random variables on $(\Omega, \mathfrak{A}, R)$ such that S, V are independent. Then there exists a random variable X on $(\Omega, \mathfrak{A}, R)$ such that

- (a) $R^X = P$, i.e. X has distribution function F
- and
- (b) X is a solution of the “stochastic equation”

$$h \circ X = S \text{ a.s. with respect to } R.$$

Remark 1.22. The situation is described by the following diagram:

$$\begin{array}{ccc} (\Omega, \mathfrak{A}, R) & \xrightarrow{S} & (\mathbb{R}^m, \mathfrak{B}^m) \\ \downarrow X & & \nearrow h \\ (\mathbb{R}^n, \mathfrak{B}^n, P) & & \end{array}$$

If the distributional equation $R^S = P^h$ holds, then there exists a *solution* X with $R^X = P$ solving the stochastic equation $h \circ X = S$ a.s. and thus making the diagram commutative. \diamond

Proof. We denote by

$$F_{i|x_1, \dots, x_{i-1}, s} = F_i(\cdot \mid x_1, \dots, x_{i-1}, s) = P^{\pi_i \mid \pi_1=x_1, \dots, \pi_{i-1}=x_{i-1}, h=s}$$

the conditional distribution function of the i -th projection π_i given $\pi_j = x_j$, $j \leq i-1$ and given $h = s$ and define inductively a random vector X by

$$\begin{aligned} X_1 &= F_1^{-1}(V_1 | S), \quad X_2 = F_2^{-1}(V_2 | X_1, S), \dots, \\ X_n &= F_n^{-1}(V_n | X_1, \dots, X_{n-1}, S). \end{aligned} \quad (1.64)$$

Equation (1.64) is an extension of the multivariate quantile transform. By the independence assumptions we obtain

$$R^{X_1|S=s} = R^{F_1^{-1}(V_1|S)|S=s} = R^{F_1^{-1}(V_1|S=s)} = P^{\pi_1|h=s}.$$

Similarly,

$$\begin{aligned} R^{X_2|X_1=x_1, S=s} &= R^{F_2^{-1}(V_2|X_1, S)|X_1=x_1, S=s} \\ &= R^{F_2^{-1}(V_2|x_1, s)} = P^{\pi_2|\pi_1=x_1, h=s} \end{aligned}$$

implying

$$\begin{aligned} R^{(X_1, X_2)|S=s} &= \int R^{X_2|X_1=x_1, S=s} dR^{X_1|S=s} \\ &= \int P^{\pi_2|\pi_1=x_1, h=s} dP^{\pi_1|h=s}(x_2) = P^{(\pi_1, \pi_2)|h=s}. \end{aligned}$$

By induction we find $R^{X|S=s} = P^{\pi|h=s}$ and, therefore, $R^X = P$.

Since almost surely w.r.t. P^h it holds that $P^{\pi|h=s}(\{x; h(x) = s\}) = 1$ we obtain $R^{X|S=s}(\{x; h(x) = s\}) = 1[R^S]$ and thus

$$R(\{h(X) = S\}) = \int R^{X|S=s}(\{x; h(x) = s\}) dR^S(x) = 1. \quad \square$$

Remark 1.23. Using a measure isomorphism argument Theorem 1.21 on the solutions of stochastic equations extends to Borel spaces E, F , replacing $\mathbb{R}^n, \mathbb{R}^m$ (see Rü (1985) and Rachev and Rü (1991)), i.e. let $(E, P) \xrightarrow{h} F$, and $(\Omega, R) \xrightarrow{f} F$ be functions such that the distributional equation

$$P^h = R^f \quad (1.65)$$

holds. If (E, P) is rich enough, i.e. it allows a uniformly $U(0, 1)$ -distributed random variable V on (E, P) independent of h , then there exists a random variable $X : E \rightarrow F$ with $P^X = R$ such that X solves the stochastic equation

$$h = f \circ X[P]. \quad (1.66)$$

◇

An interesting application of Theorem 1.21 is to the construction of optimal couplings. Let $T : (\mathbb{R}^n, \mathfrak{B}^n) \rightarrow (\mathbb{R}^m, \mathfrak{B}^m)$ and for probability measures $P, Q \in \mathcal{M}^1(\mathbb{R}^n, \mathfrak{B}^n)$ define the optimal coupling problem:

$$c_T(P, Q) := \inf \left\{ E \|T(X) - T(Y)\|^2; X \stackrel{d}{=} P, Y \stackrel{d}{=} Q \right\}. \quad (1.67)$$

