The infinite extendibility problem for exchangeable real-valued random vectors

Jan-Frederik Mai

XAIA Investment GmbH
Sonnenstr. 19, 80331 München, Germany
e-mail: mai@tum.de

Abstract: We survey known solutions to the infinite extendibility problem for (necessarily exchangeable) probability laws on $\mathbb{R}^d$, which is:

Can a given random vector $\boldsymbol{X} = (X_1, \ldots, X_d)$ be represented in distribution as the first d members of an infinite exchangeable sequence of random variables?

This is the case if and only if $\boldsymbol{X}$ has a stochastic representation that is "conditionally iid" according to the seminal de Finetti's Theorem. Of particular interest are cases in which the original motivation behind the model $\boldsymbol{X}$ is not one of conditional independence. After an introduction and some general theory, the survey covers the traditional cases when $\boldsymbol{X}$ takes values in $\{0, 1\}^d$, has a spherical law, a law with ℓ_1-norm symmetric survival function, or a law with ℓ_∞-norm symmetric density. The solutions in all these cases constitute analytical characterizations of mixtures of iid sequences drawn from popular, one-parametric probability laws on $\mathbb{R}$, like the Bernoulli, the normal, the exponential, or the uniform distribution. The survey further covers the less traditional cases when $\boldsymbol{X}$ has a Marshall-Olkin distribution, a multivariate wide-sense geometric distribution, a multivariate extreme-value distribution, or is defined as a certain exogenous shock model including the special case when its components are samples from a Dirichlet prior. The solutions in these cases correspond to iid sequences drawn from random distribution functions defined in terms of popular families of non-decreasing stochastic processes, like a Lévy subordinator, a random walk, a process that is strongly infinitely divisible with respect to time, or an additive process. The survey finishes with a list of potentially interesting open problems. In comparison to former literature on the topic, this survey purposely dispenses with generalizations to the related and larger concept of finite exchangeability or to more general state spaces than $\mathbb{R}$. Instead, it aims to constitute an up-to-date comprehensive collection of known and compelling solutions of the real-valued extendibility problem, accessible for both applied and theoretical probabilists, presented in a lecture-like fashion.

MSC2020 subject classifications: Primary 60G09, 60E05; secondary 62H99.

Keywords and phrases: Exchangeability, conditionally iid, multivariate probability distributions.

Contents

1. Introduction and general background

1.1. General notation

Before we start, let us clarify some general notation used throughout, while some section-specific notations are introduced where they appear.

TABLE 1. *Summary of main results surveyed. Whereas most appearing notations are introduced in the main body of the article, here $\mathcal{L}[X]$ denotes the Laplace transform of a random variable $X \geq 0$, and $\langle \boldsymbol{x}, \boldsymbol{y} \rangle := \sum_{k=1}^{d} x_k \, y_k$.*

law of $\boldsymbol{X}$	conditionally iid with $H_t =$	analytically	see
arbitrary in $M_+^1(\mathbb{R}^d)$	arbitrary in $M_+^1(\mathfrak{H})$	$\sup_g \left\{ \frac{\|\mathbb{E}[g(\boldsymbol{X})]\|}{\sup_{\boldsymbol{Y}} \|\mathbb{E}[g(\boldsymbol{Y})]\|} \right\} \leq 1$, g bounded, Y_k iid	Theorem 1.2
arbitrary in $M_+^1(\{0,1\}^d)$	$(1-M)\,1_{\{t \geq 0\}} + M\,1_{\{t \geq 1\}}$	$\mathbb{P}(\boldsymbol{X} = \boldsymbol{x}) = \nabla^{d - \|\boldsymbol{x}\|_1} b_{\|\boldsymbol{x}\|_1}$, $b_k = \mathbb{E}[M^k]$	Theorem 2.1
spherical law	$\Phi\left(\frac{t}{M}\right)$	$\mathbb{E}[\exp\{\mathrm{i}\,\langle \boldsymbol{x}, \boldsymbol{X} \rangle\}] = \varphi(\|\boldsymbol{x}\|_2)$, $\varphi = \mathcal{L}[M]$	Theorem 3.1
ℓ_∞-norm symmetric density	$\max\left\{0, \min\left\{\frac{t}{M}, 1\right\}\right\}$	$f_{\boldsymbol{X}}(\boldsymbol{x}) = g_d(x_{[d]})$, $g_d(x) = \mathbb{E}[M^{-d}\,1_{\{M > x\}}]$	Theorem 3.3
ranks of Dirichlet prior samples	$DP(c, \{t\}_{t \in [0,1]})$	$\mathbb{P}(\boldsymbol{X} \leq \boldsymbol{x}) = x_{[1]} \prod_{k=2}^{d} \frac{c\,x_{[k]} + k - 1}{c + k - 1}$	Lemma 6.1
law of $\boldsymbol{X}$	conditionally iid with $Z_t =$	$\mathbb{P}(\boldsymbol{X} > \boldsymbol{x}) =$	see
ℓ_1-norm symmetric survival function	$M\,t$	$\varphi(\|\boldsymbol{x}\|_2)$, $\varphi = \mathcal{L}[M]$	Theorem 3.2
Marshall–Olkin exponential law	Lévy subordinator	$\exp\left\{ \sum_{k=1}^{d} \Psi(k)\,\nabla x_{[d-k]} \right\}$, $e^{-\Psi} = \mathcal{L}[Z_1]$	Theorem 4.1($\mathcal{E}$)
multivariate wide-sense geometric distribution	$\sum_{n=1}^{\lfloor t \rfloor} Y_n$, Y_k iid	$\prod_{k=1}^{d} \mathbb{E}[\exp(-k\,Y_1)]^{x_{[d-k+1]} - x_{[d-k]}}$	Theorem 4.1($\mathcal{G}$)
min-stable multivariate exponential law	$b\,t - c\log\left\{ \prod_{n \geq 1} G_n^{(n)} \right\}$, $G^{(n)}$ iid $\sim \gamma$	$\exp\left\{ -b \sum_{k=1}^{d} x_k - c \int_{\mathfrak{H}_{+,1}} \ell_G(\boldsymbol{x})\,\gamma(\mathrm{d}G) \right\}$	Theorem 5.1
exogenous shock model	additive subordinator	$\frac{\exp\left\{ -\sum_{k=1}^{d} \Psi_{x_{[d-k+1]}}(k) \right\}}{\exp\left\{ -\Psi_{x_{[d-k+1]}}(k-1) \right\}}$, $e^{-\Psi_t} = \mathcal{L}[Z_t]$	Theorem 6.1
Sato-frailty model	self-similar, additive subordinator	$\frac{\exp\left\{ -\sum_{k=1}^{d} \Psi(k\,x_{[d-k+1]}) \right\}}{\exp\left\{ -\Psi((k-1)\,x_{[d-k+1]}) \right\}}$, $e^{-\Psi} = \mathcal{L}[Z_1]$	Lemma 6.2

Some very general mathematical definitions:

We denote by $\mathbb{N} = \{1, 2, 3, \ldots\}$ the set of natural numbers and $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$, by $\mathbb{R}$ the set of real numbers, by $\mathbb{R}^d$ the set of d-dimensional row vectors with entries in $\mathbb{R}$, for $d \in \mathbb{N}$. For $n \in \mathbb{N}_0$ we denote by $f^{(n)}$ the n-th derivative of a function $f : \mathbb{R} \to \mathbb{R}$, provided existence. For d numbers $x_1, \ldots, x_d \in \mathbb{R}$ we denote by $x_{[1]} \leq x_{[2]} \leq \ldots \leq x_{[d]}$ an ordered list. For $x \in \mathbb{R}$ we denote by $\lceil x \rceil$ the smallest integer greater or equal to x (ceiling function), and by $\lfloor x \rfloor$ the largest integer less than or equal to x (floor function). We denote by $\det[A]$ the determinant of a square matrix $A \in \mathbb{R}^{d \times d}$. We denote elements $\boldsymbol{x} = (x_1, \ldots, x_d) \in \mathbb{R}^d$ by bold letters in comparison to (one-dimensional) elements $x \in \mathbb{R}$. Expressions like $\boldsymbol{x} > \boldsymbol{y}$ for $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$ are meant component-wise, i.e. $x_k > y_k$ for each $k = 1, \ldots, d$. We furthermore use the notation $\|\boldsymbol{x}\|_p := (|x_1|^p + \ldots + |x_d|^p)^{1/p}$ for the ℓ_p-norm of $\boldsymbol{x} \in \mathbb{R}^d$, $p \geq 1$.

Some general definitions regarding probability spaces:

All random objects to be introduced are formally defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with σ-algebra $\mathcal{F}$ and probability measure $\mathbb{P}$, and the expected value of a random variable X is denoted by $\mathbb{E}[X]$. As usual, the argument $\omega \in \Omega$ of some random variable $X : \Omega \to \mathbb{R}$ will always be omitted. The symbol $\stackrel{d}{=}$ denotes equality in distribution and the symbol $\sim$ means "is distributed according to". We recall that for random vectors $(X_1, \ldots, X_d) \stackrel{d}{=} (Y_1, \ldots, Y_d)$ means $\mathbb{E}[g(X_1, \ldots, X_d)] = \mathbb{E}[g(Y_1, \ldots, Y_d)]$ for all bounded, continuous functions $g : \mathbb{R}^d \to \mathbb{R}$, where the expectation values $\mathbb{E}$ are taken on the respective probability spaces of $(X_1, \ldots, X_d)$ and $(Y_1, \ldots, Y_d)$, which might be different. Equality in law for two stochastic processes $X = \{X_t\}$ and $Y = \{Y_t\}$ means that $(X_{t_1}, \ldots, X_{t_d}) \stackrel{d}{=} (Y_{t_1}, \ldots, Y_{t_d})$ for arbitrary $d \in \mathbb{N}$ and $t_1, t_2, \ldots, t_d$. Throughout, the abbreviation *iid* stands for independent and identically distributed. The index t of a real-valued random variable f_t that belongs to some stochastic process $f = \{f_t\}_{t \in T}$ is purposely written as a sub-index, in order to distinguish it from the value $f(t)$ of some (non-random) function $f : T \to \mathbb{R}$. If F is the distribution function of some random variable taking values in $\mathbb{R}$, we denote by

$$F^{-1}(y) := \inf\{x \in \mathbb{R} \,:\, F(x) \geq y\}, \quad y \in [0, 1],$$

its *generalized inverse*, see [25] for background. Any distribution function C of a random vector $\boldsymbol{U} = (U_1, \ldots, U_d)$ whose components U_k are uniformly distributed on $[0, 1]$ is called a *copula*, see [75] for a textbook treatment. We further recall that an arbitrary survival function $\bar{F}$ of some d-variate random vector $\boldsymbol{X}$ can always be written[1] as $\bar{F}(\boldsymbol{x}) = \hat{C}\big(\mathbb{P}(X_1 > x_1), \ldots, \mathbb{P}(X_d > x_d)\big)$, where $\hat{C}$ is a copula, called a *survival copula* for $\bar{F}$, and it is uniquely determined in case the random variables $X_1, \ldots, X_d$ have continuous distribution functions. This is the survival analogue of the so-called Theorem of Sklar, due to [101]. The Theorem of Sklar itself states that the distribution function F of $\boldsymbol{X}$ can be written as $F(\boldsymbol{x}) = C\big(\mathbb{P}(X_1 \leq x_1), \ldots, \mathbb{P}(X_d \leq x_d)\big)$ for a copula C, called a

---

[1]See [81, p. 195–196].

copula for F. The relationship between a copula C and its survival copula $\hat{C}$ is that if $(U_1, \ldots, U_d) \sim C$ then $(1 - U_1, \ldots, 1 - U_d) \sim \hat{C}$.

A notation of specific interest in the present survey:

We denote by $\mathfrak{H}$ the set of all distribution functions of real-valued random variables, and by $\mathfrak{H}_+$ the subset containing all elements F such that $x < 0$ implies $F(x) = 0$, i.e. distribution functions of non-negative random variables. Elements $F \in \mathfrak{H}$ are right-continuous, and we denote by $F(x-) := \lim_{t \uparrow x} F(x)$ their left-continuous versions. If $\mathfrak{X}$ is some Hausdorff space, we denote by $M_+^1(\mathfrak{X})$ the set of all probability measures on the measurable space $(\mathfrak{X}, \mathcal{B}(\mathfrak{X}))$, where $\mathcal{B}(\mathfrak{X})$ denotes the Borel-σ-algebra of $\mathfrak{X}$. This notation is borrowed from [8]. Now recall that $\mathfrak{H}$ is metrizable (hence in particular Hausdorff) when topologized with the so-called *Lévy metric* that induces weak convergence of the associated probability distributions on $\mathbb{R}$, see [100]. Consequently, we denote by $M_+^1(\mathfrak{H})$ the set of all probability measures on $\mathfrak{H}$. A random element $H = \{H_t\}_{t \in \mathbb{R}} \sim \gamma \in M_+^1(\mathfrak{H})$ is almost surely a càdlàg stochastic process, and a common way of treating probability laws of such objects works via the so-called *Skorohod metric* on the space of càdlàg paths. However, even though the Skorohod topology and the topology induced by the Lévy metric are not identical, see [88, p. 327-328], their induced Borel-σ-algebras on the set $\mathfrak{H}$ can indeed be shown to coincide, so that our viewpoint is equivalent.

1.2. Motivation and mathematical preliminaries

Throughout, we consider by $\boldsymbol{X} = (X_1, \ldots, X_d)$ a random vector taking values in $\mathbb{R}^d$. Since we are only interested in the probability distribution of $\boldsymbol{X}$, we identify $\boldsymbol{X}$ with its probability law in the sense that we often say $\boldsymbol{X}$ has some property if and only if its probability distribution has this property. The central theme of the present survey deals with the following formal definition, using the nomenclature in [24].

Definition 1.1 (Conditionally iid). *We say that a probability measure $\mu \in M_+^1(\mathbb{R}^d)$ is* conditionally iid *if there exists a probability measure $\gamma \in M_+^1(\mathfrak{H})$ such that the equality*

$$\mu\big((-\infty, x_1] \times \ldots \times (-\infty, x_d]\big) = \int_{\mathfrak{H}} h(x_1) \cdots h(x_d) \, \gamma(\mathrm{d}h)$$

holds for all $x_1, \ldots, x_d \in \mathbb{R}$.

We say that a random vector $\boldsymbol{X} = (X_1, \ldots, X_d)$ taking values in $\mathbb{R}^d$ is conditionally iid *if its probability distribution is conditionally iid. This is equivalent to existence of a random element $H \in \mathfrak{H}$ such that almost surely*

$$\mathbb{P}(X_1 \leq x_1, \ldots, X_d \leq x_d \,|\, H) = H_{x_1} \cdots H_{x_d}, \quad x_1, \ldots, x_d \in \mathbb{R},$$

where conditioning on H means conditioning on $\sigma(H)$. Given a family of probability distributions $\mathfrak{M} \subset M_+^1(\mathbb{R}^d)$, we introduce the notation

$$\mathfrak{M}_* = \{\mu \in \mathfrak{M} \,:\, \mu \text{ is conditionally iid}\}.$$

Consider a random vector $\boldsymbol{X} = (X_1, \ldots, X_d)$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ that is defined via

$$X_k := f(U_k, H), \quad k = 1, \ldots, d,$$

with some[2] measurable "functional" f, an iid sequence of random "objects" $U_1, \ldots, U_d$, and some independent random object H that is measurable with respect to the sub-σ-algebra $\mathcal{H} = \sigma(H) \subset \mathcal{F}$ that it generates itself. Such $\boldsymbol{X}$ is always conditionally iid and the probability distribution of the stochastic process $H_t := \mathbb{P}(X_1 \leq t \,|\, \mathcal{H})$, $t \in \mathbb{R}$, plays the role of the probability measure γ in Definition 1.1. The object H, sometimes called a *latent (dependence-inducing) factor*, then induces dependence between the components, which are iid conditioned on $\mathcal{H}$. This is a Bayesian viewpoint, based on a two-step construction: first simulate an instance of H, then simulate $X_1, \ldots, X_d$ iid given H. However, it is important to be aware that a random vector that is conditionally iid according to our definition does not necessarily have to be defined by a stochastic model that relies on such a two-step construction. Our definition only requires that such a construction exists, possibly on another probability space. In fact, typical cases of interest are such that μ is defined in terms of a stochastic model or probabilistic property which is a priori unrelated to the concept of conditional independence, as we will see.

Throughout, we are interested in a solution to the following problem:

Problem 1.1 (Motivating problem). *Given a collection $\mathfrak{M} \subset M_+^1(\mathbb{R}^d)$ and $\mu \in \mathfrak{M}$, provide necessary and sufficient conditions ensuring that $\mu \in \mathfrak{M}_*$.*

Remark 1.1 (Nomenclature). *In the literature, elements of $\mathfrak{M}_*$ are not always called conditionally iid, but other names have been given. For instance, [98] calls them* positive dependent by mixture *(PDM), [103, 40, 60] call them* infinitely extendible, *and [75, Definition 1.10, p. 43] call them simply* extendible. *The nomenclature "(infinite) extendibility" refers to the fact that conditionally iid random vectors can always be thought of as finite margins of infinite conditionally iid sequences, as will be explained below. The nomenclature "PDM" becomes intuitive from Lemmata 1.2, 1.3 and 1.4 below, but is rather unusual.*

The investigation of conditionally iid random vectors is closely related to the concept of exchangeability. Recall that the probability distribution of a random vector $\boldsymbol{X}$ is called *exchangeable* if it is invariant under an arbitrary permutation of the components of $\boldsymbol{X}$. The following observation is immediate but important.

Lemma 1.1 (Exchangeability). *If $\boldsymbol{X}$ is conditionally iid, it is also exchangeable.*

Proof. Let $\boldsymbol{X} \sim \mu$ with γ as in Definition 1.1. If π is an arbitrary permutation of $\{1, \ldots, d\}$ we observe that

$$\mathbb{P}(X_1 \leq x_1, \ldots, X_1 \leq x_d) = \int_{\mathfrak{H}} h(x_1) \, \cdots \, h(x_d) \, \gamma(\mathrm{d}h)$$

---

[2]As will later be explained in more detail, we can without loss of generality choose $f(., h) = h^{-1}(.)$, $h \in \mathfrak{H}$, so that $f(\cdot, h)$ is an indexed family of quantile functions.

$$= \int_{\mathfrak{H}} h(x_{\pi(1)}) \cdots h(x_{\pi(d)}) \, \gamma(\mathrm{d}h) = \mathbb{P}(X_{\pi^{-1}(1)} \leq x_1, \ldots, X_{\pi^{-1}(d)} \leq x_d).$$

Since π was arbitrary, this shows that the distribution function (hence law) of $\boldsymbol{X}$ is invariant with respect to permutations of its components. $\square$

Exchangeability is a property which is convenient to investigate by means of Analysis, whereas the notion "conditionally iid", in which we are interested, is a priori purely probabilistic and more difficult to investigate. Unfortunately, exchangeability is only a necessary but no sufficient condition for the solution of our problem. For instance, a bivariate normal distribution is obviously exchangeable if and only if the two means and variances are identical, also for negative correlation coefficients. However, Example 1.1 and Lemma 1.2 below show that conditionally iid random vectors necessarily have non-negative correlation coefficients. One can show in general that the correlation coefficient, if existent, between two components of an exchangeable random vector on $\mathbb{R}^d$ is bounded from below by $-1/(d-1)$, see, e.g., [1, p. 7]. As the dimension d tends to infinity, this lower bound becomes zero. Even better, the difference between exchangeabilty and a conditionally iid structure vanishes completely as the dimension d tends to infinity, which is the content of de Finetti's Theorem.

Theorem 1.1 (de Finetti's Theorem). *Let $\{X_k\}_{k \in \mathbb{N}}$ be an infinite sequence of random variables on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The sequence $\{X_k\}_{k \in \mathbb{N}}$ is* exchangeable, *meaning that each finite subvector is exchangeable, if and only if it is iid conditioned on some σ-field $\mathcal{H} \subset \mathcal{F}$. In this case, $\mathcal{H}$ equals almost surely the tail-σ-field of $\{X_k\}_{k \in \mathbb{N}}$, which is given by $\cap_{n \geq 1} \sigma(X_n, X_{n+1}, \ldots)$.*

Proof. Originally due to [19]. We refer to [1] for a proof based on the reversed martingale convergence theorem, which is briefly sketched. Of course, we only need to verify that exchangeability implies conditionally iid, as the converse follows from Lemma 1.1. For the sake of a more convenient notation we assume the infinite sequence $\{X_k\}_{k \in \mathbb{N}_0}$ is indexed by $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$, and we define σ-algebras $\mathcal{F}_{-n} := \sigma(X_n, X_{n+1}, \ldots)$ for $n \in \mathbb{N}_0$. The tail-σ-filed of the sequence is $\mathcal{H} := \cap_{n \leq 0} \mathcal{F}_n$. In order to establish the claim, three auxiliary observations are helpful with an arbitrary bounded, measurable function g fixed:

(i) Exchangeability implies $(X_0, X_1, \ldots) \stackrel{d}{=} (X_0, X_{n+1}, \ldots)$ for arbitrary $n \in \mathbb{N}$. This implies $\mathbb{E}[g(X_0) \,|\, \mathcal{F}_{-1}] \stackrel{d}{=} \mathbb{E}[g(X_0) \,|\, \mathcal{F}_{-(n+1)}]$, $n \in \mathbb{N}$.

(ii) The sequence $Y_n := \mathbb{E}[g(X_0) \,|\, \mathcal{F}_n]$, $n \leq 0$, is easily checked to be a reversed martingale. The reversed martingale convergence theorem implies that Y_n converges almost surely and in L^1 to $\mathbb{E}[g(X_0) \,|\, \mathcal{H}]$. See [23, p. 264 ff] for background on reversed martingales (convergence).