Equation (1.67) is the optimal coupling of $T(X)$, $T(Y)$ over all possible couplings X, Y of P, Q . Then the following result holds (see Rü (1986)):

Corollary 1.24 (Optimal coupling of T). *Let $P_1 = P^T$, $Q_1 = Q^T$ be the distributions of T under P, Q . Then*

$$c_T(P, Q) = \inf \left\{ E \|T(X) - T(Y)\|^2; X \stackrel{d}{=} P, Y \stackrel{d}{=} Q \right\} = \ell_2^2(P_1, Q_1),$$

where $\ell_2(P_1, Q_1)$ is the “minimal ℓ_2 -metric” of P_1, Q_1 given by

$$\ell_2(P_1, Q_1) = \inf \left\{ (E \|U - V\|^2)^{1/2}; U \stackrel{d}{=} P_1, V \stackrel{d}{=} Q_1 \right\}. \quad (1.68)$$

In case $m = 1$ it holds that

$$\ell_2^2(P_1, Q_1) = \int_0^1 (F_1^{-1}(u) - G_1^{-1}(u))^2 du, \quad (1.69)$$

where F_1, G_1 are the distribution functions of P_1, Q_1 .

1.5.3 Identification and Goodness of Fit Tests

For the construction of a goodness of fit test for the hypothesis $H_0 : F = F_0$ the multivariate distributional transform allows to construct simple test statistics by checking whether the transformed random vectors $Y_i = \tau_{F_0}(X_i, V^i)$, $1 \leq i \leq n$, are uniformly distributed on the unit cube $[0, 1]^d$. Standard tests for this purpose are based on Kolmogorov–Smirnov test statistics $T_m = \sup_{t \in [0, 1]} |\hat{F}_m(t) - t|$ on Cramér–von Mises statistics $\int (\hat{F}_m(t) - t)^2 dt$ or on weighted variants of them. Here $m = dn$ is the sample size and \hat{F}_m the corresponding empirical distribution function. A detailed discussion of this principle and its practical and theoretical properties is given in Prakasa Rao (1987). A main problem for the practical application of this construction method is the calculation of conditional distribution functions.

This principle of standardization is also useful for various other kinds of identification problems and for statistical tests as for example for the test of the two-sample problem $H_0 : F = G$. For this problem we use the empirical version of

the distributional transform based on the pooled sample. We have to check whether the transformed sample is a realization of a $U([0, 1]^d)$ -distributed variety.

1.5.4 Empirical Copula Process and Empirical Dependence Function

We consider the problem of testing or describing dependence properties of multivariate distributions based on a sequence of observations. The construction of test statistics is typically based on some classical dependence measures like Kendall's τ or Spearman's ρ (see Nelsen (2006)) or related dependence functionals. Empirical versions of the dependence functionals can often be represented as functionals of the reduced empirical process, the empirical copula function and the normalized empirical copula process. The distributional transform allows to extend some limit theorems known for the case of continuous distributions to more general distribution classes.

Let $X_j = (X_{j,1}, \dots, X_{j,k})$, $1 \leq j \leq n$ be k -dimensional random vectors with distribution function $F \in \mathcal{F}(F_1, \dots, F_k)$. For the statistical analysis of dependence properties of F a useful tool is the “reduced empirical process”, which is also called “copula process”, and is defined for $t \in [0, 1]^k$ by

$$V_n(t) := \frac{1}{\sqrt{n}} \sum_{j=1}^n (I(U_{j,1} \leq t_1, \dots, U_{j,k} \leq t_k) - C(t)). \quad (1.70)$$

Here $U_j = (U_{j,1}, \dots, U_{j,k})$ are the copula vectors of X_j , $U_{j,i} = F_i(X_{j,i}, V^j)$, and C is the corresponding copula $C(t) = P(U_j \leq t)$.