(iii) Letting $n \to \infty$ in (i), we observe from (ii) that $Y_{-1} \stackrel{d}{=} \mathbb{E}[g(X_0) \,|\, \mathcal{H}]$. We can further replace this equality in law by an almost sure equality, since $\mathcal{H} \subset \mathcal{F}_{-1}$ and the second moments of Y_{-1} and $\mathbb{E}[g(X_0) \,|\, \mathcal{H}]$ coincide. Thus, the sequence $\{Y_{-n}\}_{n \in \mathbb{N}}$ is almost surely a constant sequence.

With these auxiliary observations we may now finish the argument. On the one

hand, exchangeability implies $(X_0, X_{n+1}, \ldots) \stackrel{d}{=} (X_n, X_{n+1}, \ldots)$, which gives the almost sure equality $\mathbb{E}[g(X_0) \,|\, \mathcal{F}_{-(n+1)}] = \mathbb{E}[g(X_n) \,|\, \mathcal{F}_{-(n+1)}]$. Taking $\mathbb{E}[. \,|\, \mathcal{H}]$ on both sides of this equation implies with the tower property of conditional expectation that $\mathbb{E}[g(X_0) \,|\, \mathcal{H}] = \mathbb{E}[g(X_n) \,|\, \mathcal{H}]$. Since g was arbitrary, X_0 and X_n are identically distributed conditioned on $\mathcal{H}$, and since n was arbitrary all members of the sequence are identically distributed conditioned on $\mathcal{H}$. To verify conditional independence, let g_1, g_2 be two bounded, measurable functions. For $n \geq 1$ arbitrary, using (iii) in the third equality below, we compute

$$
\begin{aligned}
\mathbb{E}[g_1(X_0)\, g_2(X_n) \,|\, \mathcal{H}] &= \mathbb{E}[\mathbb{E}[g_1(X_0)\, g_2(X_n) \,|\, \mathcal{F}_{-n}] \,|\, \mathcal{H}] \\
&= \mathbb{E}[g_2(X_n)\, \mathbb{E}[g_1(X_0) \,|\, \mathcal{F}_{-n}] \,|\, \mathcal{H}] \\
&= \mathbb{E}[g_2(X_n)\, \mathbb{E}[g_1(X_0) \,|\, \mathcal{H}] \,|\, \mathcal{H}] \\
&= \mathbb{E}[g_2(X_n) \,|\, \mathcal{H}]\, \mathbb{E}[g_1(X_0) \,|\, \mathcal{H}].
\end{aligned}
$$

The precisely same tower property argument inductively also implies

$$
\mathbb{E}\Big[\prod_{j=1}^{k} g_i(X_{i_j}) \,\Big|\, \mathcal{H} \Big] = \prod_{j=1}^{k} \mathbb{E}[g_i(X_{i_j}) \,|\, \mathcal{H}]
$$

for arbitrary $0 \leq i_1 < \ldots < i_k$ and bounded measurable functions $g_1, \ldots, g_k$. Thus, the random variables $X_0, X_1, \ldots$ are independent conditioned on $\mathcal{H}$. □

Which topics are covered in the present survey?
The present article surveys known answers to Problem 1.1 for families of multivariate probability distributions $\mathfrak{M}$ that are well known in the statistical literature, and/ or have proven useful as a mathematical model for specific applications. While several traditional results of the theory have been studied in the last century, some significant achievements have been accomplished only within the last decade, so the present author feels that this is a good time point to recap what has been achieved, hence to write this overview article. One goal of the present survey is to collect the numerous results under one common umbrella in a reader-friendly summary to make them accessible for a broader audience of applied and theoretical probabilists, and in order to inspire others to join this interesting strand of research in the future. Proofs, or at least proof sketches, are presented for most results in order to (a) demonstrate how solutions to Problem 1.1 often unravel surprising links between seemingly different fields of mathematics/probability theory, and (b) render this document a useful basis for the use as lecture notes in an advanced course on multivariate statistics or probability theory.

Which topics are not covered in the present survey?
The scope of former literature on the topic is often wider, in particular the references [1, 57, 2] are very popular surveys on the topic with wider scope. On the one hand, many references on the topic discuss conditionally iid models under the umbrella of exchangeability, which has been mentioned to be a weaker notion for finite random vectors. The characterization of the (finitely) exchangeable subfamily of $\mathfrak{M}$ is often easier than the characterization of the

(in general) smaller set $\mathfrak{M}_*$, and is typically an important first step towards a solution to Problem 1.1. However, the second (typically harder) step from (finite) exchangeability to conditionally iid is usually the more important and more interesting step from both a theoretical and practical perspective. The algebraic structure of a general theory on (finite) exchangeability is naturally of a different, often more combinatorial character, whereas "conditionally iid" by virtue of de Finetti's Theorem naturally is the concept of an infinite limit (of exchangeability) so that techniques from Analysis enter the scene. Thus, we feel it is useful to provide an account with a more narrow scope on conditionally iid, even though for some of the presented examples we are well aware that an interesting (finite) exchangeable theory is also viable. On the other hand, many references consider the case when the components of $\boldsymbol{X}$ take values in more general spaces than $\mathbb{R}$, for instance in $\mathbb{R}^n$ (i.e. lattices instead of vectors) or even function spaces. In particular, de Finetti's Theorem 1.1 can be generalized in this regard, seminal references are [30, 92]. Research in this direction is by nature more abstract and thus maybe less accessible for a broader audience, or for more practically oriented readers. One goal of the present survey is to provide an account that is not exclusively geared towards theorists but also to applicants of the theory, and in particular to point out relationships to classical statistical probability laws on $\mathbb{R}^d$. We believe that a limitation of this survey's scope to the real-valued case is still rich enough to provide a solid basis for an interesting and accessible theory. In fact, we seek to demonstrate that Problem 1.1 has been solved satisfactorily in quite a number of highly interesting cases, and the solutions contain interesting links to different probabilistic topics. Of course, it might be worthwhile to ponder about generalizations of some of the presented results to more abstract settings in the future (unless already done) – but purposely these lie outside the present survey.

Why is Problem 1.1 interesting at all?

Broadly speaking, because of two reasons: (a) conditionally iid models are convenient for applications, and (b) solutions to the extendibility problem sometimes rely on compelling relationships between a priori different theories.

(a) Roughly speaking, conditionally iid models allow to model (strong and weak) dependence between random variables in a way that features many desirable properties which are taylor-made for applications, in particular when the dimension d is large. Firstly, a conditionally iid random vector is "dimension-free" in the sense that components can be added or removed from $\boldsymbol{X}$ without altering the basic structure of the model, which simply follows from the fact that an iid sequence remains an iid sequence after addition or removal of certain members. This may be very important in applications that require a regular change of dimension, e.g. the readjustment of a large credit portfolio in a bank, when old credits leave and new credits enter the portfolio frequently. Secondly, if $\boldsymbol{X}$ has a distribution from a parametric family, the parameters of this family are typically determined by the parameters of the underlying latent probability measure γ, irrespective of the dimension d. Consequently, the number of param-

eters does usually not grow significantly with the dimension d and may be controlled at one's personal taste. This is an enormous advantage for model design in practice, in particular since the huge degree of freedom/ huge number of parameters in a high-dimensional dependence model is often more boon than bane. Thirdly, fundamental statistical theorems relying on the iid assumption, like the law of large numbers, may still be applied in a conditionally iid setting, making such models very tractable. Last but not least, in dependence modeling a "factor-model way of thinking" is very intuitive, e.g. it is well-established in the multivariate normal case (thinking of principle component analyses etc.). On a high level, if one wishes to design a multi-factor dependence model within a certain family of distributions $\mathfrak{M}$, an important first step is to determine the one-factor subfamily $\mathfrak{M}_*$. Having found a conditionally iid stochastic representation of $\mathfrak{M}_*$, the design of multi-factor models is sometimes obvious from there, see also paragraph 7.3.

(b) The solution to Problem 1.1 is often mathematically challenging and compelling. It naturally provides an interesting connection between the "static" world of random vectors and the "dynamic" cosmos of (one-dimensional) stochastic processes. The latter enter the scene because the latent factor being responsible for the dependence in a conditionally iid model for $\boldsymbol{X}$ may canonically be viewed as a non-decreasing stochastic process (a random distribution function), which is further explained in Section 1.3 below. In particular, for some classical families $\mathfrak{M}$ of multivariate laws from the statistical literature the family $\mathfrak{M}_*$ in Problem 1.1 is conveniently described in terms of a well-studied family of stochastic processes like Lévy subordinators, Sato subordinators, or processes which are infinitely divisible with respect to time. Moreover, in order to formally establish the aforementioned link between these two seemingly different fields of research the required mathematical techniques involve classical theories from Analysis like Laplace transforms, Bernstein functions, and moment problems.

Before we start, can we please study a first simple example?

It is educational to end this motivating paragraph by demonstrating the motivating problem with a simple example that all readers are familiar with. Denoting by $\mathcal{N}(\boldsymbol{\mu}, \Sigma)$ the multivariate normal law with mean vector $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_d) \in \mathbb{R}^d$ and covariance matrix $\Sigma \in \mathbb{R}^{d \times d}$, Example 1.1 provides the solution for Problem 1.1 in the case when $\mathfrak{M}$ consists of all multivariate normal laws.

Example 1.1 (The multivariate normal law)**.** *We want to solve Problem 1.1 for the family*

$$\mathfrak{M} = \{\mathcal{N}(\boldsymbol{\mu}, \Sigma) \, : \, \boldsymbol{\mu} \in \mathbb{R}^d, \, \Sigma \in \mathbb{R}^{d \times d} \text{ symmetric, positive definite}\}.$$

We claim that $\mathfrak{M}_*$ equals the set of all multivariate normal distributions satis-

fying

$$\boldsymbol{\mu} = (\mu, \ldots, \mu), \quad \Sigma = \begin{bmatrix} \sigma^2 & \rho\,\sigma^2 & \cdots & \rho\,\sigma^2 \\ \rho\,\sigma^2 & \sigma^2 & & \rho\,\sigma^2 \\ \vdots & & \ddots & \\ \rho\,\sigma^2 & \rho\,\sigma^2 & & \sigma^2 \end{bmatrix}, \quad \mu \in \mathbb{R},\, \sigma > 0,\, \rho \in [0,1].$$

Proof. Consider $\boldsymbol{X} = (X_1, \ldots, X_d) \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ for $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_d) \in \mathbb{R}^d$, and $\Sigma = (\Sigma_{i,j}) \in \mathbb{R}^{d \times d}$ a positive definite matrix. If we assume that the law of $\boldsymbol{X}$ is in $\mathfrak{M}_*$, it follows that there is a sub-σ-algebra $\mathcal{H} \subset \mathcal{F}$ such that the components $X_1, \ldots, X_d$ are iid conditioned on $\mathcal{H}$. Consequently,

$$\mu_k = \mathbb{E}[X_k] = \mathbb{E}[\mathbb{E}[X_k \,|\, \mathcal{H}]] = \mathbb{E}[\mathbb{E}[X_1 \,|\, \mathcal{H}]] = \mathbb{E}[X_1] = \mu_1, \qquad (1.1)$$

irrespectively of $k = 1, \ldots, d$. The analogous reasoning also holds for the second moment of X_k, which implies $\Sigma_{k,k} = \Sigma_{1,1}$ for all k. Moreover,

$$\mathbb{E}[X_i \, X_j] = \mathbb{E}[\mathbb{E}[X_i \,|\, \mathcal{H}]\, \mathbb{E}[X_j \,|\, \mathcal{H}]] = \mathbb{E}[\mathbb{E}[X_1 \,|\, \mathcal{H}]^2] \geq \mu_1^2, \qquad (1.2)$$

for arbitrary components $i \neq j$, where we used the conditional iid structure and Jensen's inequality. This finally implies that all off-diagonal elements of Σ are identical and non-negative.

Conversely, let $\mu \in \mathbb{R}$, $\sigma > 0$, and $\rho \in [0,1]$. Consider a probability space on which $d+1$ iid standard normally distributed random variables $M, M_1, \ldots, M_d$ are defined. We define

$$X_k := \mu + \sigma \left(\sqrt{\rho}\, M + \sqrt{1-\rho}\, M_k \right), \quad k = 1, \ldots, d. \qquad (1.3)$$

It is readily observed that $\boldsymbol{X} = (X_1, \ldots, X_d)$ has a multivariate normal law with pairwise correlation coefficients all being equal to ρ, and all components having mean μ and variance σ^2. Notice in particular that the non-negativity of ρ is important in the construction (1.3) because the square root is not well-defined otherwise. The components of $\boldsymbol{X}$ are obviously conditionally iid given the σ-algebra $\mathcal{H}$ generated by M. Hence the law of $\boldsymbol{X}$ is in $\mathfrak{M}_*$. $\qquad \square$

There are already some interesting remarks to be made about this simple example. First of all, it is observed that the family $\mathfrak{M}_*$ is always three-parametric, irrespective of the dimension d. This stands in glaring contrast to $\mathfrak{M}$, which has $d + d\,(d+1)/2$ parameters in dimension d. Second, in general a Monte Carlo simulation of a d-dimensional normal random vector $\boldsymbol{X}$ requires a Cholesky decomposition of the matrix Σ, which typically has computational complexity of order d^3, see [75, Algorithm 4.3, p. 182]. In contrast, the simulation of $\boldsymbol{X}$ with law in $\mathfrak{M}_*$ according to (1.3) has only linear complexity in the dimension d. Especially in large dimensions this can be a critical improvement of computational speed. Third, the proof above shows that each random vector with law in $\mathfrak{M}_*$ may actually be viewed as the first d components of an infinite sequence of conditionally iid random variables such that arbitrary finite n-margins have a multivariate normal law. Thus, we have actually solved the re-fined Problem 1.3 to be introduced in the upcoming paragraph.

1.3. Canonical probability spaces

We have mentioned earlier that a conditionally iid random vector $\boldsymbol{X}$ is usually constructed as $X_k := f(U_k, H)$, $k = 1, \ldots, d$, from an iid sequence $U_1, \ldots, U_d$, some independent stochastic object H, and some functional f. Clearly, this general model is inconvenient because neither the law of U_1, nor the nature of the stochastic object H or the functional f are given explicitly. However, there is a canonical choice for all three entities, which we are going to consider in the sequel. By definition, conditionally iid means that conditioned on the object H the random variables $X_1, \ldots, X_d$ are iid, distributed according to a univariate distribution function F, which may depend on H. A univariate distribution function F is nothing but a non-decreasing, right-continuous function $F : \mathbb{R} \to [0, 1]$ with $\lim_{t \to -\infty} F(t) = 0$ and $\lim_{t \to \infty} F(t) = 1$, see [12, Theorem 12.4, p. 176]. Without loss of generality we may assume that the random object $H = \{H_t\}_{t \in \mathbb{R}}$ already is the conditional distribution function itself, i.e. is a random variable in the space of distribution functions – or, in other words, a non-decreasing, right-continuous stochastic process with $\lim_{t \to -\infty} H_t = 0$ and $\lim_{t \to \infty} H_t = 1$. In other words, $H \sim \gamma \in M_+^1(\mathfrak{H})$. In this case, a canonical choice for the law of U_1 is the uniform distribution on $[0, 1]$ and the functional f may be chosen as

$$X_k = f(U_k, H) := \inf\{t \in \mathbb{R} \,:\, H_t > U_k\} = H_{U_k}^{-1}, \quad k = 1, \ldots, d. \tag{1.4}$$

Recall here that H is interpreted as a random distribution function and H^{-1} denotes its generalized inverse. In particular, $X_k \leq x$ if and only if $U_k \leq H_x$. Indeed, one verifies that $X_1, \ldots, X_d$ are iid conditioned on $\mathcal{H} := \sigma(\{H_t\}_{t \in \mathbb{R}})$, with common univariate distribution function H, since

$$\mathbb{P}(X_1 \leq t_1, \ldots, X_d \leq t_d \,|\, \mathcal{H}) = \mathbb{P}(U_1 \leq H_{t_1}, \ldots, U_d \leq H_{t_d} \,|\, \mathcal{H})$$
$$= H_{t_1} H_{t_2} \cdots H_{t_d},$$

for all $t_1, \ldots, t_d \in \mathbb{R}$. Every random vector which is conditionally iid can be constructed like this, i.e. there is a one-to-one relation between such models and random variables in the space of (one-dimensional) distribution functions, as already adumbrated in Definition 1.1. For each given $H = \{H_t\}_{t \in \mathbb{R}} \sim \gamma \in M_+^1(\mathfrak{H})$, and a given dimension $d \in \mathbb{N}$, the canonical construction (1.4) induces a multivariate probability distribution on $\mathbb{R}^d$, and we denote this mapping from $M_+^1(\mathfrak{H})$ to a subset of $M_+^1(\mathbb{R}^d)$ by Θ_d throughout, this subset being precisely the conditionally iid probability laws on $\mathbb{R}^d$. It is implicit that $\Theta_d(\gamma)$ depends on the law of $H \sim \gamma$ via

$$\Theta_d(\gamma)\big((-\infty, t_1] \times \ldots \times (-\infty, t_d]\big) = \mathbb{E}[H_{t_1} \ldots H_{t_d}].$$

Given $\mathfrak{M} \subset M_+^1(\mathbb{R}^d)$ we denote the pre-image of $\mathfrak{M}_*$ under Θ_d in $M_+^1(\mathfrak{H})$ by $\Theta_d^{-1}(\mathfrak{M}_*)$. In words, it equals the subset of $M_+^1(\mathfrak{H})$ which consists of all probability laws γ of stochastic processes $\{H_t\}_{t \in \mathbb{R}}$ such that $\boldsymbol{X}$ of the canonical construction (1.4) has a law in $\mathfrak{M}$, hence in $\mathfrak{M}_*$. From this equivalent viewpoint our motivating Problem 1.1 becomes

Problem 1.2 (Motivating problem reformulated)**.** *For a given family of d-dimensional probability distributions $\mathfrak{M}$, determine the family of stochastic processes $\Theta_d^{-1}(\mathfrak{M}_*) \subset M_+^1(\mathfrak{H})$.*

Admittedly, this reformulation in terms of the stochastic process H might appear quite artificial at this point, but we will see later that in some cases we obtain interesting correspondences between classical probability distributions on $\mathbb{R}^d$ and families of stochastic processes. On a high level, the problem of determining the intersection of a given family $\mathfrak{M}$ of distributions with the family of conditionally iid distributions may also be re-phrased as the problem of finding an increasing stochastic process whose stochastic nature induces the given multivariate distribution when inserted into a canonical stochastic model.
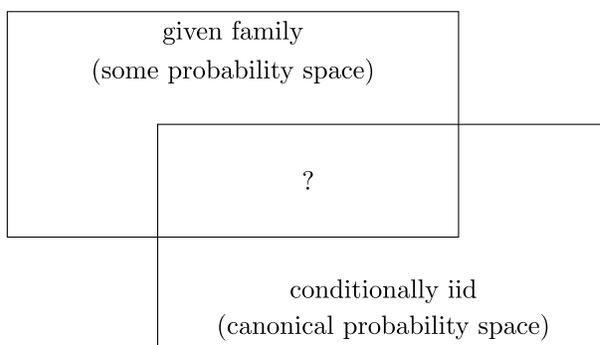
Obviously, in the stochastic model (1.4) it is possible to let d tend to infinity, since the U_k are iid. Thus, we may without loss of generality think of a conditionally iid random vector $\boldsymbol{X}$ as the first d members of an infinite sequence $\{X_k\}_{k\in\mathbb{N}}$ on $(\Omega, \mathcal{F}, \mathbb{P})$ such that conditioned on some σ-algebra $\mathcal{H} \subset \mathcal{F}$ the sequence $\{X_k\}_{k\in\mathbb{N}}$ is iid. De Finetti's Theorem thus allows us to view conditionally iid random vectors $\boldsymbol{X} = (X_1, \ldots, X_d)$ as the first d members of an infinite exchangeable sequence $\{X_k\}_{k\in\mathbb{N}}$. More clearly, a probability law $\mu \in M_+^1(\mathbb{R}^d)$ is conditionally iid if and only if there exists an infinite exchangeable sequence $\{X_k\}_{k\in\mathbb{N}}$ on some probability space such that $\boldsymbol{X} = (X_1, \ldots, X_d) \sim \mu$. At this point it is important to highlight that we deal with a fixed dimension d. In general, it is possible that two truly different probability laws $\gamma_1 \neq \gamma_2$ are mapped onto the same element $\mu \in M_+^1(\mathbb{R}^d)$ under the mapping Θ_d. By virtue of de Finetti's Theorem, this ambiguity vanishes if we let d tend to infinity in (1.4), i.e. the probability laws of H and $\{X_k\}_{k\in\mathbb{N}}$ stand in a one-to-one correspondence, see also Lemma 1.8 below. In many cases of interest, we are actually not only interested in finding *some* γ that is mapped onto a given $\mu \in \mathfrak{M}$ under Θ_d, but actually wish to find such γ which is mapped onto a probability law with a desired property for arbitrary $n \geq 1$ by Θ_n. In order to formalize this idea, we introduce the following definition.

Definition 1.2 (Conditionally iid respecting (P))**.** *Let* (P) *be some property*

which makes sense in arbitrary dimension $n \geq 1$, and define the sets

$$\mathfrak{M}_{n,(\mathrm{P})} := \{\mu \in M_+^1(\mathbb{R}^n) \,:\, \mu \text{ has property } (\mathrm{P})\} \subset M_+^1(\mathbb{R}^n).$$

We say that $\mu \in \mathfrak{M} = \mathfrak{M}_{d,(\mathrm{P})}$ is conditionally iid respecting (P) *if there exists a stochastic process $H \sim \gamma \in M_+^1(\mathfrak{H})$ whose probability law is mapped onto μ under Θ_d and, in addition, is mapped to an element of $\mathfrak{M}_{n,(\mathrm{P})}$ for arbitrary $n \geq 1$ under Θ_n. We furthermore introduce the notation*

$$\mathfrak{M}_{**} := \{\mu \in \mathfrak{M} \,:\, \mu \text{ is conditionally iid respecting } (\mathrm{P})\}.$$

Now we refine Problem 1.1.

Problem 1.3 (Motivating problem refined)**.** *Let* (P) *be a property that makes sense in any dimension, and consider $\mathfrak{M} = \mathfrak{M}_{d,(\mathrm{P})}$. For $\mu \in \mathfrak{M}$ provide necessary and sufficient conditions ensuring that $\mu \in \mathfrak{M}_{**}$.*

In the situation of Problem 1.3 we have $\mathfrak{M}_{**} \subset \mathfrak{M}_*$, and the inclusion can be proper in general, although this is unusual in cases of interest. A non-trivial example for the situation $\mathfrak{M}_{**} \neq \mathfrak{M}_*$ is presented in Example 3.3 in Section 3.2 below. A typical example for (P) is the property of "being a multivariate normal distribution (in some dimension)". For a given d-variate multivariate normal law it is a priori unclear whether there exists an infinite exchangeable sequence with d-margins being equal to the given multivariate normal law and such that all n-margins are multivariate normal as well for $n > d$. This is indeed the case and we have $\mathfrak{M}_{**} = \mathfrak{M}_*$ in this particular situation, as can be inferred from Example 1.1. The typical questions in the theory deal with subsets of $M_+^1(\mathbb{R}^d)$ of the form $\mathfrak{M} = \mathfrak{M}_{d,(\mathrm{P})}$ for a property (P) that makes sense in arbitrary dimension, so most results presented are actually solutions to Problem 1.3 rather than to Problem 1.1, see also paragraph 7.5 below for a further discussion related to this subtlety.

1.3.1. Laws with positive components

If the given family $\mathfrak{M}$ consists only of probability laws on $[0, \infty)^d$, it is convenient to slightly reformulate the stochastic model (1.4). Clearly, if we have non-negative components, necessarily $H_t = 0$ for all $t < 0$ almost surely. Therefore, without loss of generality we may assume that $H = \{H_t\}_{t \geq 0}$ is indexed by $t \in [0, \infty)$. Moreover, applying the substitution $z = -\log(1 - \bar{F})$ it trivially holds true that

$$\mathfrak{H}_+ = \big\{t \mapsto 1 - \exp(-z(t)) \,\big|\, z : [0, \infty) \to [0, \infty] \text{ non-decreasing,}$$
$$\text{right-continuous, with } z(0) \geq 0 \text{ and } \lim_{t \to \infty} z(t) = \infty\big\}.$$

One may therefore rewrite the canonical construction (1.4) as

$$X_k := \inf\{t \geq 0 \,:\, Z_t > \epsilon_k\}, \quad k = 1, \ldots, d, \tag{1.5}$$

where the $\epsilon_k := -\log(1 - U_k)$, $k = 1, \ldots, d$, are now iid exponential random variables with unit mean, and $Z = \{Z_t\}_{t \geq 0}$ is now no longer a distribution function, but instead a non-decreasing, right-continuous process with $Z_0 \geq 0$ and $\lim_{t \to \infty} Z_t = \infty$, related to H via the substitution $Z_t = -\log(1 - H_t)$. Figure 1 illustrates one realization of the simulation mechanism (1.5).
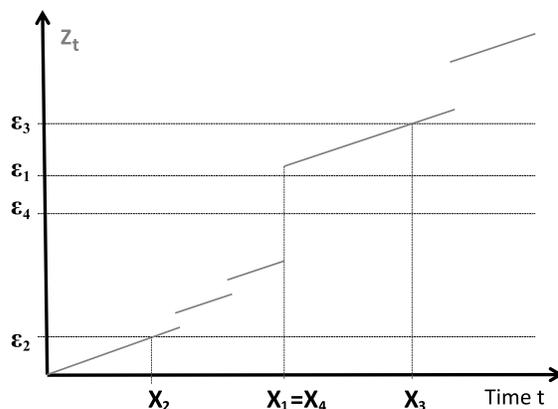
FIG 1. *Illustration of one simulation of the canonical construction (1.5) in dimension $d = 4$. One observes that the process $Z = \{Z_t\}_{t \geq 0}$ in this particular illustration has jumps, and therefore there is a positive probability that two components of $\mathbf{X}$ take the identical value. This does not happen if Z is a continuous process, see also Lemma 1.5 below.*

1.4. General properties of conditionally iid models

In this section we briefly collect some general properties of conditionally iid models. To this end, throughout this section we assume that $\mathfrak{M} = M_+^1(\mathbb{R}^d)$ denotes the family of all d-dimensional probability laws on $\mathbb{R}^d$ and we collect general properties of $\mathfrak{M}_*$.

1.4.1. Positive dependence

If the law of $\mathbf{X}$ is in $\mathfrak{M}_*$, the covariance matrix of $\mathbf{X}$, provided existence, cannot have negative entries.

Lemma 1.2 (Non-negative correlations)**.** *If the law of $\mathbf{X}$ is in $\mathfrak{M}_*$ and the covariance matrix of $\mathbf{X}$ exists, then all its entries are non-negative.*

Proof. This follows from precisely the same computations that have been carried out already in (1.1) and (1.2) for the particular example of the multivariate normal distribution. □

Correlation coefficients are sometimes inappropriate dependence measurements outside the Gaussian paradigm, see [81, 105]. For instance, their existence

depends on the existence of second moments, or we might have a correlation co-
efficient that is strictly less than one despite the fact that one component of the
random vector is a monotone function of the other, since correlation coefficients
depend on the marginal distributions as well. For these reasons, several alterna-
tive dependence measurements have been developed. One popular among them
is the concordance measurement *Kendall's Tau*. Recall that $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^2$ are called
concordant if $(x_1 - y_1)(x_2 - y_2) > 0$ and *discordant* if $(x_1 - y_1)(x_2 - y_2) < 0$.
In words, concordance means that one of the two points lies north-east of the
other, while discordance means that one of the two lies north-west of the other.
Kendall's Tau for a bivariate random vector $\boldsymbol{X}$ is defined as the difference be-
tween the probability of concordance and the probability of discordance for two
independent copies of $\boldsymbol{X}$. If $\boldsymbol{X}$ is conditionally iid, Kendall's Tau is necessarily
non-negative.

Lemma 1.3 (Non-negative Kendall's Tau). *If the law of $\boldsymbol{X} = (X_1, X_2)$ is in*
$\mathfrak{M}_*$, *then Kendall's Tau is necessarily non-negative.*

Proof. Let $\boldsymbol{X}^{(1)}$ and $\boldsymbol{X}^{(2)}$ be two independent copies of $\boldsymbol{X}$, both defined on
some common probability space. By assumption we find a σ-algebra $\mathcal{H}$ such
that conditioned on $\mathcal{H}$ all four random variables $X_1^{(1)}, X_2^{(1)}, X_1^{(2)}, X_2^{(2)}$ are in-
dependent with respective distribution functions $H^{(1)}$ (for $X_1^{(1)}, X_2^{(1)}$) and $H^{(2)}$
(for $X_1^{(2)}, X_2^{(2)}$). Notice that $H^{(1)}$ and $H^{(2)}$ are iid. We compute

$$
\begin{aligned}
\mathbb{P}&\big((X_1^{(1)} - X_1^{(2)})(X_2^{(1)} - X_2^{(2)}) > 0\big) \\
&= \mathbb{E}\big[\mathbb{P}\big((X_1^{(1)} - X_1^{(2)})(X_2^{(1)} - X_2^{(2)}) > 0 \,\big|\, \mathcal{H}\big)\big] \\
&= \mathbb{E}\big[\mathbb{P}\big(X_1^{(1)} > X_1^{(2)}, X_2^{(1)} > X_2^{(2)} \,|\, \mathcal{H}\big) + \mathbb{P}\big(X_1^{(1)} < X_1^{(2)}, X_2^{(1)} < X_2^{(2)} \,|\, \mathcal{H}\big)\big] \\
&= \mathbb{E}\Big[\Big(\int H_{x-}^{(2)} \, \mathrm{d}H_x^{(1)}\Big)^2\Big] + \mathbb{E}\Big[\Big(\int H_{x-}^{(1)} \, \mathrm{d}H_x^{(2)}\Big)^2\Big]
\end{aligned}
$$

and analogously

$$
\begin{aligned}
\mathbb{P}&\big((X_1^{(1)} - X_1^{(2)})(X_2^{(1)} - X_2^{(2)}) < 0\big) \\
&= \mathbb{E}\big[\mathbb{P}\big((X_1^{(1)} - X_1^{(2)})(X_2^{(1)} - X_2^{(2)}) < 0 \,\big|\, \mathcal{H}\big)\big] \\
&= 2\,\mathbb{E}\Big[\int H_{x-}^{(2)} \, \mathrm{d}H_x^{(1)} \int H_{x-}^{(1)} \, \mathrm{d}H_x^{(2)}\Big],
\end{aligned}
$$

so that Kendall's Tau equals

$$
\mathbb{E}\Big[\Big(\int H_{x-}^{(2)} \, \mathrm{d}H_x^{(1)} - \int H_{x-}^{(1)} \, \mathrm{d}H_x^{(2)}\Big)^2\Big] \geq 0,
$$

establishing the claim. $\square$

The next lemma is less intuitive on first glimpse, but like Lemmata 1.2 and 1.3
it qualitatively states that laws in $\mathfrak{M}_*$ exhibit some sort of "positive" depen-
dence. In order to understand it, it is useful to recall the notion of *majorization,*

see [80] for a textbook account on the topic. A vector $\boldsymbol{a} = (a_1, \ldots, a_d)$ is said to *majorize* a vector $\boldsymbol{b} = (b_1, \ldots, b_d)$ if

$$\sum_{k=n}^{d} a_{[k]} \geq \sum_{k=n}^{d} b_{[k]}, \quad n = 2, \ldots, d, \quad \sum_{k=1}^{d} a_k = \sum_{k=1}^{d} b_k.$$

Intuitively, the entries of $\boldsymbol{b}$ are "closer to each other" than the entries of $\boldsymbol{a}$, even though the sum of all entries is identical for both vectors. For instance, the vector $(1, 0, \ldots, 0)$ majorizes the vector $(1/2, 1/2, 0, \ldots, 0)$, which majorizes $(1/3, 1/3, 1/3, 0, \ldots, 0)$, and so on.

Lemma 1.4 (A link to majorization). *Consider $\boldsymbol{X}$ with law in $\mathfrak{M}_*$. Further, let $\boldsymbol{Y} = (Y_1, \ldots, Y_d)$ be a random vector with components that are iid and satisfy*

$$Y_1 \overset{d}{=} X_1.$$

We denote $F_Z(x) := \mathbb{P}(Z \leq x)$ for real-valued Z and $x \in \mathbb{R}$.

(a) *For arbitrary $x \in \mathbb{R}$ the vector $\big(F_{Y_{[1]}}(x), \ldots, F_{Y_{[d]}}(x)\big)$ majorizes the vector $\big(F_{X_{[1]}}(x), \ldots, F_{X_{[d]}}(x)\big)$.*

(b) *For any measurable, real-valued function g which is monotone on the support of X_1 the vector $\big(\mathbb{E}[g(Y_{[1]})], \ldots, \mathbb{E}[g(Y_{[d]})]\big)$ majorizes the vector $\big(\mathbb{E}[g(X_{[1]})], \ldots, \mathbb{E}[g(X_{[d]})]\big)$.*

Proof. This is [98, Theorem 2.2 and Corollary 2.3]. By definition,

$$\sum_{k=1}^{d} F_{X_{[k]}}(x) = \sum_{k=1}^{d} F_{X_k}(x) = d\, F_{X_1}(x) = d\, F_{Y_1}(x) = \sum_{k=1}^{d} F_{Y_k}(x) = \sum_{k=1}^{d} F_{Y_{[k]}}(x).$$

Since $F_{X_{[1]}}(x) \geq \ldots \geq F_{X_{[d]}}(x)$ and $F_{Y_{[1]}}(x) \geq \ldots \geq F_{X_{[d]}}(x)$, for part (a) we have to show that

$$\sum_{k=1}^{n} F_{X_{[k]}}(x) \leq \sum_{k=1}^{n} F_{Y_{[k]}}(x), \quad n = 1, \ldots, d-1.$$

First, it is not difficult to verify that

$$h_{n,d}(p) := \sum_{k=1}^{n} \sum_{i=k}^{d} \binom{d}{i} p^i (1-p)^{d-i}, \quad p \in [0,1],$$

is concave for arbitrary $1 \leq n \leq d$. Second, concavity implies that

$$\sum_{k=1}^{n} F_{X_{[k]}}(x) = \mathbb{E}\Big[\sum_{k=1}^{n} \mathbb{P}(X_{[k]} \leq x \mid \mathcal{H}) \Big] = \mathbb{E}[h_{n,d}(H_x)]$$

$$\leq h_{n,d}(\mathbb{E}[H_x]) = h_{n,d}(\mathbb{P}(Y_1 \leq x)) = \sum_{k=1}^{n} F_{Y_{[k]}}(x),$$

where Jensen's inequality has been used. Making use of the relation $\mathbb{E}[Z] = \int_0^\infty 1 - F_Z(z)\,\mathrm{d}z - \int_{-\infty}^0 F_Z(z)\,\mathrm{d}z$ for real-valued random variables Z, part (b) is obtained from (a) for the case $g(x) = x$. For the general case, one simply has to observe that the law of $\big(g(X_1), \ldots, g(X_d)\big)$ is also in $\mathfrak{M}_*$ and due to monotonicity of g we have either $g(X_{[1]}) \leq \ldots \leq g(X_{[d]})$ in the non-decreasing case or $g(X_{[d]}) \leq \ldots \leq g(X_{[1]})$ in the non-increasing case. $\qquad\square$

Intuitively, statement (b) in case $g(x) = x$ states that the expected values of the order statistics $\mathbb{E}[X_{[k]}]$, $k = 1, \ldots, d$, are closer to each other than the respective values if the components of $\boldsymbol{X}$ were iid (and not only conditionally iid). Intuitively, the components of a random vector $\boldsymbol{X}$ with components that are conditionally iid are thus less spread out than the components of a random vector with iid components. Thus, Lemmata 1.2, 1.3 and 1.4 show that dependence models built from a conditionally iid setup can only capture the situation of components being "more clustered" than independence, which is loosely interpreted as "positive dependence". Generally speaking, negative dependence concepts are more complicated than positive dependence concepts in dimensions $d \geq 3$, the interested reader is referred to [89] for a nice overview and references dealing with such concepts.

Whereas Lemmata 1.2, 1.3 and 1.4 provide three particular quantifications for positive dependence of a conditionally iid probability law, many other possible concepts of positive dependence can be found in the literature, a textbook account on the topic is [86]. [95, Theorem 4] claims that if $\boldsymbol{X} = (X_1, \ldots, X_d)$ is conditionally iid and $x \mapsto \mathbb{P}(X_1 \leq x)$ is continuous, then

$$\mathbb{P}(\boldsymbol{X} \leq \boldsymbol{x}) \geq \prod_{k=1}^d \mathbb{P}(X_k \leq x_k), \quad \boldsymbol{x} \in \mathbb{R}^d,$$

a positive dependence property called *positive lower orthant dependency*. However, here is a counterexample showing that [95, Theorem 4] is not correct and conditionally iid random vectors need not exhibit positive lower orthant dependency in general.

Example 1.2 (Conditionally iid $\not\Longrightarrow$ positive lower orthant dependency). *Let M be uniformly distributed on $[0, 1/2]$. Conditioned on M let $\boldsymbol{X} = (X_1, X_2)$ be a vector of two iid random variables which have distribution function*

$$H_t = \frac{1}{2}\,1_{\{-M+\frac{1}{2} \leq t < M+\frac{1}{2}\}} + 1_{\{t \geq M+\frac{1}{2}\}}, \quad t \in \mathbb{R}.$$

It is not difficult to compute that

$$\mathbb{P}(\boldsymbol{X} \leq \boldsymbol{x}) = \mathbb{E}[H_{x_1}\,H_{x_2}] = \frac{1}{2}\,x_{[1]} + \frac{1}{2}\,\max\{0, x_1 + x_2 - 1\}, \quad x_1, x_2 \in [0, 1],$$

and the distribution function of $\boldsymbol{X}$ is a copula, i.e. has standard uniform one-dimensional marginals. In particular,

$$\mathbb{P}\Big(X_1 \leq \frac{1}{4}, X_2 \leq \frac{3}{4}\Big) = \frac{1}{8} < \frac{3}{16} = \mathbb{P}\Big(X_1 \leq \frac{1}{4}\Big)\,\mathbb{P}\Big(X_2 \leq \frac{3}{4}\Big),$$

contradicting positive lower orthant dependency. Notice that Kendall's Tau for $\boldsymbol{X}$ *is exactly equal to zero, and also the correlation coefficient between the components of* $\boldsymbol{X}$ *equals zero. Figure 2 depicts a scatter plot of* 1000 *samples from* $\boldsymbol{X}$.

In contrast to Example 1.2, [24] prove that the weaker property $\mathbb{P}(X_1 \in A, \ldots, X_d \in A) \geq \mathbb{P}(X_1 \in A)^d$ holds indeed true for conditionally iid $\boldsymbol{X}$ and an arbitrary measurable set $A \subset \mathbb{R}$. This makes clear that a decisive point in Example 1.2 is that the considered x_i are different.
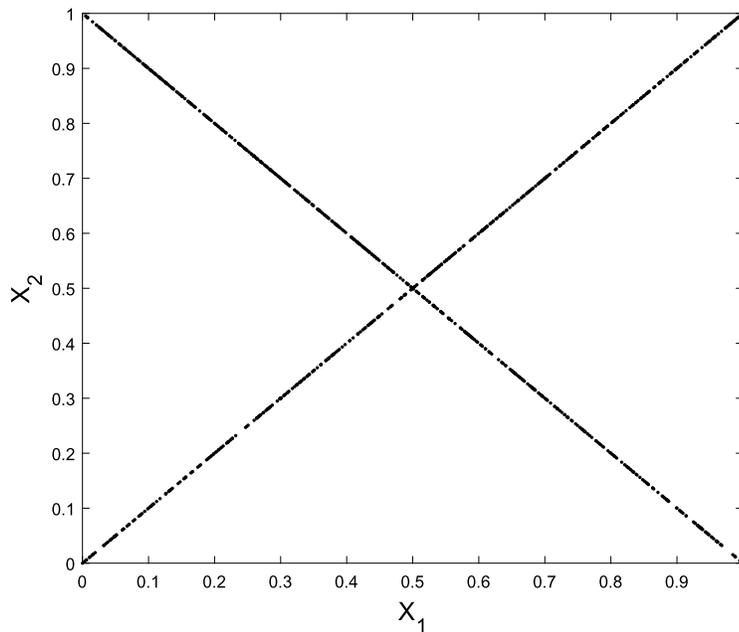
Fig 2. 1000 *samples of* (X_1, X_2) *from Example 1.2.*

1.4.2. Further properties

Even though it is obvious, we find it educational to point out explicitly that path continuity of H corresponds to the absence of a singular component in the law of $\boldsymbol{X}$.

Lemma 1.5 (Path continuity of H)**.** *Let* $H \sim \gamma \in M_+^1(\mathfrak{H})$ *and consider the random vector* $\boldsymbol{X} = (X_1, \ldots, X_d)$ *constructed in Equation (1.4) for arbitrary* $d \geq 2$. *Then* $\mathbb{P}(X_1 = X_2) = 0$ *if and only if the paths of* H *are almost surely continuous.*

Proof. Conditioned on the σ-algebra $\mathcal{H}$ generated by H, the random variables X_1, X_2 are iid with distribution function H. Since two iid random variables take

exactly the same value with positive probability if and only if their common distribution function has at least one jump, the claim follows. □

The following result is shown in [98, Proposition 4.2], but we present a slightly different proof.

Lemma 1.6 (Closure under convergence in distribution)**.** *If* $\boldsymbol{X}^{(n)}$ *are conditionally iid and converge in distribution to* $\boldsymbol{X}$*, then the law of* $\boldsymbol{X}$ *is also conditionally iid.*

Proof. Since we only deal with a statement in distribution, we are free to assume that each $\boldsymbol{X}^{(n)}$ is represented as in (1.4) from some stochastic process $H^{(n)} = \{H_t^{(n)}\}_{t \in \mathbb{R}}$, and all objects are defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The random objects $H^{(n)}$ take values in the set of distribution functions of random variables taking values in $[-\infty, \infty]$. This set is compact by Helly's Selection Theorem and Hausdorff when equipped with the topology of pointwise convergence at all continuity points of the limit, see [100]. Thus, the probability measures on this set are a compact set by [3, Corollary II.4.2, p. 104]. This implies that we find a convergent subsequence $\{n_k\}_{k \in \mathbb{N}} \subset \mathbb{N}$ such that $H^{(n_k)}$ converges in distribution to some limiting stochastic process H, which takes itself values in the set of distribution functions of random variables taking values in $[-\infty, \infty]$. It is now not difficult to see that

$$\mathbb{P}(X_1 \leq x_1, \ldots, X_d \leq x_d) = \lim_{k \to \infty} \mathbb{P}(X_1^{(n_k)} \leq x_1, \ldots, X_d^{(n_k)} \leq x_d)$$
$$= \lim_{k \to \infty} \mathbb{E}[H_{x_1}^{(n_k)} \cdots H_{x_d}^{(n_k)}] = \mathbb{E}[H_{x_1} \cdots H_{x_d}],$$

where bounded convergence is used in the last equality. This implies that the law of $\boldsymbol{X}$ can be constructed canonically like in (1.4), hence $\boldsymbol{X}$ is conditionally iid. Finally, since $\boldsymbol{X}$ is assumed to take values in $\mathbb{R}^d$, necessarily H is almost surely the distribution function of a random variable taking values in $\mathbb{R}$ (instead of $[-\infty, \infty]$). □

Recall that a random vector $(X_1, \ldots, X_d)$ is called *radially symmetric* if there exist $\mu_1, \ldots, \mu_d \in \mathbb{R}$ such that

$$(X_1 - \mu_1, \ldots, X_d - \mu_d) \overset{d}{=} (\mu_1 - X_1, \ldots, \mu_d - X_d).$$

If $(X_1, \ldots, X_d)$ is constructed as in Equation (1.4), then radial symmetry can be translated into a symmetry property of the random distribution function H, which is the content of the following lemma.