The construction of the distributional transforms $U_{j,i}$ is based on knowing the marginal distribution functions F_i . If F_i are not known it is natural to use empirical versions of them. Let

$$\hat{F}_i(x_i) = \frac{1}{n} \sum_{j=1}^n 1_{(-\infty, x_i]}(X_{j,i}) \quad (1.71)$$

denote the empirical distribution functions of the i -th components of X_1, \dots, X_n . Then in the case of a continuous distribution function F the empirical counterparts of the distributional transforms are

$$\hat{U}_{j,i} := \hat{F}_i(X_{j,i}), \quad \hat{U}_j = (\hat{U}_{j,1}, \dots, \hat{U}_{j,k}). \quad (1.72)$$

For continuous distribution function F_i we have that

$$n\hat{U}_{j,i} = n\hat{F}_i(X_{j,i}) = R_{j,i}^n \quad (1.73)$$

are the ranks of $X_{j,i}$ in the n -tuple of i -th components $X_{1,i}, \dots, X_{n,i}$ of X_1, \dots, X_n and the ranks $R_{1,i}, \dots, R_{n,i}$ are a.s. a permutation of $1, \dots, n$. The “empirical copula function” is then given by

$$\widehat{C}_n(t) = \frac{1}{n} \sum_{j=1}^n I(\widehat{U}_j \leq t), \quad t \in [0, 1]^k. \quad (1.74)$$

\widehat{C}_n is an estimator of the copula function C . \widehat{C}_n induces the “normalized empirical copula process”

$$\begin{aligned} L_n(t) &:= \sqrt{n} (\widehat{C}_n(t) - C(t)) \\ &= \frac{1}{\sqrt{n}} \sum_{j=1}^n \{I(R_{j,1}^n \leq nt_1, \dots, R_{j,k}^n \leq nt_k) - C(t)\}, \quad t \in [0, 1]^k. \end{aligned} \quad (1.75)$$

This normalized empirical copula process was introduced in Rü (1974, 1976) under the name *multivariate rank order process*. In that paper more generally the sequential version of the process

$$L_n(s, t) = \frac{1}{\sqrt{n}} \sum_{j=1}^{[ns]} \{I(\widehat{U}_j \leq t) - C(t)\}, \quad s \in [0, 1], t \in [0, 1]^k \quad (1.76)$$

was introduced and analysed for nonstationary and mixing random variables.

The empirical copula function \widehat{C}_n was also introduced in Deheuvels (1979) and called “empirical dependence function”. Based on limit theory for the reduced empirical process it is shown in Rü (1974, 1976) and also in a series of papers of Deheuvels starting with Deheuvels (1979) that the normalized empirical copula process converges to a Gaussian process. Several nonparametric measures of dependence like Spearman’s ρ or Kendall’s τ have corresponding empirical versions which can be represented as functionals of L_n . As a consequence one obtains asymptotic distributions for these test statistics for testing dependence properties.

The distributional transform suggests to consider an extension of the empirical copula process to the case of general distribution functions F . The empirical versions of the $U_{j,i}$ are now defined as

$$\widehat{U}_{j,i} = \tau_{F_i}(\widehat{X}_{j,i}, V^j) \quad (1.77)$$

which are exactly $U(0, 1)$ distributed. In order to avoid artificial dependence it is natural to let the copula $C_j(t) = P(U_j \leq t)$, $t \in [0, 1]^k$, be based on the same randomization V^j in all components of the j -th random vector such that $C_j(t) = C(t)$, $1 \leq j \leq n$. We define the normalized empirical copula process by

$$L_n(t) = \sqrt{n}(\widehat{C}_n(t) - C(t)), \quad t \in [0, 1]^k. \quad (1.78)$$

The copula C has bounded nondecreasing partial derivatives a.s. on $[0, 1]^k$ (see Nelsen (2006, p. 11)). Now the proof of Theorem 3.3 in Rü (1976) extends to the case of general distributions.

The basic assumption of this theorem is convergence of the reduced sequential empirical process, the sequential version of V_n in (1.70) (defined as in (1.76) for L_n). This assumption has been established for various classes of independent and mixing sequences of random vectors.

(A) Assume that the reduced sequential process $V_n(s, t)$ converges weakly to an a.s. continuous Gaussian process V_0 in the Skorohod space D_{k+1} .

The additional assumptions on V_0 made in Rü (1976) served there to obtain stronger convergence results or to deal with more general assumptions on the distributions.