Lemma 1.7 (Radial symmetry)**.** *Let* $H \sim \gamma \in M_+^1(\mathfrak{H})$*. The random vector* $(X_1, \ldots, X_d)$ *constructed in Equation (1.4) is radially symmetric if and only if there is some* $\mu \in \mathbb{R}$ *such that*

$$\{H_{\mu-t}\}_{t \in \mathbb{R}} \overset{d}{=} \left\{1 - H_{(t+\mu)-}\right\}_{t \in \mathbb{R}}.$$

Proof. On the one hand, we observe

$$\mathbb{P}(\mu - X_1 \leq x_1, \ldots, \mu - X_d \leq x_d) = \mathbb{P}(H_{(\mu+x_1)-} \leq U_1, \ldots, H_{(\mu+x_d)-} \leq U_d)$$
$$= \mathbb{E}\Big[(1 - H_{(\mu+x_1)-}) \cdots (1 - H_{(\mu+x_d)-})\Big].$$

On the other hand, we have

$$\mathbb{P}(X_1 - \mu \leq x_1, \ldots, X_d - \mu \leq x_d) = \mathbb{E}[H_{\mu-x_1} \cdots H_{\mu-x_d}],$$

from where the claimed equivalence can now be deduced easily. Notice that the conditionally iid structure implies that d can be chosen arbitrary and the law of H is determined uniquely by the law of an infinite exchangeable sequence $\{X_k\}_{k\in\mathbb{N}}$ constructed as in (1.4) with $d \to \infty$. $\qquad\square$

Example 1.3 (The multivariate normal law, again). *The most prominent radially symmetric distribution is the multivariate normal law. Recalling Example 1.1, it follows from (1.3) that $\mathcal{N}(\boldsymbol{\mu}, \Sigma)_*$, the conditionally iid normal laws, are induced by the stochastic process $\{H_t\}_{t\geq 0}$ given by*

$$H_t = \Phi\left(\frac{\frac{t-\mu}{\sigma} - \sqrt{\rho}\, M}{\sqrt{1-\rho}}\right), \quad t \in \mathbb{R}, \tag{1.6}$$

for some $\mu \in \mathbb{R}$, $\sigma > 0$, and $\rho \in [0,1]$, and a random variable $M \sim \Phi =$ distribution function of a standard normal law. The reader may check herself that this random distribution function H satisfies the property of Lemma 1.7.

An immediate but quite useful property of a conditionally iid model is the following corollary to the classical Glivenko-Cantelli Theorem.

Lemma 1.8 (Conditional Glivenko-Cantelli). *Let $\{X_k\}_{k\in\mathbb{N}}$ be an infinite exchangeable sequence defined by the canonical construction (1.4) from an infinite iid sequence $\{U_k\}_{k\in\mathbb{N}}$ and an independent random distribution function $H \sim \gamma \in M^1_+(\mathfrak{H})$. It holds almost surely and uniformly in $t \in \mathbb{R}$ that*

$$\frac{1}{d}\sum_{k=1}^{d} 1_{\{X_k \leq t\}} \longrightarrow H_t, \quad \text{as } d \to \infty.$$

Proof. Follows immediately from the classical Glivenko-Cantelli Theorem, which is applied in the second equality below:

$$\mathbb{P}\Big(\lim_{d\to\infty}\sup_{t\in\mathbb{R}}\Big|\frac{1}{d}\sum_{k=1}^{d} 1_{\{X_k\leq t\}} - H_t\Big| = 0\Big)$$
$$= \mathbb{E}\Big[\mathbb{P}\Big(\lim_{d\to\infty}\sup_{t\in\mathbb{R}}\Big|\frac{1}{d}\sum_{k=1}^{d} 1_{\{X_k\leq t\}} - H_t\Big| = 0 \,\Big|\, \mathcal{H}\Big)\Big] = \mathbb{E}[1] = 1.$$

The claim is established. $\qquad\square$

The stochastic nature of the process $\{H_t\}_{t\in\mathbb{R}}$ clearly determines the law of $\boldsymbol{X}$. Conversely, Lemma 1.8 tells us that the law of the d-dimensional vector $\boldsymbol{X}$ does not determine the law of the underlying latent factor $\{H_t\}_{t\in\mathbb{R}}$ in general, but accomplishes this in the limit as $d \to \infty$. Given some infinite exchangeable sequence of random variables $\{X_k\}_{k\in\mathbb{N}}$, it shows how we can recover its latent random distribution function H.

A rather obvious property of the set $\mathfrak{M}_*$ is convexity.

Lemma 1.9 ($\mathfrak{M}_*$ is convex with extremal boundary the product measures). *If $\mu_1, \mu_2 \in \mathfrak{M}_*$ and $\epsilon \in (0,1)$, then $\epsilon\,\mu_1 + (1-\epsilon)\,\mu_2 \in \mathfrak{M}_*$. Furthermore, if $\mu \in \mathfrak{M}_*$ is* extremal, *meaning that $\mu = \epsilon\,\mu_1 + (1-\epsilon)\,\mu_2$ for some $\epsilon \in (0,1)$ and $\mu_1, \mu_2 \in \mathfrak{M}_*$ necessarily implies $\mu = \mu_1 = \mu_2$, then μ is a product measure*[3].

Proof. The convexity of $\mathfrak{M}_*$ is an immediate transfer from the (obvious) convexity of $M^1_+(\mathfrak{H})$ under the mapping Θ_d, as the reader can readily check herself. That product measures are extremal is also obvious. Finally, consider an extremal element $\mu \in \mathfrak{M}_*$. Since μ is conditionally iid, there is a probability measure $\gamma \in M^1_+(\mathfrak{H})$ such that $\mu\big((-\boldsymbol{\infty}, \boldsymbol{x}]\big) = \int_{\mathfrak{H}} h(x_1) \cdots h(x_d)\,\gamma(\mathrm{d}h)$. We choose a Borel set $A \in \mathfrak{H}$ with $\gamma(A) > 0$. If $\gamma(A) = 1$ is the only possible choice, γ is actually a Dirac measure at some element $h \in \mathfrak{H}$ and μ is a product measure, as claimed. Let us derive a contradiction otherwise, in which case $\gamma = \gamma(A)\,\gamma(.\,|\,A) + \gamma(A^c)\,\gamma(.\,|\,A^c)$ and both $\gamma(.\,|\,A)$ and $\gamma(.\,|\,A^c)$ are elements of $M^1_+(\mathfrak{H})$. We obtain a convex combination of μ, to wit

$$\mu((-\boldsymbol{\infty}, \boldsymbol{x}]) = \gamma(A) \int_{\mathfrak{H}} h(x_1) \cdots h(x_d)\,\gamma(\mathrm{d}h\,|\,A)$$
$$+ (1 - \gamma(A)) \int_{\mathfrak{H}} h(x_1) \cdots h(x_d)\,\gamma(\mathrm{d}h\,|\,A^c).$$

Since μ is extremal and $\gamma(.\,|\,A)$ and $\gamma(.\,|\,A^c)$ are different by definition, we obtain the desired contradiction. $\quad\square$

For the sake of completeness, the following remark gives two equivalent conditions for exchangeability of an infinite sequence of random variables.

Remark 1.2 (Conditions equivalent to infinite exchangeability). *A result due to [93] states that an infinite sequence $\{X_k\}_{k\in\mathbb{N}}$ of random variables is exchangeable (or, equivalently, conditionally iid by de Finetti's Theorem) if and only if the law of the infinite sequence $\{X_{n_k}\}_{k\in\mathbb{N}}$ is invariant with respect to the choice of (increasing) subsequence $\{n_k\}_{k\in\mathbb{N}} \subset \mathbb{N}$. Another equivalent condition to exchangeability is that $\{X_k\}_{k\in\mathbb{N}} \stackrel{d}{=} \{X_{\tau+k}\}_{k\in\mathbb{N}}$ for an arbitrary finite stopping time τ with respect to the filtration $\mathcal{F}_n := \sigma(X_1, \ldots, X_n)$, $n \in \mathbb{N}$, see [52].*

1.5. A general (abstract) solution to Problem 1.1

[58] solve Problem 1.1 on an abstract level for the whole family $\mathfrak{M} = M^1_+(\mathbb{R}^d)$

---

[3]Meaning that the components of $\boldsymbol{X} \sim \mu$ are iid.

of all probability laws on $\mathbb{R}^d$. Their result is formulated in the next theorem in our notation[4].

Theorem 1.2 (General solution to Problem 1.1). *The law of $\boldsymbol{X} = (X_1, \ldots, X_d)$ is conditionally iid if and only if*

$$\sup_{g \neq 0} \Big\{ \frac{|\mathbb{E}[g(\boldsymbol{X})]|}{\sup_{\boldsymbol{Y}} |\mathbb{E}[g(\boldsymbol{Y})]|} \Big\} \leq 1,$$

where the outer supremum is taken over all (non-zero) bounded, measurable functions $g : \mathbb{R}^d \to \mathbb{R}$, and the inner supremum in the denominator is taken over all random vectors $\boldsymbol{Y} = (Y_1, \ldots, Y_d)$ with iid components.

Proof. The proof of sufficiency is the difficult part, relying on functional analytic methods, and we refer the interested reader to [58, Theorem 5.1], but provide some intuition below. Necessity of the condition in Theorem 1.2 is the easy part, as will briefly be explained. Without loss of generality we may assume that $\boldsymbol{X}$ is represented by (1.4) with some stochastic process $H \sim \gamma \in M_+^1(\mathfrak{H})$ and an independent sequence of iid variates $U_1, \ldots, U_d$ uniformly distributed on $[0, 1]$. For arbitrary bounded and measurable g we observe

$$|\mathbb{E}[g(\boldsymbol{X})]| = |\mathbb{E}[\mathbb{E}[g(\boldsymbol{X}) \,|\, H]]| = |\mathbb{E}[g(H_{U_1}^{-1}, \ldots, H_{U_d}^{-1})]|$$
$$\leq \sup_{G(.) \in \mathfrak{H}} \big| \mathbb{E}\big[g\big(G^{-1}(U_1), \ldots, G^{-1}(U_d)\big)\big] \big| = \sup_{\boldsymbol{Y}} |\mathbb{E}[g(\boldsymbol{Y})]|. \qquad \square$$

Regarding the intuition of the sufficiency of the condition in Theorem 1.2, we provide one demonstrating example. With X standard normal, we have already seen in Example 1.1 that the random vector $\boldsymbol{X} = (X, -X)$ is not conditionally iid, since it is bivariate normal with negative correlation coefficient. So how does this random vector violate the condition? Considering the bounded measurable function $g(x_1, x_2) = 1_{\{x_1 < 0 < x_2\}}$, we readily observe that $\mathbb{E}[g(\boldsymbol{X})] = \mathbb{P}(X < 0) = 1/2$. If $\boldsymbol{Y} = (Y_1, Y_2)$ is an arbitrary vector with iid components, we observe that

$$\mathbb{E}[g(\boldsymbol{Y})] = \underbrace{\mathbb{P}(Y_1 < 0)}_{\leq \mathbb{P}(Y_1 \leq 0)} \underbrace{\mathbb{P}(Y_2 > 0)}_{=\mathbb{P}(Y_1 > 0)} \leq \mathbb{P}(Y_1 \leq 0) \big(1 - \mathbb{P}(Y_1 \leq 0)\big)$$
$$\leq \sup_{p \in [0,1]} \{p\,(1 - p)\} = 1/4.$$

Consequently, the supremum over all such $\boldsymbol{Y}$ is bounded from above by $1/4$, hence the supremum over all g in the condition of Theorem 1.2 is at least two, hence larger than one. The intuition behind this counterexample is that we have found one particular bounded measurable g that addresses a distributional property of $\boldsymbol{X}$ that sets it apart from any iid sequence. Indeed, the proof of [58] relies on the Hahn-Banach Theorem and thus on a separation argument,

---

[4]In addition to Theorem 1.2, [58] even consider more abstract spaces than $\mathbb{R}$, and also provide a necessary and sufficient criterion for finite extendibility of the law of $\boldsymbol{X} = (X_1, \ldots, X_d)$ to an exchangeable law on $\mathbb{R}^n$ for $n > d$ arbitrary.

since the set of conditionally iid laws can be viewed as a closed convex subset of $M_+^1(\mathbb{R}^d)$ with extremal boundary comprising the laws with iid components, see Lemma 1.9.

On the one hand, Theorem 1.2 is clearly a milestone with regards to the present survey as it solves Problem 1.1 in the general case. On the other hand, it is difficult to apply the derived condition in particular cases of Problem 1.1, when the family $\mathfrak{M}$ is some (semi-)parametric family of interest – simply because the involved suprema are hard to evaluate, see also Example 1.4 below. On a high level, Theorem 1.2 solves Problem 1.1 but not the refined Problem 1.3, which depends on an additional dimension-independent property (P). However, the most compelling results of the theory deal precisely with certain dimension-independent properties (P) of interest, see the upcoming sections as well as paragraph 7.5 for a further discussion. This is because the additional structure provided by some property (P) and the search for structure-preserving extensions is in many cases a more natural and more interesting problem than to simply find *some* extension. We will see that the algebraic structure of this problem is highly case-specific in general, i.e. heavily dependent on (P).

The following example shows that the supremum condition of Theorem 1.2 can lead to an NP-hard problem in general.

Example 1.4 (In general, the extendibility problem is difficult). *If a random vector $\boldsymbol{X} = (X_1, X_2)$ takes values in $\{x_1, \ldots, x_n\}^2 \subset \mathbb{R}^2$, its joint probability distribution is fully described in terms of the matrix $A \in [0, 1]^{n \times n}$ defined via $A_{ij} := \mathbb{P}(X_1 = x_i, X_2 = x_j)$, $1 \leq i, j \leq n$. The probability law of $\boldsymbol{X}$ is exchangeable if and only if $A = A^T$, and the law of $\boldsymbol{X}$ is conditionally iid if and only if[5] there are (row vectors) $\boldsymbol{\lambda} \in S_m$ and $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m \in S_n$ such that $A = \sum_{i=1}^m \lambda_i \boldsymbol{x}_i^T \boldsymbol{x}_i$. Up to normalization, which is only due to the fact that we deal with a probabilistic interpretation, this property is called* complete positivity*. A completely positive matrix A is necessarily also* doubly non-negative, *meaning that it is symmetric, element-wise non-negative and positive semi-definite, and its elements sum up to one. The set of completely positive matrices is a proper subset of doubly non-negative matrices in dimensions $d \geq 5$, and to decide for a given matrix A whether or not it is completely positive is known to be NP-hard, see [21]. Theorem 1.2 implies that $\boldsymbol{X}$, given in terms of A, is conditionally iid if and only if*

$$\sup_{G \in \mathbb{R}^{n \times n} \setminus \{\boldsymbol{0}\}} \left\{ \frac{\left| \sum_{i,j=1}^n G_{ij} A_{ij} \right|}{\sup_{\boldsymbol{y} \in S_n} |\boldsymbol{y}^T G \boldsymbol{y}|} \right\} \leq 1.$$

Notice that the denominator is equal to the absolute value of the maximal eigenvalue of G, the so-called spectral radius of G. As outlined before, this optimization problem must be NP-hard, unless P=NP.

---

[5]We denote by $S_m := \{\boldsymbol{y} \in [0, 1]^m : \|\boldsymbol{y}\|_1 = 1\}$ the m-dimensional unit simplex.

2. Binary sequences

We study probability laws on $\{0,1\}^d$, i.e. on the set of finite binary sequences. We start with a short digression on the little moment problem, because it occupies a commanding role, not only in this section but also in Section 4 below. For a further discussion between the little moment problem and de Finetti's Theorem, the interested reader is also referred to [17].

2.1. Hausdorff's moment problem

If $(b_0, \ldots, b_d)$ is a finite sequence of real numbers, we write $\nabla b_k = b_k - b_{k+1}$ for $k = 0, \ldots, d-1$. The (reversed) difference operator ∇ may be iterated, yielding $\nabla^2 b_k = \nabla(\nabla b_k) = \nabla b_k - \nabla b_{k+1}$ for $k = 0, \ldots, d-2$, and so on. In general we obtain the formula

$$\nabla^j b_k := \sum_{i=0}^{j} (-1)^i \binom{j}{i} b_{k+i}, \quad 0 \le j + k \le d,$$

with $\nabla = \nabla^1$ and ∇^0 the identity.

Definition 2.1 (*d*-monotone sequences). *For $d \in \mathbb{N}$, we say that a finite sequence $(b_0, b_1, \ldots, b_d) \in [0, \infty)^{d+1}$ is d-monotone if $\nabla^{d-k} b_k \ge 0$ for $k = 0, 1, \ldots, d$. An infinite sequence $\{b_k\}_{k \in \mathbb{N}_0}$ with positive members is said to be completely monotone if $(b_0, \ldots, b_d)$ is d-monotone for each $d \ge 2$.*

If $(b_0, \ldots, b_d)$ is d-monotone, then $\nabla^j b_k \ge 0$ for all $0 \le j + k \le d$. In particular, if for $d \ge 2$ the sequence $(b_0, \ldots, b_d)$ is d-monotone, then the shorter sequences $(b_0, \ldots, b_{d-1})$ and $(b_1, \ldots, b_d)$ are both $(d-1)$-monotone. Intuitively, when viewing $(b_0, \ldots, b_d)$ as a function $\{0, \ldots, d\} \to [0, \infty)$, then $(-1)^j \nabla^j b_k$ is something like the j-th derivative at k. With this interpretation in mind, d-monotonicity means that the higher-order derivatives alternate in sign, i.e. first derivative is non-positive, second derivative is non-negative, third derivative is non-positive, and so on. For instance, a 2-monotone sequence is non-increasing ($b_k \ge b_{k+1}$) and "convex" (b_{k+1} is smaller or equal than the arithmetic mean of its neighbors b_k and b_{k+2}). The set of all d-monotone sequences starting with $b_0 = 1$ will be denoted by $\mathcal{M}_d$ in the sequel. Similarly, $\mathcal{M}_\infty$ denotes the set of completely monotone sequences starting with $b_0 = 1$.

Finite sequences in $\mathcal{M}_d$ arise quite naturally in the context of certain discrete probability laws, as will briefly be explained. Consider a probability distribution on the power set (including the empty set) of $\{1, \ldots, d\}$ with the property that subsets with the same cardinality are equally likely outcomes. Concretely, the probability of some subset $I \subset \{1, \ldots, d\}$ only depends on the cardinality $|I|$ of I, and there are only $d + 1$ possible cardinalities. Denote the probability of a subset with cardinality k by p_k, $k = 0, \ldots, d$. Then $p_0, \ldots, p_d$ are non-negative numbers satisfying

$$\sum_{k=0}^{d} \binom{d}{k} p_k = \sum_{I \subset \{1, \ldots, d\}} p_{|I|} = 1. \tag{2.1}$$

Defining the sequence

$$b_k := \sum_{i=0}^{d-k} \binom{d-k}{i} p_{d-i}, \quad k = 0, \ldots, d, \tag{2.2}$$

it follows that $\nabla^{d-k} b_k = p_k \geq 0$ for $k = 0, \ldots, d$. In particular, $b_0 = 1$, so $(b_0, \ldots, b_d) \in \mathcal{M}_d$. Furthermore, the construction (2.2) can be inverted, i.e. is general enough to construct all elements of $\mathcal{M}_d$. To wit, if $(b_0, \ldots, b_d)$ is an arbitrary element in $\mathcal{M}_d$, then the vector of non-negative numbers $(p_0, \ldots, p_d) = (\nabla^d b_0, \nabla^{d-1} b_1, \ldots, \nabla^0 b_d)$ satisfies (2.1), i.e. defines a probability law on the power set of $\{1, \ldots, d\}$ with the aforementioned property. Thus, these probability laws on the power set of $\{1, \ldots, d\}$ and $\mathcal{M}_d$ stand in a one-to-one correspondence. Of course, the power set of $\{1, \ldots, d\}$ can naturally be identified with $\{0, 1\}^d$, when identifying $\boldsymbol{x} \in \{0, 1\}^d$ with the subset $I = \{k : x_k = 1\}$. This explains the occurrence of d-monotonicity in the present section.