Theorem 1.25 (Limit theorem for the normalized empirical copula process, Rü (1976, 2009)). *Under condition (A) the sequential version $L_n(s, t)$ of the normalized empirical copula process converges weakly to the a.s. continuous Gaussian process L_0 given by*

$$L_0(s, t) = V_0(s, t) - s \sum_{i=1}^n \frac{\partial C(t)}{\partial t_i} V_0(1, \dots, 1, t_i, \dots, 1). \quad (1.79)$$

Based on this convergence result asymptotic distributions of test statistics testing dependence properties can be derived as in the continuous case. The proofs are based on representations or approximations of these statistics by functionals of the empirical copula process L_n . For examples of this type see Rü (1974, 1976) and Deheuvels (1979, 1981). For applications to the estimation of dependence functionals and extensions to the empirical tail copula process see Schmidt and Stadtmüller (2006).

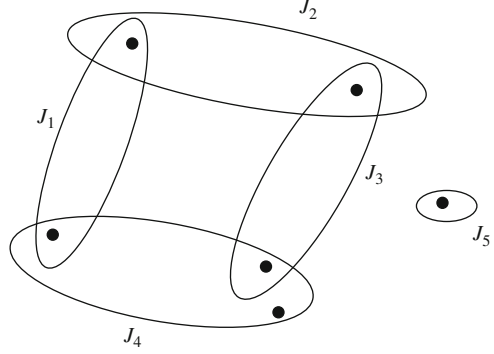
1.6 Multivariate and Overlapping Marginals

In this section we consider the case that not only one-dimensional (marginal) distributions of the risk vector X are known. We assume that also for certain subsets J of the components the joint distribution of $(X_j)_{j \in J}$ is known. This is motivated from available empirical information contained in certain historical data sets or from functional knowledge of the random mechanism.

1.6.1 Generalized Fréchet Class

Let (E_j, \mathfrak{A}_j) , $1 \leq j \leq n$ be n measure spaces and let $\mathcal{E} \subset \mathcal{P}(\{1, \dots, n\})$ be a system of subsets $J \subset \{1, \dots, n\}$ such that $\cup_{J \in \mathcal{E}} J = \{1, \dots, n\}$. Let $P_J \in$

Figure 1.4 Multivariate marginals



$\mathcal{M}^1(E_J, \mathfrak{A}_J)$, $J \in \mathcal{E}$, be a consistent system of probability distributions on $(E_J, \mathfrak{A}_J) = \bigotimes_{j \in J} (E_j, \mathfrak{A}_j)$. We assume that we know the joint distributions of the components in J for all $J \in \mathcal{E}$.

This assumption is a restriction on the joint dependence structure in the model. In comparison the Fréchet class (with only single marginals fixed) includes the set of all possible dependence structures (Figure 1.4).

Definition 1.26 (Generalized Fréchet class). To a given consistent system $(P_J)_{J \in \mathcal{E}}$ of probability measures we define the “generalized Fréchet class” $\mathcal{M}_{\mathcal{E}}$ by

$$\mathcal{M}_{\mathcal{E}} = \mathcal{M}(P_J, J \in \mathcal{E}) = \{P \in \mathcal{M}^1(E, \mathfrak{A}); P^{\pi_J} = P_J, J \in \mathcal{E}\}, \quad (1.80)$$

where $(E, \mathfrak{A}) = \bigotimes_{i=1}^n (E_i, \mathfrak{A}_i)$ and π_J are the projections on the components in J .

Using the generalized Fréchet class as a model class for a risk vector X means that the distribution of $X_J = (X_j)_{j \in J}$ is specified to be P_J for all sets $J \in \mathcal{E}$.

In the particular case that $\mathcal{E} = \{\{1\}, \dots, \{n\}\}$ where \mathcal{E} consists of singletons we get the usual Fréchet class $\mathcal{M}(P_1, \dots, P_n)$. If $\mathcal{E} = \mathcal{E}^s = \{\{i, i+1\}, 1 \leq i \leq n-1\}$ we get the “series case”, where all neighboured pairwise distributions are known. A system \mathcal{E} is called “decomposable” (or “regular”), if there do not exist cycles in \mathcal{E} . The most simple nondecomposable (nonregular) system is given by $\mathcal{E}_3^2 = \{\{1, 2\}, \{2, 3\}, \{1, 3\}\}$ or in more general form by the “pairwise system” $\mathcal{E} = \mathcal{E}_n^2 = \{\{i, j\}; 1 \leq i < j \leq n\}$.

The “marginal problem” is the question whether there exist joint distributions with the given multivariate marginals. A classical result due to Vorobev (1962) and Kellerer (1964) states in the situation of Polish spaces:

Theorem 1.27 (Marginal problem). Let $\mathcal{E} \subset \mathcal{P}(\{1, \dots, n\})$ and (E_j, \mathfrak{A}_j) be Polish spaces. Then the statement

$$\text{Consistency of } (P_J) \text{ implies } \mathcal{M}_{\mathcal{E}} \neq \emptyset$$

is equivalent to the condition that \mathcal{E} is decomposable.

Thus in general consistency is not enough to imply the existence of joint distributions. A simple counterexample is the following. Let $\mathcal{E} = \{\{1, 2\}, \{2, 3\}, \{1, 3\}\}$ and let $P_{12} = P_{23} = P_{13}$ be the distribution of the pair $(U, 1 - U)$ where $U \sim U(0, 1)$. Then this system is consistent but $\mathcal{M}_{\mathcal{E}} = \emptyset$. If there would exist some element $P \in \mathcal{M}_{\mathcal{E}}$ and $X = (X_1, X_2, X_3) \sim P$ then we would get

$$\text{Cov}(X_1, X_3) = \text{Cov}(X_1, 1 - X_2) = -\text{Cov}(X_1, X_2) > 0,$$

a contradiction. Some characterizations of nonemptiness of $\mathcal{M}_{\mathcal{E}}$ are known (see Rü (1991a)), which however are not easy to apply but may serve to produce counterexamples.

Assuming $\mathcal{M}_{\mathcal{E}} \neq \emptyset$, a natural idea to construct submodels $\mathcal{P} \subset \mathcal{M}_{\mathcal{E}}$ describing the dependence structure in a correct way is the following extension of the projection method discussed in Section 1.2 for the simple marginal case. Let $\{P_{\vartheta}; \vartheta \in \Theta\}$ be a parametric class of probability measures on (E, \mathfrak{A}) with densities $f_{\vartheta} \sim P_{\vartheta}$ describing the dependence of the components correctly. Then one may try to determine the projections $P_{\vartheta} \rightarrow P'_{\vartheta} \in \mathcal{M}_{\mathcal{E}}$ w.r.t. some suitable distance in order to fit the marginal structure. The hope is that even after projection the dependence structure is essentially not changed (see Section 1.2).

In the case that $\mathcal{M}_{\mathcal{E}} = \mathcal{M}(P_1, \dots, P_n)$ one can interpret Sklar's Theorem in this sense, i.e. transforming an element $G \in \mathcal{F}(G_1, \dots, G_n)$ to some $F \in \mathcal{F}(F_1, \dots, F_n)$ with the correct marginals F_1, \dots, F_n via the copula C ,

$$G \rightarrow C \rightarrow F = C(F_1, \dots, F_n). \quad (1.81)$$

For the Kullback–Leibler distance the projection is characterized by a density of product form

$$\frac{dP}{dQ} = \prod_{i=1}^n f_i.$$

The “iterative proportional fitting algorithm (IPF)” has been shown to converge to the projection (see Rü (1995a)) under some conditions.

In the case of general overlapping marginals a characterization of all L^2 -projections (with restriction to the probability measures continuous w.r.t. the product measure) is given in Rü (1985). For the Kullback–Leibler distance a product form of the density

$$\frac{dP}{dQ}(x) = \prod_{J \in \mathcal{E}} f_J(x_J) \quad (1.82)$$

is sufficient for the projection and up to a closedness property also necessary.

In general however a natural extension of Sklar's Theorem giving a construction

$$\mathcal{M}(Q_J, J \in \mathcal{E}) \rightarrow \mathcal{M}(P_J, J \in \mathcal{E}) \quad (1.83)$$

for two marginal systems is still an open question.

There are some particular results on the connection described in (1.83). For $P \in \mathcal{M}^1(\mathbb{R}^n, \mathfrak{B}^n)$ denote $\mathcal{C}(P)$ the set of all copulas of P , then the following relations can be easily seen.