The so-called *Hausdorff moment problem* (also known as *little moment problem*) states that the sequences $\mathcal{M}_\infty$ stand in one-to-one correspondence with the moment sequences of random variables taking values on the unit interval $[0, 1]$. Concretely, the sequence $\{b_k\}_{k \in \mathbb{N}_0}$ with $b_0 = 1$ is completely monotone if and only if there is a random variable M taking values in $[0, 1]$ such that $b_k = \mathbb{E}[M^k]$, $k \in \mathbb{N}_0$. Furthermore, the sequence $\{b_k\}_{k \in \mathbb{N}_0}$ uniquely determines the probability law of M. This result is originally due to [44, 45]. See also [32, p. 225] for a proof. Uniqueness of the probability law of M relies heavily on the boundedness of the interval $[0, 1]$ and is due to the fact that polynomials are dense in the space of continuous functions on a bounded interval (Stone-Weierstrass).

It is important to observe that not every d-monotone sequence can be extended to a completely monotone sequence. Being given a d-monotone sequence $(b_0, \ldots, b_d)$, to check whether there exists an extension $b_{d+1}, b_{d+2}, \ldots$ to an infinite completely monotone sequence $\{b_k\}_{k \in \mathbb{N}_0}$ is a purely analytical, highly non-trivial problem, and luckily already solved. This problem is known as the *truncated Hausdorff moment problem*. Its solution, due to [53], states that $(b_0, \ldots, b_d)$ with $b_0 = 1$ can be extended to an element in $\mathcal{M}_\infty$ if and only if the *Hankel determinants* $\hat{H}_1, \check{H}_1, \ldots, \hat{H}_{d-1}, \check{H}_{d-1}$ are all non-negative, which are defined as

$$\hat{H}_{2\ell} := \det \begin{bmatrix} b_0 & \ldots & b_\ell \\ \vdots & & \vdots \\ b_\ell & \ldots & b_{2\ell} \end{bmatrix}, \quad \check{H}_{2\ell} := \det \begin{bmatrix} \nabla b_1 & \ldots & \nabla b_\ell \\ \vdots & & \vdots \\ \nabla b_\ell & \ldots & \nabla b_{2\ell-1} \end{bmatrix},$$

$$\hat{H}_{2\ell+1} := \det \begin{bmatrix} b_1 & \ldots & b_{\ell+1} \\ \vdots & & \vdots \\ b_{\ell+1} & \ldots & b_{2\ell+1} \end{bmatrix}, \quad \check{H}_{2\ell+1} := \det \begin{bmatrix} \nabla b_0 & \ldots & \nabla b_\ell \\ \vdots & & \vdots \\ \nabla b_\ell & \ldots & \nabla b_{2\ell} \end{bmatrix},$$
$$\tag{2.3}$$

for all $\ell \in \mathbb{N}_0$ with $2\ell \leq d$, respectively $2\ell + 1 \leq d$. To provide an example, the sequence $(1, 1/2, \epsilon)$ is 2-monotone for all $\epsilon \in [0, 1/2]$, but can only be extended to a completely monotone sequence if $\epsilon \in [1/4, 1/2]$.

2.2. Extendibility of exchangeable binary sequences

Actually, before Bruno de Finetti published his seminal Theorem 1.1 in 1937, he first published in [18] the same result for the simpler case of binary sequences. In fact, he showed that there is a one-to-one correspondence between exchangeable probability laws on infinite binary sequences and the set $M_+^1([0,1])$ of probability laws on $[0,1]$.

We start with a random vector $\boldsymbol{X} = (X_1, \ldots, X_d)$ taking values in $\{0,1\}^d$. We know from Lemma 1.1 that $\boldsymbol{X}$ needs to be exchangeable in order to possibly be conditionally iid, so we concentrate on the exchangeable case. Let $\boldsymbol{1}_m, \boldsymbol{0}_m$ denote m-dimensional row vectors with all entries equal to one and zero, respectively, and define

$$p_k := \mathbb{P}\big(\boldsymbol{X} = (\boldsymbol{1}_k, \boldsymbol{0}_{d-k})\big), \quad k = 0, \ldots, d.$$

Exchangeability implies that $\mathbb{P}(\boldsymbol{X} = \boldsymbol{x}) = p_{\|\boldsymbol{x}\|_1}$ for arbitrary $\boldsymbol{x} \in \{0,1\}^d$. Consequently, the probability law of $\boldsymbol{X}$ is fully determined by $p_0, \ldots, p_d$.

Theorem 2.1 (Extendibility of exchangeable binary sequences). *Let $\boldsymbol{X}$ be an exchangeable random vector taking values in $\{0,1\}^d$. We denote*

$$p_k := \mathbb{P}\big(\boldsymbol{X} = (\boldsymbol{1}_k, \boldsymbol{0}_{d-k})\big), \quad k = 0, \ldots, d.$$

The following statements are equivalent:

(a) $\boldsymbol{X}$ is conditionally iid.
(b) There is a random variable M taking values in $[0,1]$ such that

$$p_k = \nabla^{d-k} b_k, \quad k = 0, \ldots, d,$$

where $b_k := \mathbb{E}[M^k]$ for $k = 0, \ldots, d$.
(c) The Hankel determinants in (2.3) are all non-negative, for all $\ell \in \mathbb{N}_0$ with $2\ell \leq d$, respectively $2\ell + 1 \leq d$, where

$$b_k := \sum_{i=0}^{d-k} \binom{d-k}{i} p_{d-i}, \quad k = 0, \ldots, d.$$

If one (hence all) of these conditions are satisfied, and $\boldsymbol{U} = (U_1, \ldots, U_d)$ is an iid sequence of random variables that are uniformly distributed on $[0,1]$, independent of M in part (b), then

$$\boldsymbol{X} \overset{d}{=} (1_{\{U_1 \leq M\}}, \ldots, 1_{\{U_d \leq M\}}).$$

Proof. The equivalence of (c) and (b) relies on the truncated Hausdorff moment problem and the identities

$$\mathbb{E}[M^k (1 - M)^{d-k}] = \nabla^{d-k} b_k = p_k, \quad k = 0, \ldots, d,$$

which are all readily verified. To show that (b) implies (a) works precisely along the stochastic model with $\boldsymbol{U}$ as claimed, which is easily checked. To verify the essential part (a) $\implies$ (b) we may simply apply de Finetti's Theorem 1.1 in the special case of a binary sequence[6]: (a) implies that we may without loss of generality assume that the given random vector equals the first d members of an infinite exchangeable binary sequence $\{X_k\}_{k\in\mathbb{N}}$. De Finetti's Theorem 1.1, and as a corollary Lemma 1.8, give us a random variable $H \sim \gamma \in M^1_+(\mathfrak{H})$. Since each X_k takes values only in $\{0,1\}$, necessarily almost every path of H has only one value different from $\{0,1\}$, which is H_t for $t \in [0,1)$. So we define $M := 1 - H_{1/2}$ and observe that conditioned on M, the random variables X_k are iid Bernoulli with success probability M. This implies the claim. $\qquad\square$

In words, the canonical stochastic model for conditionally iid $\boldsymbol{X}$ with values in $\{0,1\}^d$ is a sequence of d independent coin tosses with success probability M which is identical for all coin tosses, but simulated once before the first coin toss. We end this section with two examples of particular interest.

Example 2.1 (Pólya's urn). *Let $r \in \mathbb{N}$ and $b \in \mathbb{N}$ denote the numbers of red and blue balls in an urn. Define a random vector $\boldsymbol{X} \in \{0,1\}^d$ as follows:*

(i) Set $k := 1$.
(ii) Draw a ball at random from the urn.
(iii) Set $X_k := 1$ if the ball is red, and $X_k := 0$ otherwise.
(iv) Put the ball back into the urn with 1 additional ball of the same color.
(v) Increment $k := k + 1$.
(vi) If $k = d + 1$, stop, otherwise go to step (ii).

It is not difficult to observe that $\boldsymbol{X}$ is exchangeable, since

$$\mathbb{P}(\boldsymbol{X} = \boldsymbol{x}) = \frac{\prod_{k=0}^{\|\boldsymbol{x}\|_1-1}(r+k)\,\prod_{k=0}^{d-\|\boldsymbol{x}\|_1-1}(b+k)}{\prod_{k=0}^{d-1}(r+b+k)}, \quad \boldsymbol{x} \in \{0,1\}^d,$$

depends on $\boldsymbol{x}$ only through $\|\boldsymbol{x}\|_1$. Like in Theorem 2.1 we denote by p_k the probability $\mathbb{P}(\boldsymbol{X} = \boldsymbol{x})$ if $\|\boldsymbol{x}\|_1 = k$, $k = 0, \ldots, d$. Using induction over $k = d, d-1, \ldots, 0$ in order to verify (*) below and knowledge about the moments of the Beta-distribution[7] in (**) below, we observe that

$$b_k := \sum_{i=0}^{d-k}\binom{d-k}{i}p_{d-i} = \sum_{i=0}^{d-k}\binom{d-k}{i}\frac{(r+b-1)!\,(r+d-i-1)!\,(b+i-1)!}{(r-1)!\,(b-1)!\,(r+b+d-1)!}$$

$$\overset{(*)}{=} \frac{(r+k-1)!\,(r+b-1)!}{(r-1)!\,(b+r+k-1)!} = \frac{\Gamma(r+k)\,\Gamma(r+b)}{\Gamma(r)\,\Gamma(r+b+k)} \overset{(**)}{=} \mathbb{E}[M^k],$$

where M is a random variable with Beta-distribution whose density is given by

$$f_M(x) = \frac{\Gamma(r+b)}{\Gamma(r)\,\Gamma(b)}\,x^{r-1}\,(1-x)^{b-1}, \quad 0 < x < 1.$$

---

[6]Alternatively, one may construct a completely monotone sequence $\{b_k\}_{k\in\mathbb{N}}$ from an infinite extension of $\boldsymbol{X}$, as demonstrated in [17, Equation (1)], and then make use of Hausdorff's moment problem to obtain M.

[7]See, e.g., [29, p. 35].

Thus, the probability law of $\boldsymbol{X}$ has a conditionally iid representation like in Theorem 2.1. This is one of the traditional examples, in which the conditionally iid structure is a priori not easy to guess from the original motivation of $\boldsymbol{X}$ – in this case a simple urn replacement model.

Example 2.2 (Ferromagnetic Curie-Weiss Ising model)**.** *Motivated by several models in statistical mechanics, [60] study random vectors which admit a density with respect to the law of a vector with iid components which is the exponential of a quadratic form. Concretely, they consider the situation*

$$\mathbb{P}(\boldsymbol{X} \in \mathrm{d}\boldsymbol{x}) = \frac{1}{c_d}\, e^{\frac{1}{2}\left(\sum_{k=1}^{d} x_k\right)^2} \mathbb{P}(\boldsymbol{Y} \in \mathrm{d}\boldsymbol{x}), \qquad (2.4)$$

where $\boldsymbol{Y} = (Y_1, \ldots, Y_d)$ is a vector with iid components and Y_1 is assumed to satisfy

$$\psi(v) := \mathbb{E}\Big[e^{v\, Y_1}\Big] < \infty \text{ for all } v \in \mathbb{R}, \quad c_d := \mathbb{E}\Big[e^{\frac{1}{2}\left(\sum_{k=1}^{d} Y_k\right)^2}\Big] < \infty. \quad (2.5)$$

Of particular interest are cases in which Y_1 takes only finitely many different values. Especially if $Y_1 \in \{0,1\}$, the vector $\boldsymbol{X}$ is a binary sequence like in the present section.

A prominent model motivating the investigation of [60] is the so-called Curie-Weiss Ising model. In probabilistic terms, this model is a probability law on $\{-1,1\}^d$ with two parameters $J, h \in \mathbb{R}$, and the components of a random vector $\boldsymbol{Z}$ with this probability law models the so-called spins at d different sites. These spins can either have the value -1 or 1 (so $\boldsymbol{X} := (1_{\{Z_1>0\}}, \ldots, 1_{\{Z_d>0\}})$ is a transformation from $\{-1,1\}^d$ to $\{0,1\}^d$). We denote for $\boldsymbol{n} \in \{-1,1\}^d$ by $N(\boldsymbol{n})$ the number of 1's in $\boldsymbol{n}$, so that $d - N(\boldsymbol{n})$ equals the number of -1's. For $\boldsymbol{n} \in \{-1,1\}^d$ we define

$$\mathbb{P}(\boldsymbol{Z} = \boldsymbol{n}) = \frac{e^{h\left(2\, N(\boldsymbol{n})-d\right) + \frac{J}{2}\left(2\, N(\boldsymbol{n})-d\right)^2}}{\sum_{k=0}^{d}\binom{d}{k} e^{h\,(2\, k-d) + \frac{J}{2}\,(2\, k-d)^2}}, \quad \boldsymbol{n} \in \{-1,1\}^d, \qquad (2.6)$$

which is an exchangeable probability law on $\{-1,1\}^d$. The exponent of the numerator can be re-written as

$$h\left(2\, N(\boldsymbol{n}) - d\right) + \frac{J}{2}\left(2\, N(\boldsymbol{n}) - d\right)^2 = h\sum_{k=1}^{d} n_k + \frac{J}{2}\sum_{k=1}^{d}\sum_{i=1}^{d} n_k\, n_i$$

and is called the Hamilton operator *of the model. The parameter h determines the external magnetic field and the parameter J denotes a coupling constant. If $J \geq 0$ the model is called ferromagnetic, and for $J < 0$ it is called antiferromagnetic. The ferromagnetic case arises as special case of (2.4), if Y_1 takes values in $\{-\sqrt{J}, \sqrt{J}\}$ with respective probabilities $\mathbb{P}(Y_1 = \sqrt{J}) = 1 - \mathbb{P}(Y_1 = -\sqrt{J}) = \exp(h)/(\exp(h) + \exp(-h))$. Then the law of $\boldsymbol{Z}/\sqrt{J}$ on $\{-1,1\}^d$ is precisely given by the Curie-Weiss Ising model in (2.6) with $J \geq 0$. Notice that for the antiferromagnetic case $J < 0$ this construction is impossible.*

[60, Theorem 1.2] shows that $\boldsymbol{X}$ as defined in (2.4) is conditionally iid. More concretely, conditioned on a random variable M with density[8]

$$f_M(v) := \frac{\psi(x)}{c_d} \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\,\pi}}, \quad x \in \mathbb{R},$$

the components of $\boldsymbol{X}$ are iid with common distribution

$$\mathbb{P}(X_k \in \mathrm{d}x \,|\, M) = \frac{e^{M\,x}}{\psi(M)}\, \mathbb{P}(Y_1 \in \mathrm{d}x), \quad k = 1, \ldots, d,$$

as can easily be checked. In particular, this shows that the aforementioned ferromagnetic Curie-Weiss Ising model is conditionally iid, a result originally due to [87].

3. Classical results for static factor models

Besides the seminal de Finetti's Theorem 1.1, the most popular results in the theory on conditionally iid models concern latent factor processes H of a very special form to be discussed in the present section. To this end, we consider a popular one-parametric family of one-dimensional distribution functions $x \mapsto F_m(x)$ on the real line and put a prior distribution on the parameter $m \in \mathbb{R}$. Then define $H = \{H_t\}_{t \in \mathbb{R}}$ in the canonical construction (1.4) by $H_t = F_M(t)$, where M is some random variable taking values in the set of admissible values for the parameter m. For some prominent families, for example the zero mean normal law or the exponential law, the resulting distribution of the random vector $\boldsymbol{X}$ belongs to a prominent multivariate family of distributions $\mathfrak{M}$, and in fact defines the subset $\mathfrak{M}_* \subset \mathfrak{M}$. Of particular interest is the case when the subset $\mathfrak{M}_*$ of $\mathfrak{M}$ admits a convenient analytical description within the framework of the analytical description of the larger family $\mathfrak{M}$. By construction, in this method of generating conditionally iid laws the dependence-inducing latent factor process H is fully determined already by a single random parameter M, so that it appears unnatural to formulate the model in terms of a "stochastic process" H at all. Since we investigate situations in which this appears to be more natural in later sections, we purposely do this anyway in order to present all results of the present article under one common umbrella. The "single-parameter construction" just described can then be classified as some kind of "static" process within the realm of all possible processes with laws in $M_+^1(\mathfrak{H})$.

More rigorously, let $\{H_t\}_{t \geq 0}$ be the stochastic process from the canonical stochastic representation (1.4) of some multivariate law in $\mathfrak{M}_* \subset \mathfrak{M}$. Equivalently, we view this probability law as a d-dimensional marginal law of some infinite exchangeable sequence of random variables $\{X_k\}_{k \in \mathbb{N}}$, and define $\{H_t\}_{t \geq 0}$ according to Lemma 1.8 as the uniform limit of $\left\{\sum_{k=1}^d 1_{\{X_k \leq t\}}/d\right\}_{t \geq 0}$ as $d \to \infty$. We call the probability law of $\boldsymbol{X} = (X_1, \ldots, X_d)$ *static*, if the natural filtration generated by $\{H_t\}_{t \geq 0}$, i.e. $\mathcal{H}_t := \sigma(H_s \,|\, s \leq t)$, $t \in \mathbb{R}$, is trivial, meaning

---

[8]Completing the square shows that f_M defines a proper density function on $\mathbb{R}$.

that there is some $T \in [-\infty, \infty)$ such that $\mathcal{H}_t = \{\emptyset, \Omega\}$ for $t \leq T$ ("zero information before T") and $\mathcal{H}_t = \mathcal{H}$ for $t > T$ ("total information after T"). The present section reviews well-known families of distributions $\mathfrak{M}$, for which the set $\mathfrak{M}_*$ consists only of static laws. As already mentioned, this situation typically occurs when the random distribution function $H \sim \gamma \in M_+^1(\mathfrak{H})$ is itself given by $H_t = F_M(t)$, for a popular family F_m of one-dimensional distribution functions and a single random variable M representing a random parameter pick.

Example 3.1 (The multivariate normal law revisited). *It follows from Examples 1.1 and 1.3 that $\mathcal{N}(\boldsymbol{\mu}, \Sigma)_*$, the conditionally iid normal laws, are static. The random distribution function H as given by (1.6) obviously satisfies $\mathcal{H} = \sigma(H_t : t \in \mathbb{R}) = \sigma(M) = \mathcal{H}_t$ for arbitrary $t \in \mathbb{R}$.*

Example 3.2 (Binary sequences revisited). *If one (hence all) of the conditions of Theorem 2.1 is satisfied, the law of the binary sequence $\boldsymbol{X} \in \{0, 1\}^d$ is static. Using the notation in Theorem 2.1, the random distribution function H equals $H_t := (1 - M) \, 1_{\{t \geq 0\}} + M \, 1_{\{t \geq 1\}}$. Obviously, $\mathcal{H} = \sigma(H_t : t \in \mathbb{R}) = \sigma(M) = \mathcal{H}_t$ for arbitrary $t \geq 0$.*

In the remaining section we treat the mixture of iid zero mean normals in paragraph 3.1 and the mixture of iid exponentials in paragraph 3.2, since these are the best-studied cases of the theory with nice analytical characterizations. The interested reader is also referred to [20, 90] who additionally study mixtures of iid geometric variables, iid Poisson variables, and iid uniform variables. Mixtures of uniform random variables are discussed in more detail also in Section 3.3 below.

3.1. Spherical laws (aka ℓ_2-norm symmetric laws)

A random vector $\boldsymbol{X} \in \mathbb{R}^d$ is called *spherical* if its probability distribution remains invariant under unitary transformations, such as rotations or reflections, i.e. $\boldsymbol{X} \stackrel{d}{=} \boldsymbol{X} O$ for an arbitrary orthogonal matrix $O \in \mathbb{R}^{d \times d}$. A spherical random vector $\boldsymbol{X}$ has a canonical stochastic representation

$$\boldsymbol{X} \stackrel{d}{=} R \, \boldsymbol{S}, \tag{3.1}$$

where R is a non-negative random variable and the random vector $\boldsymbol{S}$ is independent of R and uniformly distributed on the Euclidean unit sphere $\{\boldsymbol{x} \in \mathbb{R}^d : \|\boldsymbol{x}\|_2 = 1\}$, see [31, Chapter 2]. Hence, realizations of spherical laws must be thought of as being the result of a two-step simulation algorithm: first draw one completely random point on the unit d-sphere, and then scale this point according to some one-dimensional probability distribution on the positive half-axis. In analytical terms, spherical laws are most conveniently treated via their (multivariate) characteristic functions. In particular, it is not difficult to see that $\boldsymbol{X}$ has a spherical law if and only if there exists a real-valued function $\varphi : [0, \infty) \to \mathbb{R}$ in one variable such that

$$\mathbb{E}\left[e^{\mathrm{i} \, (u_1 \, X_1 + \ldots + u_d \, X_d)}\right] = \varphi(\|\boldsymbol{u}\|_2^2), \quad \boldsymbol{u} = (u_1, \ldots, u_d) \in \mathbb{R}^d,$$

see, e.g., [75, Lemma 4.1, p. 161]. The function φ is called the *characteristic generator*. If the components of $\boldsymbol{X}$ are conditionally iid, the function φ is of a very special form, see Schoenberg's Theorem 3.1 below.

If the components of $\boldsymbol{Y} = (Y_1, \ldots, Y_d)$ are iid standard normally distributed, and $M \in (0, \infty)$ is an independent random variable, the random vector $\boldsymbol{X} = M\,\boldsymbol{Y}$ is spherical, because $\boldsymbol{Y}\,O$ is a vector of iid standard normal components for any orthogonal matrix O. Furthermore, the components of $\boldsymbol{X}$ are iid conditioned on the σ-algebra generated by the mixture variable M. Schoenberg's Theorem states that the converse is true as well, i.e. all conditionally iid spherical laws are mixtures of zero-mean normals.

Theorem 3.1 (Schoenberg's Theorem). *Let $\mathfrak{M}$ be the family of d-dimensional spherical laws, and let the law of $\boldsymbol{X}$ be in $\mathfrak{M}$, and assume $\boldsymbol{X}$ is not identically equal to a vector of zeros. The following are equivalent*

(a) *The law of $\boldsymbol{X}$ lies in $\mathfrak{M}_*$.*
(b) *There are iid standard normal random variables $Y_1, \ldots, Y_d$ and an independent positive random variable $M \in (0, \infty)$ such that*

$$\boldsymbol{X} \stackrel{d}{=} M\,(Y_1, \ldots, Y_d).$$

In other words, this means that $\boldsymbol{X}$ has a stochastic representation as in (1.4) with $H_t := \Phi(t/M)$, where Φ denotes the distribution function of a standard normally distributed random variable.
(c) *There is a random variable Z with χ^2-law with d degrees of freedom, a positive random variable $M \in (0, \infty)$, and $\boldsymbol{S}$ uniformly distributed on the Euclidean unit sphere, all three being mutually independent, such that*

$$\boldsymbol{X} \stackrel{d}{=} M\,\sqrt{Z}\,\boldsymbol{S}.$$

In other words, the random variable R of the general representation (3.1) is of the special form $R \stackrel{d}{=} M\,\sqrt{Z}$.
(d) *The (multivariate) characteristic function of $\boldsymbol{X}$ has the form*

$$\mathbb{E}\left[e^{\mathrm{i}\,(u_1\,X_1 + \ldots + u_d\,X_d)}\right] = \varphi(\|\boldsymbol{u}\|_2^2), \quad \boldsymbol{u} = (u_1, \ldots, u_d) \in \mathbb{R}^d,$$

where φ is the Laplace transform φ of some positive random variable.

Proof. Named after [97], see also [56] or [1, p. 22] for further references. An alternative proof is also given in [20]. Statement (c) is only included in order to highlight how the random radius R must be chosen in the canonical representation (3.1) such that the law of $\boldsymbol{X}$ is in $\mathfrak{M}_*$, see also Remark 3.1 below; the interested reader can find a proof for the equivalence (b) $\Leftrightarrow$ (c) in [75, Lemma 4.2, p. 166]. Similarly, the equivalence (b) $\Leftrightarrow$ (d) is obvious, and φ in (d) equals the Laplace transform of the positive random variable $M^2/2$ with M from (b). Trivially, (b) implies (a). We only verify the non-obvious implication (a) $\Rightarrow$ (b), and the proof consists of two steps, following the lines of [1, p. 22].

(i) As a first step we show Maxwell's Theorem, i.e. if $X_1, \ldots, X_d$ are independent and $(X_1, \ldots, X_d)$ is spherically symmetric, then all components X_k are actually iid sharing a normal distribution with mean zero. Since $(X_1, \ldots, X_d)$ is spherically symmetric, its characteristic function can be written as

$$\mathbb{E}\Big[e^{\mathrm{i}\,(u_1\,X_1 + \ldots + u_d\,X_d)}\Big] =: \varphi(\|\boldsymbol{u}\|_2^2), \quad \boldsymbol{u} = (u_1, \ldots, u_d) \in \mathbb{R}^d,$$

for some function φ in one variable, see, e.g., [75, Lemma 4.1, p. 161]. Denoting the characteristic function of X_k by f_k, $k = 1, \ldots, d$, independence of the components implies that $\varphi(\|\boldsymbol{u}\|_2^2) = f_1(u_1) \ldots f_d(u_d)$. Taking the derivative[9] w.r.t. u_k and dividing by $\varphi(\|\boldsymbol{u}\|_2^2)$ on both sides of the last equation implies for arbitrary $k = 1, \ldots, d$ that

$$\frac{f_k'(u_k)}{f_k(u_k)\,2\,u_k} = \frac{\varphi'(\|\boldsymbol{u}\|_2^2)}{\varphi(\|\boldsymbol{u}\|_2^2)}. \tag{3.2}$$

Let $u, y \in \mathbb{R}$ arbitrary. Plugging $\boldsymbol{u} = (u, \ldots, u)$ into (3.2) shows that

$$\frac{f_k'(u)}{f_k(u)\,2\,u} = \frac{\varphi'(\|\boldsymbol{u}\|_2^2)}{\varphi(\|\boldsymbol{u}\|_2^2)} = \frac{f_j'(u)}{f_j(u)\,2\,u}, \quad \text{arbitrary } 1 \le k, j \le d. \tag{3.3}$$

Plugging some $\boldsymbol{u}$ which has u as its k-th and y as its j-th component into (3.2), we observe

$$\frac{f_k'(u)}{f_k(u)\,2\,u} = \frac{\varphi'(\|\boldsymbol{u}\|_2^2)}{\varphi(\|\boldsymbol{u}\|_2^2)} = \frac{f_j'(y)}{f_j(y)\,2\,y} \overset{(3.3)}{=} \frac{f_k'(y)}{f_k(y)\,2\,y}.$$

Since u, y were arbitrary, the functions $x \mapsto f_k'(u)/(f_k(u)\,2\,u)$ are therefore shown to equal some constant c independent of k. Since $f_k(0) = 1$, solving the resulting ordinary differential equation implies that $f_k(u) = \exp(c\,u^2)$. Left to show is now only that $c \le 0$, because this would imply that f_k equals the characteristic function of a zero-mean normal. Since f_k is a characteristic function and as such must be positive semi-definite, the inequality

$$\det \begin{bmatrix} f_k(0-0) & f_k(0-1) \\ f_k(1-0) & f_k(1-1) \end{bmatrix} = f_k(0)^2 - f_k(1)\,f_k(-1) = 1 - e^{2\,c} \ge 0$$

must hold. Clearly, this is only possible for $c \le 0$. The case $c = 0$ is ruled out by the assumption that $\boldsymbol{X}$ is not identical to a vector of zeros.

(ii) If the law of $\boldsymbol{X}$ lies in $\mathfrak{M}_*$ we can without loss of generality assume that $\boldsymbol{X}$ equals the first d members of an infinite exchangeable sequence $\{X_k\}_{k \in \mathbb{N}}$. Conditioned on the tail-σ-field $\mathcal{H} := \cap_{n \ge 1} \sigma(X_n, X_{n+1}, \ldots)$ the random

---

[9]Notice that characteristic functions are differentiable.

variables $X_1, \ldots, X_d$ are iid according to de Finetti's Theorem 1.1. We observe for an arbitrary orthogonal matrix $O \in \mathbb{R}^{d \times d}$ that

$$(\boldsymbol{X}\, O, X_{d+1}, X_{d+2}, \ldots) \stackrel{d}{=} (\boldsymbol{X}, X_{d+1}, X_{d+2}, \ldots),$$

since $\boldsymbol{X}$ is spherical. Since $\mathcal{H}$ does not depend on $\boldsymbol{X}$ (but only on the tail of the infinite sequence), this implies that the conditional distribution of $\boldsymbol{X}$ and $\boldsymbol{X}\, O$ given $\mathcal{H}$ are identical. As O was arbitrary, $\boldsymbol{X}$ conditioned on $\mathcal{H}$ is spherical. Maxwell's Theorem now implies that $\boldsymbol{X}$ conditioned on $\mathcal{H}$ is an iid sequence of zero mean normals. Thus, only the standard deviation may still be a $\mathcal{H}$-measurable random variable, which we denote by M.   □

If (P) in Problem 1.3 is the property of "having a spherical law (in some dimension)", then Schoenberg's Theorem 3.1 also implies that $\mathfrak{M}_* = \mathfrak{M}_{**}$, which follows trivially from the equivalence of (a) and (b), since the stochastic construction in (b) clearly works for arbitrary $n > d$ as well. Furthermore, it is observed that the random distribution function $H_t = \Phi(t/M)$ in part (b) satisfies the condition in Lemma 1.7 with $\mu = 0$, so conditionally iid spherical laws are radially symmetric. In fact, (arbitrary) spherical laws are always radially symmetric, since $(X_1, \ldots, X_d) \stackrel{d}{=} (-X_1, \ldots, -X_d)$ follows immediately from the definition.

Remark 3.1 (Realization of uniform law on Euclidean unit sphere)**.** *Denoting* $\boldsymbol{Y} = (Y_1, \ldots, Y_d)$, *the equivalence (b) $\Leftrightarrow$ (c) in Theorem 3.1 implies*

$$\boldsymbol{S} \stackrel{d}{=} \Big(\frac{Y_1}{\|\boldsymbol{Y}\|_2}, \ldots, \frac{Y_d}{\|\boldsymbol{Y}\|_2} \Big),$$

which shows how to generate realizations of the uniform law on the Euclidean unit sphere from a list of iid standard normals.

Remark 3.2 (Elliptical laws)**.** *Spherical laws are always exchangeable, which is easy to see. A popular method to enrich the family of spherical laws to obtain a larger family beyond the exchangeable paradigm is linear transformation. To wit, for $\boldsymbol{X} \in \mathbb{R}^k$ spherical with characteristic generator φ, $A \in \mathbb{R}^{k \times d}$ some matrix with $\Sigma := A'\, A \in \mathbb{R}^{d \times d}$ and rank of Σ equal to $k \leq d$, and with $\boldsymbol{b} = (b_1, \ldots, b_d)$ some real-valued row vector, the random vector*

$$\boldsymbol{Z} = (Z_1, \ldots, Z_d) = \boldsymbol{X}\, A + \boldsymbol{b} \tag{3.4}$$

is said to have an elliptical law *with parameters $(\varphi, \Sigma, \boldsymbol{b})$. This generalization from spherical laws to elliptical laws is especially well-behaved from an analytical viewpoint, since the apparatus of linear algebra gets along perfectly well with the definition of spherical laws. The most prominent elliptical law is the multivariate normal distribution, which is obtained in the special case when $\varphi(x) = \exp(-x/2)$ is the Laplace transform of the constant $1/2$. The case when $\mathbb{E}[\|\boldsymbol{X}\|_2^2] < \infty$ is of most prominent importance, since the random vector $\boldsymbol{Z}$ then has existing covariance matrix given by $\mathbb{E}[\|\boldsymbol{X}\|_2^2]\, \Sigma / k$.*

Since the normal distribution special case occupies a commanding role when deciding whether or not a spherical law is conditionally iid according to Theorem 3.1(b), and since we have also solved our motivating Problem 1.1 for the multivariate normal law in Example 1.1, it is not difficult to decide when an elliptical law is conditionally iid as well. To wit, in the most important case when $\mathbb{E}[\|\boldsymbol{X}\|_2^2] < \infty$ the random vector $\boldsymbol{Z}$ in (3.4) has a stochastic representation that is conditionally iid if and only if $b_1 = \ldots = b_d$, and $\boldsymbol{Z} \stackrel{d}{=} R\boldsymbol{Y} + \boldsymbol{b}$ with R some positive random variable with finite second moment and $\boldsymbol{Y} = (Y_1, \ldots, Y_d)$ multivariate normal with zero mean vector and covariance matrix such as in Example 1.1, i.e. with identical diagonal elements $\sigma^2 > 0$ and identical off-diagonal elements $\rho\,\sigma^2 \geq 0$.

3.2. ℓ_1-norm symmetric laws

According to [83], a random vector $\boldsymbol{X} \in [0,\infty)^d$ is called ℓ_1-*norm symmetric* if it has a stochastic representation

$$\boldsymbol{X} \stackrel{d}{=} R\,\boldsymbol{S},$$

where R is a non-negative random variable and the random vector $\boldsymbol{S}$ is independent of R and uniformly distributed on the unit simplex $S_d = \{\boldsymbol{x} \in [0,\infty)^d : \|\boldsymbol{x}\|_1 = 1\}$. Comparing this representation to (3.1), the only difference is that $\boldsymbol{S}$ is now uniformly distributed on the unit sphere with respect to the ℓ_1-norm (restricted to the positive orthant $[0,\infty)^d$), rather than on the unit sphere with respect to the Euclidean norm. Consequently, quite similar to spherical laws, realizations of ℓ_1-norm symmetric distributions must be thought of as being the result of the following two-step simulation algorithm: first draw one completely random point on the d-dimensional unit simplex, and then scale this point according to some one-dimensional probability distribution on the positive half-axis.

Remark 3.1 points out an important relationship between the (univariate) standard normal distribution and the uniform law on the Euclidean unit sphere (w.r.t. the Euclidean norm $\|.\|_2$). It is not difficult to observe that the (univariate) standard exponential law plays the analogous role for the uniform law on the unit simplex (w.r.t. the ℓ_1-norm $\|.\|_1$). More precisely, if the components of $\boldsymbol{E} = (E_1, \ldots, E_d)$ are iid exponentially distributed with unit mean, then

$$\boldsymbol{S} := \Big(\frac{E_1}{\|\boldsymbol{E}\|_1}, \ldots, \frac{E_d}{\|\boldsymbol{E}\|_1}\Big)$$

is uniformly distributed on the unit simplex, cf. [75, Lemma 2.2(2), p. 77] or [31, Theorem 5.2(2), p. 115]. An arbitrary ℓ_1-norm symmetric random vector $\boldsymbol{X}$ is represented as

$$\boldsymbol{X} \stackrel{d}{=} R\Big(\frac{E_1}{\|\boldsymbol{E}\|_1}, \ldots, \frac{E_d}{\|\boldsymbol{E}\|_1}\Big) \tag{3.5}$$

with independent R and $\boldsymbol{E}$. With the analogy to the spherical case in mind, heuristic reasoning suggests that $\boldsymbol{X}$ is extendible if and only if R is chosen such that it "cancels" out the denominator of $\boldsymbol{S}$ in distribution. Since $\|\boldsymbol{E}\|_1$ has a unit-scale Erlang distribution with parameter d, this would imply that R should be chosen as $R = Z/M$ for some positive random variable M and an independent random variable Z with Erlang distribution and parameter d. This is precisely the case, as Theorem 3.2 below shows.

Generally speaking, it follows from the canonical stochastic representation (3.5) that

$$\mathbb{P}(X_k > x) = \mathbb{P}\Big(E_k > \frac{x}{R-x} \sum_{i \neq k} E_i, R > x\Big) = \mathbb{E}\Big[e^{-\frac{x}{R-x} \sum_{i \neq k} E_i} 1_{\{R > x\}}\Big]$$

$$= \mathbb{E}\Big[\max\Big\{1 - \frac{x}{R}, 0\Big\}^{d-1}\Big] =: \varphi_{d,R}(x), \quad k = 1, \dots, d,$$

where the last equality uses knowledge about the Laplace transform of the Erlang-distributed random variable $\sum_{i \neq k} E_i$. This means that the marginal survival functions of the components X_k are given by the so-called *Williamson d-transform* $\varphi_{d,R}$ of R. It has been studied in [106], who shows in particular that the law of R is uniquely determined by $\varphi_{d,R}$. A similar computation as above shows that the joint survival function of $\boldsymbol{X}$ is given by

$$\mathbb{P}(\boldsymbol{X} > \boldsymbol{x}) = \varphi_{d,R}(\|\boldsymbol{x}\|_1), \quad \boldsymbol{x} = (x_1, \dots, x_d) \in [0, \infty)^d.$$

Theorem 3.2 solves Problem 1.3 for the property (P) of "having an ℓ_1-norm symmetric law (in some dimension)".

Theorem 3.2 (Conditionally iid ℓ_1-norm symmetric laws)**.** *Let $\varphi : [0, \infty) \to [0, 1]$ be a function in one variable. The following statements are equivalent:*

(a) *There is an infinite sequence of random variables $\{X_k\}_{k \in \mathbb{N}}$ such that for arbitrary $d \in \mathbb{N}$ we have*

$$\mathbb{P}(\boldsymbol{X} > \boldsymbol{x}) = \varphi(\|\boldsymbol{x}\|_1), \quad \boldsymbol{x} \in [0, \infty)^d.$$

(b) *The function φ equals the Laplace transform of some positive random variable M, i.e. $\varphi(x) = \mathbb{E}[\exp(-x\,M)]$.*

In this case, for arbitrary $d \in \mathbb{N}$ we have

$$\boldsymbol{X} = (X_1, \dots, X_d) \stackrel{d}{=} \frac{1}{M} Z \boldsymbol{S} \stackrel{d}{=} \frac{1}{M} \boldsymbol{E},$$

where $\boldsymbol{X}$ as in (a), M as in (b), $\boldsymbol{S}$ uniformly distributed on the unit simplex, $\boldsymbol{E} = (E_1, \dots, E_d)$ a vector of iid unit exponentials, and Z a unit-scale Erlang distributed variate with parameter d, all mutually independent. In other words, $\boldsymbol{X}$ has a stochastic representation as in (1.5) with $Z_t := M\,t$, in particular is conditionally iid.

Proof. The implication (b) $\Rightarrow$ (a) works precisely along the stochastic model claimed, and is readily observed. The implication (a) $\Rightarrow$ (b) is known as Kimberling's Theorem, see [54]. We provide a proof sketch in the sequel. From $d = 1$ we observe that φ is the survival function of some positive random variable. Consequently, due to Bernstein's Theorem[10], it is sufficient to prove that φ is completely monotone, meaning that $(-1)^d \varphi^{(d)} \geq 0$ for all $d \in \mathbb{N}_0$. To this end, recall that

$$(-1)^d \varphi^{(d)}(x) = \Delta_h^d[\varphi](x) + O(h), \quad \Delta_h^d[\varphi](x) := \sum_{k=0}^d \binom{d}{k} (-1)^{d-k} \varphi(x - k\,h),$$

so that it is sufficient to show that $\Delta_h^d[\varphi](x) \geq 0$ for arbitrary $d \in \mathbb{N}_0$ and x, h such that $0 \leq x - d\,h$. To this end, we consider the infinite sequence of random variables $\{U_k\}_{k \in \mathbb{N}}$ with $U_k := \varphi(X_k)$, $k \in \mathbb{N}$, and with $\alpha := \varphi(x/d)$ and $\beta := \varphi(x/d - h) > \alpha$ define the events

$$A_I := \Big(\cap_{j \in I} \{U_j \leq \alpha\} \Big) \cap \Big(\cap_{j \notin I} \{U_j \leq \beta\} \Big), \quad I \subset \{1, \ldots, d\}.$$

A lengthy but straightforward computation, with one application of the inclusion exclusion principle, shows that

$$\Delta_h^d[\varphi](x) = \ldots = \mathbb{P}\Big(A_\emptyset \setminus \big(\cup_{k=1}^d \{U_k \leq \alpha\} \big) \Big) \geq 0,$$

which implies the claim. $\qquad\square$

Remark 3.3 (On involved probability transforms). *In Theorem 3.2, the function φ in part (b) equals the Laplace transform of the random variable M. Furthermore, the survival function of any element in $\mathfrak{M}$ has the form as claimed in (a), only the parameterizing function φ needs not be a Laplace transform in general. Instead, φ always equals the Williamson d-transform of some positive random variable (namely of R). The Williamson d-transform of some random variable is also a Williamson $(d+1)$-transform (of some other random variable), and Laplace transforms can be viewed as a proper subset of Williamson d-transforms given by*

$$\{Laplace\ transforms\} = \bigcap_{d \in \mathbb{N}} \{Williamson\ d\text{-}transforms\}.$$

The most important example for a Williamson d-transform, which is not a Laplace transform (in fact, not even a Williamson $(d+1)$-transform), is given by $\varphi(x) = \varphi(x; d, r) = (1 - x/r)_+^{d-1}$, with a constant $r > 0$. In fact, [106] shows that the set of Williamson d-transforms is a simplex with extremal boundary given by $\{\varphi(.; d, r)\}_{r > 0}$, which is just another way to say that the function $\varphi_{d,R}$ determines the probability law of the positive random variable R uniquely. Similarly, Laplace transforms form a simplex with extremal boundary given by the

---

[10]The original reference is [10], a detailed proof can be found in [8] or [96].

functions $x \mapsto \exp(-m\,x)$ for $m > 0$, which is just another way to say that the function $\varphi(x) = \mathbb{E}[\exp(-x\,M)]$ determines the law of the positive random variable M uniquely. Typical parametric examples for Laplace transforms in the context of ℓ_1-norm symmetric distributions are $\varphi(x) = (1+x)^{-\theta}$ with $\theta > 0$, corresponding to a Gamma distribution of M, or $\varphi(x) = \exp(-x^{\theta})$ with $\theta \in (0,1)$, corresponding to a stable distribution of M.