- Proposition 1.28.** (a) If $\mathcal{M}_{\mathcal{E}}(P_J, J \in \mathcal{E}) \neq \phi$, then there exist $C_J \in \mathcal{C}(P_J)$, $J \in \mathcal{E}$ such that $\mathcal{M}_{\mathcal{E}}^{\mathcal{C}} := \mathcal{M}_{\mathcal{E}}(C_J, J \in \mathcal{E}) \neq \phi$.
 (b) If C_J are copulas of P_J , $J \in \mathcal{E}$ and $\mathcal{M}_{\mathcal{E}}^{\mathcal{C}} = \mathcal{M}_{\mathcal{E}}(C_J, J \in \mathcal{E}) \neq \phi$, then $\mathcal{M}_{\mathcal{E}} = \mathcal{M}_{\mathcal{E}}(P_J, J \in \mathcal{E}) \neq \phi$.
 (c) In general $\mathcal{M}_{\mathcal{E}}(P_J, J \in \mathcal{E}) \neq \phi$ does not imply $\mathcal{M}_{\mathcal{E}}^{\mathcal{C}}(C_J, J \in \mathcal{E}) \neq \phi$ for any choice of copulas $C_J \in \mathcal{C}(P_J)$.

1.6.2 Copulas with Given Independence Structure

In Section 1.2 the projections w.r.t. L^2 -distance have been described in the simple marginal case. These results have been extended to the multivariate marginal case in Rü (1985). Let \mathcal{E} be an index class $\mathcal{E} = \{T_1, \dots, T_k\}$ with $\cup T_j = \{1, \dots, n\}$ and let P_{T_1}, \dots, P_{T_k} be a given consistent system of distributions on $[0, 1]^{|T_i|}$, $1 \leq i \leq k$. We assume that all univariate marginals are uniform $U(0, 1)$ and thus the generated Fréchet class $\mathcal{M}_{\mathcal{E}}$ is a subclass of the set of all copulas. As in Section 1.2 we restrict in the following to the Lebesgue-continuous elements in $\mathcal{M}_{\mathcal{E}}$

$$\mathcal{M}_{\mathcal{E}}(\mathbb{X}^n) = \{P \in \mathcal{M}_{\mathcal{E}}; P \ll \mathbb{X}^n\}; \quad (1.84)$$

the signed version of this class we denote by $\mathcal{M}_{\mathcal{E}}^s(\mathbb{X}^n)$.

In the first part we consider the special case of distributions with given independence structure, i.e. we assume that

$$P_{T_i} = \mathbb{X}^{|T_i|}, \quad 1 \leq i \leq k. \quad (1.85)$$

Thus we consider the class of probability models for a random vector X such that $(X_j)_{j \in T_i}$ are iid $U(0, 1)$ -distributed for any $i \leq k$.

To describe the corresponding generalized Fréchet class we need a second linear operator V supplementing the operator S defined in Section 1.2 in (1.27).

Define $V : L^1(\mathbb{X}^n) \rightarrow L^1(\mathbb{X}^n)$ inductively by

$$f_{(1)} := f_{R_1}, f_{(m+1)} := f_{(m)} - (f_{(m)})_{R_{m+1}} \text{ for } m < k \text{ and } V(f) := f_{(k)}, \quad (1.86)$$

where $R_m = T_m^c = \{1, \dots, n\} \setminus T_m$ and f_R is obtained from f by integrating over the components in R .

Theorem 1.29 (Distributions with given independence structure). *For the independence structure \mathcal{E} given in (1.85) we have the representation of the generalized Fréchet class $\mathcal{M}_n^s(\mathcal{E})$ by*

$$\mathcal{M}_n^s(\mathcal{E}) = \{(1 + V \circ Sf)\mathbb{X}^n; f \in L^1(\mathbb{X}^n)\}$$

respectively

$$\mathcal{M}_n(\mathcal{E}) = \{(1 + V \circ Sf)\mathbb{X}^n; f \in L^1(\mathbb{X}^n), 1 + V \circ Sf \geq 0\}.$$

The proof of Theorem 1.29 is similar to that of Theorem 1.7 in Section 1.2 (see Rü (1985)).

Special attention has been given in the literature to the case that $\mathcal{E} = \mathcal{E}_k = \{T \subset \{1, \dots, n\}; |T| = k\}$ i.e. the case that all choices of k -components are independent. In this case a more compact representation of the solutions is possible.

Define for $f \in L^1(\mathbb{X}^n)$ and $1 \leq k < n$ inductively linear operators V_1, \dots, V_n by

$$V_1 f = f, V_{k+1} f = V_k f - \sum_{|T|=n-k+1} (V_k f)_T. \quad (1.87)$$

Call a signed measure P “ k -independent” if the distribution of any k -components $(\pi_{i_1}, \dots, \pi_{i_k})$ is $\mathbb{X}^k/[0, 1]^k$.