Remark 3.4 (Archimedean copulas). *Considering $\boldsymbol{X} = (X_1, \ldots, X_d)$ with ℓ_1-norm symmetric law associated with the Williamson d-transform $\varphi = \varphi_{d,R}$, the random vector $(U_1, \ldots, U_d) := \big(\varphi(X_1), \ldots, \varphi(X_d)\big)$ has distribution function*

$$C_\varphi(u_1, \ldots, u_d) := \mathbb{P}(U_1 \leq u_1, \ldots, U_d \leq u_d) = \varphi\big(\varphi^{-1}(u_1) + \ldots + \varphi^{-1}(u_d)\big),$$

for $u_1, \ldots, u_d \in [0,1]$. Recall that φ^{-1} denotes the generalized inverse of φ. The function C_φ is called an Archimedean copula *and the study of ℓ_1-norm symmetric distributions can obviously be translated into an analogous study of Archimedean copulas. In the statistical and applied literature Archimedean copulas have received considerably more attention. For instance, nested and hierarchical extensions of (exchangeable) Archimedean copulas have become quite popular, see, e.g. [16, 47, 49, 82, 107, 65].*

Remark 3.5 (Extension to Liouville distributions). *Analyzing the analogy between spherical laws (aka ℓ_2-norm symmetric laws) and ℓ_1-norm symmetric laws, there is one common mathematical fact on which the analytical treatment of both families relies. To wit, for both families the uniform distribution on the d-dimensional unit sphere can be represented as the normalized vector of iid random variables. In the spherical case the normalized vector $\boldsymbol{Y}/\|\boldsymbol{Y}\|_2$ of d iid standard normals $\boldsymbol{Y} = (Y_1, \ldots, Y_d)$ is uniform on the $\|.\|_2$-sphere, whereas in the ℓ_1-norm symmetric case the normalized vector $\boldsymbol{E}/\|\boldsymbol{E}\|_1$ of d iid standard exponentials $\boldsymbol{E} = (E_1, \ldots, E_d)$ is uniform on the $\|.\|_1$-sphere restricted to the positive orthant $[0,\infty)^d$. Furthermore, in both cases the normalization can be "canceled out" in distribution, that is*

$$\sqrt{Z}\,\frac{\boldsymbol{Y}}{\|\boldsymbol{Y}\|_2} \stackrel{d}{=} \boldsymbol{Y}, \quad R\,\frac{\boldsymbol{E}}{\|\boldsymbol{E}\|_1} \stackrel{d}{=} \boldsymbol{E},$$

where $\sqrt{Z} \stackrel{d}{=} \|\boldsymbol{Y}\|_2$ is independent of $\boldsymbol{Y}$ and Z has a χ^2-law with d degrees of freedom and $R \stackrel{d}{=} \|\boldsymbol{E}\|_1$ is independent of $\boldsymbol{E}$ and has an Erlang distribution with parameter d. The so-called Lukacs Theorem, *due to [63], states that the exponential distribution of the E_k in the last distributional equality can be generalized to a Gamma distribution (but no other law on $(0,\infty)$ is possible). More precisely, if $\boldsymbol{G} = (G_1, \ldots, G_d)$ are independent random variables with Gamma distributions with the same scale parameter, then $\|\boldsymbol{G}\|_1$ is independent of $\frac{\boldsymbol{G}}{\|\boldsymbol{G}\|_1}$, which means that*

$$R\,\frac{\boldsymbol{G}}{\|\boldsymbol{G}\|_1} \stackrel{d}{=} \boldsymbol{G}, \ \text{where } R \stackrel{d}{=} \|\boldsymbol{G}\|_1 \ \text{is independent of } \boldsymbol{G}. \tag{3.6}$$

The random vector $\boldsymbol{S} := \boldsymbol{G}/\|\boldsymbol{G}\|_1$ on the unit simplex is not uniformly distributed unless the G_k happen to be iid exponential. In general, the law of $\boldsymbol{S}$ is called Dirichlet distribution, *parameterized by the d values $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_d)$, where the Gamma density of G_k is given by*

$$f_k(x) = x^{\alpha_k - 1} e^{-x}/\Gamma(\alpha_k), \quad x > 0, \quad k = 1, \ldots, d. \tag{3.7}$$

Notice that the scale parameter of this Gamma distribution is without loss of generality set to one, since it has no influence on the law of $\boldsymbol{S}$. A d-parametric generalization of ℓ_1-norm symmetric laws is obtained by replacing the uniform law of $\boldsymbol{S}$ on the unit simplex (which is obtained for $\alpha_1 = \ldots = \alpha_d$) with a Dirichlet distribution (with arbitrary $\alpha_k > 0$). One says that the random vector $\boldsymbol{X} = R\,\boldsymbol{S}$ with R some positive random variable and $\boldsymbol{S}$ an independent Dirichlet-distributed random vector on the unit simplex, follows a Liouville distribution. *It is precisely the property (3.6) that implies that the generalization to Liouville distributions is still analytically quite convenient to work with, see [84] for a detailed study. Analogous to the ℓ_1-norm symmetric case, the components of $\boldsymbol{X}$ are conditionally iid if $\alpha_1 = \ldots = \alpha_d$ and R satisfies $R \overset{d}{=} Z/M$ with $Z \overset{d}{=} \|\boldsymbol{G}\|_1$ and M some independent positive random variable.*

Having at hand the apparatus of Archimedean copulas, we are now in the position to provide a non-trivial example for the situation $\mathfrak{M}_{**} \subsetneq \mathfrak{M}_*$.

Example 3.3 (In general, $\mathfrak{M}_{**} \subsetneq \mathfrak{M}_*$). *Consider the family $\mathfrak{M} \subset M_+^1([0,1]^2)$ defined by the property* (P) *of "having an Archimedean copula as distribution function and being radially symmetric". It is well-known since [35, Theorem 4.1] that the set $\mathfrak{M}$ comprises precisely Frank's copula family, that is the bivariate distribution function of an element in $\mathfrak{M}$ is either given by $C_{-\infty}(u_1, u_2) := \max\{u_1 + u_2 - 1, 0\}$, by $C_0(u_1, u_2) := u_1\,u_2$, by $C_\infty(u_1, u_2) := \min\{u_1, u_2\}$, or by*

$$C_\theta(u_1, u_2) := -\frac{1}{\theta} \log \left\{ 1 + \frac{\left(e^{-\theta\,u_1} - 1\right)\left(e^{-\theta\,u_2} - 1\right)}{\left(e^{-\theta} - 1\right)} \right\}, \quad u_1, u_2 \in [0, 1],$$

for some parameter $\theta \in (-\infty, 0) \cup (0, \infty)$. Since Kendall's Tau of the copula C_θ is negative in the case $\theta < 0$, Lemma 1.3 implies that the subset $\mathfrak{M}_*$ can at best contain the elements corresponding to $\theta \in [0, \infty]$. Indeed, the cases $\theta \in \{0, \infty\}$ are obviously contained in $\mathfrak{M}_{**} \subset \mathfrak{M}_*$, and for $\theta \in (0, \infty)$ membership in $\mathfrak{M}_*$ follows via the canonical construction (1.4) with the choice $H \sim \gamma \in M_+^1(\mathfrak{H}_+)$, given by

$$H_t = \left(\frac{1 - e^{-\theta\,t}}{1 - e^{-\theta}}\right)^M, \quad t \in [0, 1], \tag{3.8}$$

for a random variable M with logarithmic distribution $\mathbb{P}(M = m) = (1 - \exp(-\theta))^m/(m\,\theta)$, $m \in \mathbb{N}$. Furthermore, we can deduce from Theorem 3.2 that the property of "having an Archimedean copula as distribution function (in arbitrary dimension)" implies that potential elements in $\mathfrak{M}_{**}$ must necessarily

be induced by a stochastic process of the form (3.8) with some positive random variable M, which must necessarily be logarithmic in the radially symmetric case by the result of Frank. The only thing left to check is whether the multivariate Archimedean copula derived from the canonical construction via H defined by (3.8) with logarithmic M is radially symmetric in arbitrary dimension $d \geq 2$. According to Lemma 1.7 this is the case if and only if

$$\left\{ \left(\frac{1 - e^{-\theta\left(\frac{1}{2}-t\right)}}{1 - e^{-\theta}} \right)^M \right\}_{t \in \left[-\frac{1}{2}, \frac{1}{2}\right]} = \{H_{\frac{1}{2}-t}\}_{t \in \left[-\frac{1}{2}, \frac{1}{2}\right]}$$

$$\stackrel{d}{=} \{1 - H_{t+\frac{1}{2}}\}_{t \in \left[-\frac{1}{2}, \frac{1}{2}\right]} = \left\{ 1 - \left(\frac{1 - e^{-\theta\left(t+\frac{1}{2}\right)}}{1 - e^{-\theta}} \right)^M \right\}_{t \in \left[-\frac{1}{2}, \frac{1}{2}\right]}.$$

This statement is false, however, as will briefly be explained. Assuming it was true, then in particular for $t = 0$ we observe that the law of the random variable $H_{\frac{1}{2}}$ was symmetric about $\frac{1}{2}$. In particular, this symmetry would imply

$$0 = \mathbb{E}\left[\left(H_{\frac{1}{2}} - \frac{1}{2} \right)^3 \right] = [\ldots] = \varphi_\theta\left(3\,\varphi_\theta^{-1}(1/2)\right) - \frac{3}{2}\,\varphi_\theta\left(2\,\varphi_\theta^{-1}(1/2)\right) + \frac{1}{4},$$

with $\varphi_\theta(x) = -\log\left(e^{-x}\left(e^{-\theta} - 1\right) + 1\right)/\theta$. Numerically, it is easily verified that the last equality does not hold for any $\theta \in (0, \infty)$, since the right-hand side is strictly smaller than zero. Thus, we see that $\mathfrak{M}_{**}$ consists only of two elements, namely those corresponding to $\{0, \infty\}$. Thus, $\mathfrak{M}_{**} \subsetneq \mathfrak{M}_* \subsetneq \mathfrak{M}$, since $\{0, \infty\} \subsetneq [0, \infty] \subsetneq [-\infty, \infty]$.

3.3. ℓ_∞-norm symmetric laws

[40, Theorem 2] studies random vectors $\boldsymbol{X} = (X_1, \ldots, X_d)$ which are absolutely continuous with density given by

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = g_d(x_{[d]}), \quad \boldsymbol{x} \in (0, \infty)^d, \tag{3.9}$$

with some measurable function $g_d : (0, \infty) \to [0, \infty)$. Recall that

$$x_{[d]} := \max\{x_1, \ldots, x_d\} = \|\boldsymbol{x}\|_\infty$$

equals the ℓ_∞-norm of $\boldsymbol{x} \in (0, \infty)^d$. Since $f_{\boldsymbol{X}}$ is invariant with respect to permutations of the components of $\boldsymbol{x}$, the random vector $\boldsymbol{X}$ is exchangeable. But whether or not it is conditionally iid depends on the choice of g_d. First of all, since $f_{\boldsymbol{X}}$ is a probability density,

$$1 = \iint_{(0,\infty)^d} g_d(x_{[d]})\,\mathrm{d}\boldsymbol{x} = d \int_0^\infty g_d(x)\,x^{d-1}\,\mathrm{d}x, \tag{3.10}$$

constituting a necessary and sufficient integrability condition on g_d such that $f_{\boldsymbol{X}}$ defines a proper probability density. Furthermore, lower-dimensional margins of $\boldsymbol{X}$ have a density of the same structural form, since

$$\int_0^\infty f_{\boldsymbol{X}}(x_1, \ldots, x_{d-1}, x)\,\mathrm{d}x = g_{d-1}(x_{[d-1]}), \quad x_1, \ldots, x_{d-1} > 0,$$

$$\text{where } g_{d-1}(x) := \int_x^\infty g_d(u)\,\mathrm{d}u + x\,g_d(x), \tag{3.11}$$

and the function g_{d-1} is easily checked to satisfy (3.10) in dimension $d-1$, that is $1 = (d-1)\int_0^\infty g_{d-1}(x)x^{d-2}\,\mathrm{d}x$. It is further not difficult to verify that g_d is given in terms of g_{d-1} as

$$g_d(x) = \frac{g_{d-1}(x)}{x} - \int_x^\infty \frac{g_{d-1}(u)}{u^2}\,\mathrm{d}u. \tag{3.12}$$

If $\mathfrak{M}$ denotes the family of all laws with density of the form (3.9), i.e. with a function g_d satisfying (3.10), the following result provides necessary and sufficient conditions on g_d to define a law in $\mathfrak{M}_*$.

Theorem 3.3 (Conditionally iid ℓ_∞-norm symmetric densities). *Let $\mathfrak{M}$ be the family of probability laws on $(0,\infty)^d$ with densities of the form (3.9) with a measurable function $g_d : (0,\infty) \to [0,\infty)$ satisfying (3.10). For $\boldsymbol{X}$ with law in $\mathfrak{M}$, the following statements are equivalent:*

(a) The law of $\boldsymbol{X}$ lies in $\mathfrak{M}_*$.
(b) g_d is non-increasing.
(c) For a vector $\boldsymbol{U} = (U_1,\ldots,U_d)$ whose components are iid uniform on $[0,1]$ and an independent, positive random variable M we have

$$\boldsymbol{X} \stackrel{d}{=} M\,\boldsymbol{U}.$$

Proof. This is [40, Theorem 2]. Clearly, (c) $\Rightarrow$ (a) is obvious. In order to see (b) $\Rightarrow$ (c), we first conclude from (3.10) that

$$0 = \lim_{x\to\infty} g_d(x)\,x^d = \lim_{x\to\infty} g_d\Big(\frac{1}{x}\Big)\frac{1}{x^d}. \tag{3.13}$$

By non-increasingness, we may without loss of generality assume that g_d is right-continuous (otherwise, change to its right-continuous version, which does not change the density $f_{\boldsymbol{X}}$ essentially) and apply integration by parts, then (3.13) and (3.10) imply

$$\int_0^\infty x^d\,\mathrm{d}\big(-g_d(x)\big) = d\int_0^\infty g_d(x)\,x^{d-1}\,\mathrm{d}x = 1.$$

Consequently, $x \mapsto \int_0^x y^d\,\mathrm{d}\big(-g_d(y)\big)$ defines the distribution function of a positive random variable M, and we see that

$$\mathbb{E}[1_{\{M>x\}}\,M^{-d}] = \int_x^\infty \frac{y^d}{y^d}\,\mathrm{d}\big(-g_d(y)\big) = g_d(x).$$

Now let $\boldsymbol{U}$ as claimed be independent of M. Conditioned on M, the density of $M\,\boldsymbol{U}$ is

$$\boldsymbol{x} \mapsto \prod_{k=1}^d \frac{1_{\{0<x_k<M\}}}{M} = 1_{\{0<x_{[d]}<M\}}\frac{1}{M^d}.$$

Integrating out M, the density of $M\boldsymbol{U}$ is found to be

$$\int_0^\infty 1_{\{x_{[d]}<m\}} \frac{1}{m^d} \mathrm{d}\mathbb{P}(M \le m) = \mathbb{E}[1_{\{M>x_{[d]}\}} M^{-d}] = g_d(x_{[d]}),$$

which shows (c). The hardest part is (a) $\Rightarrow$ (b). Fix $\epsilon > 0$ arbitrary. Due to measurability of g_d, Lusin's Theorem guarantees continuity of g_d on a set C_ϵ whose complement has Lebesgue measure less than ϵ. Without loss of generality we may assume that all points t in C_ϵ are density points, i.e. satisfy

$$\lim_{\delta \searrow 0} \frac{\lambda(C_\epsilon \cap [t-\delta, t+\delta])}{2\,\delta} = 1,$$

where λ denotes Lebesgue measure. Let $\{X_k\}_{k\in\mathbb{N}}$ an infinite exchangeable sequence such that d-margins have the density $f_{\boldsymbol{X}}$. Fix $t \ge s$ arbitrary. We define the sequence of random variables $\{\xi_k\}_{k\in\mathbb{N}}$ by

$$\xi_k := \frac{1}{2\,\delta}\big(1_{\{X_k \in A_s\}} - 1_{\{X_k \in A_t\}}\big), \quad k \in \mathbb{N},$$

where $A_x := C_\epsilon \cap [x-\delta, x+\delta]$ for $x \in \{s, t\}$. Notice that the ξ_k are square-integrable and

$$0 \le \mathbb{E}[(\xi_1 + \ldots + \xi_d)^2] = d\,\mathbb{E}[\xi_1^2] + d\,(d-1)\,\mathbb{E}[\xi_1\,\xi_2].$$

If we divide by d^2 and let $d \to \infty$, it follows that $\mathbb{E}[\xi_1\,\xi_2] \ge 0$. Denoting by g_2 the marginal density of (X_1, X_2), we observe

$$0 \le \mathbb{E}[\xi_1\,\xi_2] = \frac{1}{4\,\delta^2} \big\{ \mathbb{E}[1_{\{X_1, X_2 \in A_s\}}] + \mathbb{E}[1_{\{X_1, X_2 \in A_t\}}] - 2\,\mathbb{E}[1_{\{X_1 \in A_t, X_2 \in A_s\}}] \big\}$$

$$= \frac{1}{4\,\delta^2} \Big\{ \iint_{A_s \times A_s} g_2(x_{[2]})\,\mathrm{d}\boldsymbol{x} + \iint_{A_t \times A_t} g_2(x_{[2]})\,\mathrm{d}\boldsymbol{x} - 2 \iint_{A_t \times A_s} g_2(x_{[2]})\,\mathrm{d}\boldsymbol{x} \Big\}$$

$$= g_2(\eta_s) + g_2(\eta_t) - 2\,g_2(\tilde{\eta}_t),$$

for certain values $s - \delta \le \eta_s \le s + \delta$ and $t - \delta \le \eta_t, \tilde{\eta}_t \le t + \delta$ by the mean value theorem for Lebesgue integration. As $\delta \searrow 0$, we thus observe that $g_2(s) \ge g_2(t)$, i.e. g_2 is non-increasing. Making use of (3.12) and integrating by parts, we observe that

$$g_3(x) = \int_x^\infty \frac{1}{u}\,\mathrm{d}\big(-g_2(u)\big),$$

which implies that g_3 is non-increasing as well. Inductively, the same argument implies that $g_4, \ldots, g_d$ are all non-increasing. $\qquad\square$

From the equivalence of (a) and (c) in Theorem 3.3 we observe easily that $\mathfrak{M}_* = \mathfrak{M}_{**}$, when considering the property (P) of "having a density of the form (3.9) (in some dimension $d \in \mathbb{N}$)" in Problem 1.3. Notice furthermore that the

law of $M\,\boldsymbol{U}$ is static in the sense defined in the beginning of this section, and we have

$$H_t := \mathbb{P}(X_k \le t \,|\, M) = \max\Big\{0, \min\Big\{1, \frac{t}{M}\Big\}\Big\}, \quad t \in \mathbb{R},$$

for $X_k := M\,U_k$ as defined in part (c) of Theorem 3.3.

Remark 3.6 (Common umbrella of ℓ_p-norm symmetry results)**.** *Theorem 3.3 on ℓ_∞-norm symmetric densities is very similar in nature to Schoenberg's Theorem 3.1 on ℓ_2-norm symmetric characteristic functions and Theorem 3.2 on ℓ_1-norm symmetric survival functions, which makes it a beautiful result with regards to the present survey. The reference [90] considers all these three cases under one common umbrella, and even manages to generalize them in some meaningful sense to the case of arbitrary ℓ_p-norm, with $p \in [1, \infty]$ arbitrary[11]. More precisely, it is shown that an infinite exchangeable sequence $\{X_k\}_{k \in \mathbb{N}}$ of the form $X_k := M\,Y_k$, $k \in \mathbb{N}$, with $M > 0$ and an independent iid sequence $\{Y_k\}_{k \in \mathbb{N}}$ of positive random variables is ℓ_p-norm symmetric in some meaningful sense[12] if and only if the random variables Y_k have density f_p given by*

$$f_p(x) := \frac{p^{1-\frac{1}{p}}}{\Gamma(1/p)}\, e^{-\frac{x^p}{p}}, \quad 0 < p, x < \infty, \quad f_\infty(x) := 1_{\{x \in (0,1)\}}.$$

Notice that f_1, f_2, and f_∞ are the densities of the unit exponential law, the absolute value of a standard normal law, and the uniform law on $[0, 1]$, respectively. This parametric family in the parameter p is further investigated, and might for instance be characterized by the fact that f_p for $p < \infty$ has maximal entropy among all densities on $(0, \infty)$ with p-th moment equal to one, and f_∞ has maximal entropy among all densities with support $(0, 1)$, which is [90, Theorem 3.5].

An analogous result to Theorem 3.3 on mixtures of the form $M\,\boldsymbol{U}$, when the components of $\boldsymbol{U}$ are iid uniform on $[-1, 1]$, is also presented in [40]. The resulting densities depend on two arguments, $x_{[1]}$ and $x_{[d]}$. Furthermore, [90, Corollary 4.3] prove that an infinite exchangeable sequence $\{X_k\}_{k \in \mathbb{N}}$ satisfies

$$\{X_k\}_{k \in \mathbb{N}} \stackrel{d}{=} \{M\,U_k\}_{k \in \mathbb{N}}, \quad U_1, U_2, \dots \text{ iid uniform on } [0, 1], \ M > 0 \text{ independent},$$

if and only if for arbitrary $d \in \mathbb{N}$ and almost all $s > 0$ the law of $\boldsymbol{X} = (X_1, \dots, X_d)$ conditioned on the event $\{\|\boldsymbol{X}\|_\infty = s\}$ is uniformly distributed on the sphere $\{\boldsymbol{x} \in (0, \infty)^d : \|\boldsymbol{x}\|_\infty = s\}$. This provides an alternative characterization of densities that are ℓ_∞-norm symmetric and conditionally iid.

Remark 3.7 (Relation to non-homogeneous pure birth processes)**.** *[99] provide an interesting interpretation of ℓ_∞-norm symmetric densities, which is briefly explained. Every non-negative function g_d satisfying (3.10) is of the form*

$$g_d(x) = c_d\, r_d(x)\, e^{-\int_0^x r_d(u)\,\mathrm{d}u}$$

---

[11] The authors even allow for $p \in (0, 1)$, but in this case $\|.\|_p$ is no longer a norm.