Theorem 1.30 (k -independent probability measures). *The set of all k -independent (signed) measures has the representation*

$$\mathcal{M}_n^s(\mathcal{E}_k) = \{(1 + V_k \circ Sf)\mathbb{X}^n; f \in L^1(\mathbb{X}^n)\}$$

respectively

$$\mathcal{M}_n(\mathcal{E}_k) = \{(1 + V_k \circ Sf)\mathbb{X}^n; f \in L^1(\mathbb{X}^n), 1 + V_k \circ Sf \geq 0\}. \quad (1.88)$$

This result follows by reduction from Theorem 1.29.

Example 1.31 (FGM-distributions). Consider the generalized FGM-distributions defined as $(1 + f)\mathbb{X}^n$, where $f(x) = \prod_{i=1}^n v_i(x_i)$ such that $\int v_i(x_i) dx_i = 0$, $1 \leq i \leq n$. If $\mathbb{X}^n(\{f = 0\}) < 1$, then $Sf = f$ and furthermore

$$V_1 f = V_2 f = \dots = V_{n-1} f = f. \quad (1.89)$$

This implies that the FGM-distribution $(1 + f)\mathbb{X}^n$ is $(n - 1)$ -independent but not n -independent. This observation indicates the lack of strong dependence in higher dimensional FGM-families. Simultaneously, it gives some natural examples of $(n - 1)$ -independent distributions which are not n -independent. Similarly, $(1 + f)\mathbb{X}^n$ is k -independent but not $(k + 1)$ -independent, where we define f as

$$f(x) = \alpha \sum_{|T|=k} \prod_{j \in T} v_j(x_j), \quad (1.90)$$

α being a factor such that $1 + f \geq 0$.

◇

1.6.3 Copulas, Overlapping Marginals, and L^2 -Projections

The construction in Theorem 1.29 can be extended to the construction in the general Fréchet class case. Define for $J \subset \{1, \dots, k\}$, $T_J := \bigcap_{j \in J} T_j$, and for $f \in L^1(\mathbb{X}^n)$, f_T by integrating over the T components. f_T again is considered as a function on $[0, 1]^n$. Let for $T \in \mathcal{E}$, P_T have densities g^T , $P_T = g^T \mathbb{X}^{|T|}$ and define

$$h(x) = \sum_{m=1}^k (-1)^{m-1} \sum_{\substack{J \subset \{1, \dots, k\} \\ |J|=m}} g^{T_J}(x), \quad x \in [0, 1]^n, \quad (1.91)$$

where $g^{T_J}(x) = 0$ if T_J is empty. Then we have the following representation of the general Fréchet class (see Rü (1985)). Define the operator $T_{\mathcal{E}}$ by

$$T_{\mathcal{E}} f := h + V \circ S f. \quad (1.92)$$

Theorem 1.32 (Representation of general Fréchet class). *The class of all (signed) measures in the generalized Fréchet class which are Lebesgue-continuous has the representation*

$$\mathcal{M}_n^s(\mathcal{E}) = \{(T_{\mathcal{E}} f) \mathbb{X}^n; \quad f \in L^1(\mathbb{X}^n)\}$$

respectively

$$M_n(\mathcal{E}) = \{(T_{\mathcal{E}} f) \mathbb{X}^n; \quad f \in L^1(\mathbb{X}^n), T_{\mathcal{E}} f \geq 0\}. \quad (1.93)$$

Proof. In the first step we prove that $h \mathbb{X}^n \in \mathcal{M}_n^s(\mathcal{E})$ or, equivalently, that $h_{R_i} = g^{T_i}$, where $R_i = T_i^c$, $1 \leq i \leq k$. Without loss of generality we consider the case $i = 1$. By definition of h we get

$$\begin{aligned} h &= g^{T_1} + \sum_{m=1}^k (-1)^{m-1} \sum_{\substack{|J|=m, \\ J \neq \{1\}}} g^{T_J} \\ &= g^{T_1} + \sum_{m=1}^k (-1)^{m-1} \left(\sum_{\substack{|J|=m, \\ 1 \in J, J \neq \{1\}}} g^{T_{J \cup \{1\}}} + \sum_{\substack{|J|=m, \\ 1 \notin J}} g^{T_J} \right) \\ &= g^{T_1} + \sum_{m=1}^k (-1)^{m-1} \left(\sum_{\substack{|J|=m-1, \\ 1 \notin J, J \neq \emptyset}} g^{T_{J \cup \{1\}}} + \sum_{\substack{|J|=m, \\ 1 \notin J}} g^{T_J} \right) \\ &= g^{T_1} + \sum_{m=1}^k (-1)^m \left(\sum_{\substack{|J|=m, \\ 1 \notin J}} g^{T_{J \cup \{1\}}} - \sum_{\substack{|J|=m, \\ 1 \notin J}} g^{T_J} \right) \end{aligned}$$

and from the relation $(g^{T_{J \cup \{1\}}})_{R_1} = (g^{T_J})_{R_1}$ we obtain the assertion $h_{R_1^c} = g^{T_1}$.

Let now $P = g\lambda^n \in \mathcal{M}_n^s(\mathcal{E})$, then $g = h + (g - h) = h + V \circ S(g - h)$, since $g - h$ is by the first part of this proof a fixpoint of $V \circ S$. Conversely, for $f \in L^1(\lambda^n)$ and $T \in \mathcal{E}$

$$(h + V \circ Sf)_{T^c} = h_{T^c} = g^T$$

by definition of V , i.e. $(h + V \circ Sf)\lambda^n \in \mathcal{M}(\mathcal{E})$. \square

Theorem 1.32 allows in certain cases to construct families of probability measures with given multivariate marginals. The idea is to find a function $f \in L^1(\lambda^n)$, such that $V \circ Sf$ is balancing the negative parts of h . Some natural candidates for f are functions which allow an explicit and simple determination of the transform $V \circ Sf$, such as e.g. linear combinations of functions of the type $\prod_{i=1}^n v_i(x_i)$ where $\int v_i(x_i) dx_i = 0$, $1 \leq i \leq n$. The following is an example of this kind of construction.

Example 1.33. Let $n = 3$, $\mathcal{E} = \{\{1, 2\}, \{2, 3\}, \{1, 3\}\}$.

- (a) When the marginal densities are $f_{12}(x_1, x_2) = 1$, $f_{23}(x_1, x_3) = 1 + (x_2 - \frac{1}{2}) \cdot (x_3 - \frac{1}{2})$, $f_{13}(x_1, x_3) = 1 + (x_1 - \frac{1}{2}) (x_3 - \frac{1}{2})$, then

$$h(x_1, x_2, x_3) = \frac{3}{2} + x_1 x_2 + x_2 x_3 - x_3 - \frac{x_1 + x_2}{2}$$

is already a non-negative density with the given marginals.

- (b) If $f_{12}(x_1, x_2) = 1 + 3(x_1 - \frac{1}{2})(x_2 - \frac{1}{2})$, $f_{13}(x_1, x_3) = 1 - 3(x_1 - \frac{1}{2}) \cdot (x_3 - \frac{1}{2})$, and $f_{23}(x_2, x_3) = 1$, then

$$h(x_1, x_2, x_3) = f_{13}(x_1, x_3) + f_{12}(x_1, x_2) - 1$$

and $\min\{h(x_1, x_2, x_3)\} = \frac{1}{2} = h(1, 0, 1) = h(0, 1, 0)$. A function balancing these negative parts is given by

$$f(x_1, x_2, x_3) = -6(x_1 - \frac{1}{2})(x_2 - \frac{1}{2})(x_3 - \frac{1}{2}),$$

so that

$$\begin{aligned} h(x_1, x_2, x_3) + f(x_1, x_2, x_3) \\ = 1 - 6(x_1 - \frac{1}{2})(x_2 - \frac{1}{2})(x_3 - \frac{1}{2}) \\ + 3(x_1 - \frac{1}{2})(x_2 - \frac{1}{2}) - 3(x_1 - \frac{1}{2})(x_3 - \frac{1}{2}) \end{aligned}$$

gives a non-negative density with the given marginals as can easily be seen by discussing the cases $x_1, x_2 \leq \frac{1}{2}$, $x_3 \geq \frac{1}{2}$, etc. Instead of the factor 6 in the balancing function, one can use a factor a in an interval around 6, in this way obtaining a parametric class of distributions with given multivariate marginals.

\diamond

Mathematical Risk Analysis

Dependence, Risk Bounds, Optimal Allocations and
Portfolios

Rüschendorf, L.

2013, XII, 408 p., Hardcover

ISBN: 978-3-642-33589-1