[12] See [90] for details.

for some non-negative function r_d satisfying $\int_0^\infty r_d(x)\,\mathrm{d}x = \infty$, and some normalizing constant $c_d > 0$. To wit, a function

$$r_d(x) := \frac{g_d(x)}{c_d \int_x^\infty g_d(u)\,\mathrm{d}u}, \quad x > 0, \tag{3.14}$$

for some normalizing constant $c_d > 0$ does the job, as can readily be checked. From such a function r_d we iteratively define functions $r_{d-1}, \ldots, r_1$ by solving the equations

$$r_k(x)\,e^{-R_k(x)} = \frac{e^{-R_{k+1}(x)}}{\int_0^\infty e^{-R_{k+1}(u)}\,\mathrm{d}u}, \quad k = d-1, \ldots, 1, \tag{3.15}$$

where $R_k(x) := \int_0^x r_k(u)\,\mathrm{d}u$ for $k = 1, \ldots, d$. Notice that r_k is related to the right-hand side of (3.15) exactly in the same way as r_d is related to g_d, so the solution (3.14) shows how the r_k look like in terms of r_{k+1}. We define independent positive random variables $E_1, \ldots, E_d$ with survival functions $\mathbb{P}(E_k > x) = \exp(-R_k(x))$, $k = 1, \ldots, d$, $x \geq 0$. Independently, let Π be a random permutation of $\{1, \ldots, d\}$ with $\mathbb{P}(\Pi = \pi) = 1/d!$ for each permutation π of $\{1, \ldots, d\}$, i.e. Π is uniformly distributed on the set of all $d!$ permutations. We consider the increasing sequence of random variables $T_1 < T_2 < \ldots < T_d$ defined by $T_k := E_1 + \ldots + E_k$. Then the (obviously exchangeable) random vector $\boldsymbol{X} = (X_1, \ldots, X_d) := (T_{\Pi(1)}, \ldots, T_{\Pi(d)})$ has density (3.9). If $E_1, E_2, \ldots$ is an arbitrary sequence of independent, absolutely continuous, positive random variables the counting process

$$N_t := \sum_{k \geq 1} 1_{\{E_1 + \ldots + E_k \leq t\}}, \quad t \geq 0,$$

is called non-homogeneous pure birth process *with intensity rate functions given by $r_k(x) := -\frac{\partial}{\partial x}\log\{\mathbb{P}(E_k > x)\}$, $k \geq 1$. A random permutation of the first d jump times $T_k := E_1 + \ldots + E_k$, $k = 1, \ldots, d$, of a pure birth process N thus has an ℓ_∞-norm symmetric density if the intensities $r_1, \ldots, r_{d-1}$ can be retrieved recursively from r_d via (3.15). The case of arbitrary intensities $r_1, \ldots, r_d$ hence provides a natural generalization of the family of ℓ_∞-norm symmetric densities. It appears to be an interesting open problem to determine necessary and sufficient conditions on $r_1, \ldots, r_d$ such that the respective exchangeable density is conditionally iid, see also paragraph 7.1 below.*

Example 3.4 (Pareto mixture of uniforms)**.** *Let M in Theorem 3.3 have survival function $\mathbb{P}(M > x) = \min\{1, x^{-\alpha}\}$ for some $\alpha > 0$. The associated function g_d generating the ℓ_∞-norm symmetric density is given by*

$$g_d(x) = \mathbb{E}[1_{\{M > x\}}\,M^{-d}] = \alpha \int_{\max\{x,1\}}^\infty u^{-d-1-\alpha}\,\mathrm{d}u = \frac{\alpha}{d+\alpha}\max\{1,x\}^{-d-\alpha}.$$

The components X_k of $\boldsymbol{X}$ have the following one-dimensional distribution function $G(x) := \mathbb{P}(X_k \leq x)$, and respective inverse G^{-1}, given by

$$G(x) = \begin{cases} \frac{\alpha}{1+\alpha}\,x & ,\ \text{if } x < 1, \\ 1 - \frac{1}{1+\alpha}\,x^{-\alpha} & ,\ \text{if } x \geq 1 \end{cases},$$

$$G^{-1}(y) = \begin{cases} \frac{1+\alpha}{\alpha}\, y & \text{, if } 0 < y < \frac{\alpha}{1+\alpha} \\ \left((1-y)\,(1+\alpha)\right)^{-\frac{1}{\alpha}} & \text{, if } \frac{\alpha}{1+\alpha} \le y < 1 \end{cases}.$$

This induces the one-parametric bivariate copula family defined by

$$C_\alpha(u_1, u_2) := \mathbb{P}\big(G(X_1) \le u_1, G(X_2) \le u_2\big)$$

$$= \begin{cases} \frac{(1+\alpha)^2}{\alpha\,(\alpha+2)}\, u_1\, u_2 & \text{, if } u_1,\, u_2 \le \frac{\alpha}{1+\alpha} \\ u_{[1]} - \frac{(1+\alpha)^{1+1/\alpha}}{2+\alpha}\, u_{[1]}\,(1 - u_{[2]})^{1+1/\alpha} & \text{, if } u_{[1]} \le \frac{\alpha}{1+\alpha} \le u_{[2]} \\ u_{[1]} - \frac{\alpha}{2+\alpha}\,(1 - u_{[1]})^{-1/\alpha}\,(1 - u_{[2]})^{1+1/\alpha} & \text{, else} \end{cases}.$$

Scatter plots from this copula for different values of α are depicted in Figure 3, visualizing the dependence structure behind pairs of $\boldsymbol{X}$. The dependence decreases with α, and the limiting cases $\alpha = 0$ and $\alpha = \infty$ correspond to perfect positive association and independence, respectively. One furthermore observes that the dependence is highly asymmetric, i.e. large values of $G(X_1), G(X_2)$ are more likely jointly close to each other than small values, which behave like independence. This effect can be quantified in terms of the so-called upper- and lower-tail dependence coefficients, given by

$$\lim_{x \to \infty} \mathbb{P}(X_1 > x \,|\, X_2 > x) = \frac{2}{2+\alpha}, \quad \lim_{x \searrow 0} \mathbb{P}(X_1 \le x \,|\, X_2 \le x) = 0,$$

respectively.

4. The multivariate lack-of-memory property

A random vector $\boldsymbol{X} = (X_1, \ldots, X_d)$ with non-negative components is said to satisfy the *(multivariate) lack-of-memory property* if for arbitrary $1 \le i_1 < \ldots < i_n \le d$ we have that

$$\mathbb{P}(X_{i_1} > t_{i_1} + t, \ldots, X_{i_n} > t_{i_n} + t \,|\, X_{i_1} > t, \ldots, X_{i_n} > t)$$
$$= \mathbb{P}(X_{i_1} > t_{i_1}, \ldots, X_{i_n} > t_{i_n}),$$

with the $t, t_1, \ldots, t_d$ either in $(0, \infty)$ (continuous support case) or in $\mathbb{N}_0$ (discrete support case). The lack-of-memory property is very intuitive when the k-th component X_k of $\boldsymbol{X}$ is interpreted as the future time point at which the k-th component in a system of d components fails. In words, it means that conditioned on the survival of an arbitrary sub-system $(i_1, \ldots, i_n)$ until time t, the residual lifetimes of the components $i_1, \ldots, i_n$ are identical in distribution to the lifetimes at inception of the system. Needless to mention that such intuitive property occupies a commanding role in reliability theory, see [6] for a textbook treatment, but is also important in other contexts such as financial risk management, e.g., [28, 62]. An alternative way to formulate the multivariate lack-of-memory property, due to [13], is the following. For $1 \le i_1 < \ldots < i_n \le d$ we denote by
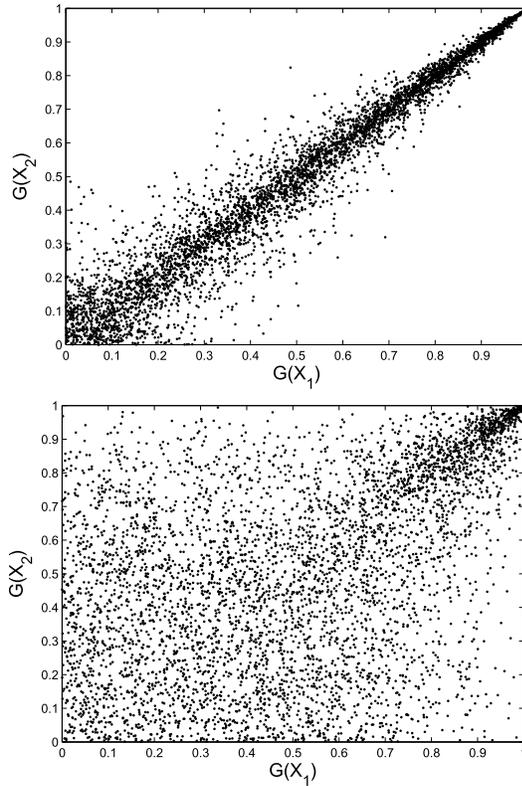
Fig 3. *Top:* 5000 *samples of* $(G(X_1), G(X_2))$ *for* $\alpha = 0.1$ *in Example* 3.4. *Bottom:* 5000 *samples of* $(G(X_1), G(X_2))$ *for* $\alpha = 1$ *in Example* 3.4.

$Z_{i_1,\ldots,i_n}(t) := (1_{\{X_{i_1} > t\}}, \ldots, 1_{\{X_{i_n} > t\}})$, $t \geq 0$, the stochastic process which indicates for each of the n components $i_1, \ldots, i_n$ whether it is still working or already dysfunctional. The random vector $\boldsymbol{X}$ has the lack-of-memory property if and only if $Z_{i_1,\ldots,i_n}$ is a continuous-time Markov chain for all $1 \leq i_1 < \ldots < i_n \leq d$.

From a theoretical point of view, studying the (multivariate) lack-of-memory property is also natural as it generalizes very popular one-dimensional probability distributions to the multivariate case. Indeed, if $d = 1$ we abbreviate $X := X_1$ and recall the following classical characterizations.

Lemma 4.1 (Characterization of lack-of-memory for $d = 1$)**.**

- *($\mathcal{E}$)* *If the support of* X *equals* $[0, \infty)$, *then* X *satisfies the lack-of-memory property if and only if* X *has an exponential distribution, that is* $\mathbb{P}(X > t) = \exp(-\lambda t)$ *for some* $\lambda > 0$.
- *($\mathcal{G}$)* *If the support of* X *equals* $\mathbb{N}$, *then* X *satisfies the lack-of-memory property if and only if* X *has a geometric distribution, that is* $\mathbb{P}(X > n) = (1-p)^n$ *for some* $p \in (0, 1)$.

Proof. In the geometric case, inductively we see that $\bar{F}(n) := \mathbb{P}(X > n)$ satisfies $\bar{F}(n) = \bar{F}(1)^n$, $n \in \mathbb{N}_0$, and the claim follows with $p := 1 - \bar{F}(1)$. Notice that $\bar{F}(1) \in \{0, 1\}$ is ruled out by the assumption of support equal to $\mathbb{N}$. The exponential case follows similarly, see [12, p. 190].   □

4.1. Marshall-Olkin and multivariate geometric distributions

The well known characterizations of univariate lack-of-memory in Lemma 4.1 have been lifted to the multivariate case in [79] and [4, 77], respectively, which is briefly recalled. First of all, we introduce the multivariate exponential models of [79] and [4]. To this end, we denote by $\mathcal{E}(\lambda)$ the univariate exponential law with rate $\lambda > 0$, and by $\mathcal{G}(p)$ the univariate geometric distribution with parameter $p \in (0, 1)$, i.e. with survival function $\mathbb{P}(X > n) = (1 - p)^n$. In order to include boundary cases, we denote by $\mathcal{E}(0), \mathcal{G}(0)$ the probability law of a degenerate random variable that is identically equal to infinity, and by $\mathcal{G}(1)$ the probability law of a degenerate random variable that is identically equal to one.

Example 4.1 (Probability laws with multivariate lack-of-memory).

($\mathcal{E}$) *For each non-empty $I \subset \{1, \ldots, d\}$ let $\lambda_I \geq 0$ with $\sum_{I\,:\,k \in I} \lambda_I > 0$ for each $k = 1, \ldots, d$. With $E_I \sim \mathcal{E}(\lambda_I)$ a list of $2^d - 1$ independent random variables, we define $\boldsymbol{X}$ via*

$$X_k := \min\{E_I \,:\, k \in I\}, \quad k = 1, \ldots, d.$$

Then $\boldsymbol{X}$ satisfies the multivariate lack-of-memory property, which is easy to see while noticing that the survival function of $\boldsymbol{X}$ equals

$$\bar{F}(\boldsymbol{x}) = \mathbb{P}(\boldsymbol{X} > \boldsymbol{x}) = \exp\Big\{ - \sum_{\emptyset \neq I \subset \{1,\ldots,d\}} \lambda_I \max_{k \in I}\{x_k\} \Big\}.$$

($\mathcal{G}$) *For each (possibly empty) $I \subset \{1, \ldots, d\}$ let $p_I \in [0, 1]$ with $\sum_{I\,:\,k \notin I} p_I < 1$ for each $k = 1, \ldots, d$ and $\sum_I p_I = 1$. The probabilities p_I define a probability law on the power set of $\{1, \ldots, d\}$. Let $S_1, S_2, \ldots$ be an iid sequence drawn from this law and denote by G_I the smallest $n \in \mathbb{N}$ such that $S_n = I$. Notice that $G_I \sim \mathcal{G}(p_I)$. We define the random vector $\boldsymbol{X}$ with values in $\mathbb{N}^d$ by*

$$X_k := \min\{G_I \,:\, k \in I\}, \quad k = 1, \ldots, d.$$

Then $\boldsymbol{X}$ satisfies the multivariate lack-of-memory property. Furthermore, the survival function of $\boldsymbol{X}$ equals

$$\bar{F}(\boldsymbol{n}) = \mathbb{P}(\boldsymbol{X} > \boldsymbol{n}) = \prod_{k=1}^d \Big(\sum_{I\,:\,\{\sigma_{\boldsymbol{n}}(k),\ldots,\sigma_{\boldsymbol{n}}(d)\} \cap I = \emptyset} p_I \Big)^{n_{[k]} - n_{[k-1]}},$$

where $\sigma_{\boldsymbol{n}}$ denotes a permutation of $\{1, \ldots, d\}$ such that $n_{\sigma_{\boldsymbol{n}}(1)} \leq \ldots \leq n_{\sigma_{\boldsymbol{n}}(d)}$, and $n_0 := 0$, for $\boldsymbol{n} = (n_1, \ldots, n_d) \in \mathbb{N}_0^d$.

The probability distribution in part $(\mathcal{E})$ of Example 4.1 is called *Marshall-Olkin distribution*. It is named after [79]. The probability distribution in part $(\mathcal{G})$ of Example 4.1 is called *wide-sense geometric distribution*. The stochastic model has been introduced in [4]. The presented form of the survival function is computed in [77]. The following lemma shows that the multivariate stochastic models in Example 4.1 define precisely the multivariate analogues of the univariate exponential and geometric laws, when defined via the lack-of-memory property. Thus, it constitutes a multivariate extension of Lemma 4.1.

Lemma 4.2 (Characterization of lack-of-memory for $d \geq 1$)**.**

- $(\mathcal{E})$ *The d-variate Marshall-Olkin distribution is the only probability law with support $[0, \infty)^d$ satisfying the lack-of-memory property.*
- $(\mathcal{G})$ *The d-variate wide-sense geometric distribution is the only probability law with support $\mathbb{N}^d$ satisfying the lack-of-memory property.*

Proof. Part $(\mathcal{E})$ is due to the original reference [79], while part $(\mathcal{G})$ is shown in [77]. □

Example 4.2 (Narrow-sense geometric law)**.** *If $\boldsymbol{Y}$ has a Marshall-Olkin distribution and we define $\boldsymbol{X} := (\lceil Y_1 \rceil, \ldots, \lceil Y_d \rceil)$, then $\boldsymbol{X}$ is said to have the* narrow-sense geometric distribution. *As the nomenclature suggests, the narrow-sense geometric distribution is a subset of the wide-sense geometric distribution in dimensions $d \geq 2$ (and identical for $d = 1$), which is very easy to see by the characterizing lack-of-memory property of the Marshall-Olkin law. Not every wide-sense geometric law can be constructed like this, i.e. the narrow-sense family defines a proper subset of the wide-sense family. This indicates that for $d \geq 2$ the structure of the discrete lack-of-memory property is more delicate than the structure of its continuous counterpart. For example, while two components of a random vector with Marshall-Olkin distribution or narrow-sense geometric distribution cannot be negatively correlated, two components of a random vector with wide-sense geometric distribution can be, see [77] for details.*

4.2. Infinite divisibility and Lévy subordinators

The concept of infinite divisibility is of fundamental importance in the present section, but also in Sections 5 and 6 below. Thus, we briefly recall the required background in the present paragraph. For an elaborate textbook treatment we refer to [94]. The concept of a Lévy subordinator plays an essential role when studying the conditionally iid subfamily of the Marshall-Olkin distribution, a result first discovered in [72]. Recall that a càdlàg stochastic process $Z = \{Z_t\}_{t \geq 0}$ with $Z_0 = 0$ is called a *Lévy process* if it has stationary and independent increments, which means that:

- (i) The law of $Z_{t+h} - Z_t$ is independent of $t \geq 0$ for each $h \geq 0$, i.e. $Z_{t+h} - Z_t \stackrel{d}{=} Z_h$.
- (ii) $Z_{t_2} - Z_{t_1}, \ldots, Z_{t_n} - Z_{t_{n-1}}$ are independent for $0 \leq t_1 < \ldots < t_n$.

Hence, Lévy processes are the continuous-time equivalents of discrete-time random walks. A non-decreasing Lévy process is called a *Lévy subordinator*. However, there is one fundamental difference between a random walk and a Lévy process: the probability law of the increments in a random walk is arbitrary on $\mathbb{R}$, whereas the law of the increments in a Lévy process needs to satisfy a certain compatibility condition with respect to time, as increments of arbitrarily large time span can be considered. Concretely, it is immediate from the definition of a Lévy process $Z = \{Z_t\}_{t \geq 0}$ that the probability law of Z_1 is infinitely divisible. Recall that a random variable X is called *infinitely divisible* if for each $n \in \mathbb{N}$ there exist iid random variables $X_1^{(n)}, \ldots, X_n^{(n)}$ such that $X \overset{d}{=} X_1^{(n)} + \ldots + X_n^{(n)}$. Furthermore, if X has an infinitely divisible probability law, there exists a Lévy process $Z = \{Z_t\}_{t \geq 0}$, which is uniquely determined in law, such that $Z_1 \overset{d}{=} X$. As a consequence, a Lévy subordinator Z is uniquely determined in distribution by the law of Z_1, or analytically by the function $\Psi(x) := -\log(\mathbb{E}[\exp(-x\,Z_1)])$, $x \geq 0$. One calls Ψ the *Laplace exponent* of the infinitely divisible random variable Z_1 (or of the Lévy subordinator Z). The function Ψ is a so-called *Bernstein function*, which means that it is infinitely often differentiable on $(0, \infty)$ and the derivative $\Psi^{(1)}$ is completely monotone, i.e. $(-1)^{k+1}\,\Psi^{(k)} \geq 0$ for all $k \geq 1$, see [8, 96] for textbook treatments on the topic. The value $\Psi(0)$ is by definition equal to zero but we might have a jump at zero meaning that $\Psi(x) > \epsilon > 0$ for all $x > 0$ is possible. Intuitively, this is the case if and only if $\mathbb{P}(Z_t = \infty) > 0$ for $t > 0$, and in this case one sometimes also speaks of a *killed* Lévy subordinator.

4.3. Analytical characterization of exchangeability and conditionally iid

By Lemma 1.1 a random vector $\boldsymbol{X}$ with either Marshall-Olkin distribution or wide-sense geometric distribution can only be conditionally iid if it is exchangeable. An elementary computation shows that the Marshall-Olkin distribution (resp. wide-sense geometric distribution) is exchangeable if and only if its parameters λ_I (resp. p_I) depend on the indexing subsets I only through their cardinality $|I|$. In this exchangeable case, we denote these parameters by $\lambda_1, \ldots, \lambda_d$ (resp. $p_0, p_1, \ldots, p_d$), with subindices denoting the possible cardinalities, i.e. $\lambda_k := \lambda_{\{1,\ldots,k\}}$ and $p_k := p_{\{1,\ldots,k\}}$, and combinatorial computations show that the survival function $\bar{F}$ of $\boldsymbol{X}$ takes the convenient algebraic form

$$\bar{F}(\boldsymbol{x}) = \prod_{k=1}^{d} b_k^{x_{[d-k+1]} - x_{[d-k]}}, \quad \bar{F}(\boldsymbol{n}) = \prod_{k=1}^{d} b_k^{n_{[d-k+1]} - n_{[d-k]}}, \tag{4.1}$$

for either $x_1, \ldots, x_d \in [0, \infty)$ with $x_0 := 0$ (in the Marshall-Olkin case) or $n_1, \ldots, n_d \in \mathbb{N}_0$ with $n_0 := 0$ (in the wide-sense geometric case), and with[13]

$$b_k := \prod_{i=1}^{k} \exp\left\{ -\sum_{j=0}^{d-i} \binom{d-i}{j} \lambda_{j+1} \right\}, \qquad \text{(Marshall-Olkin case)} \tag{4.2}$$

---

[13]The empty product is conveniently defined to be equal to one, i.e $\prod_{i=1}^{0} := 1$.

$$b_k := \sum_{i=0}^{d-k} \binom{d-k}{i} p_i, \qquad \text{(wide-sense geometric case)}, \quad (4.3)$$

for $k = 0, \ldots, d$. While the parameters λ_k (resp. p_k) are intuitive since they allow for the probabilistic interpretations according to Example 4.1, the re-parameterization in terms of the new parameters b_k is more convenient with regards to finding an answer to the question: when is $\boldsymbol{X}$ conditionally iid? The main result in this regard is stated in Theorem 4.1 below, which requires the notion of d-monotone sequences and log-d-monotone sequences. The concept of d-monotonicity as well as the notations $\mathcal{M}_d$ and $\mathcal{M}_\infty$ have already been introduced in paragraph 2.1, the related concept of log-d-monotonicity is introduced in the following definition.

Definition 4.1 (Log- monotone sequences). *For $d \in \mathbb{N}$, a finite sequence $(b_0, b_1, \ldots, b_d) \in (0, \infty)^{d+1}$ is said to be log-d-monotone if $\nabla^{d-k} \log(b_k) \geq 0$ for $k = 0, 1, \ldots, d-1$. An infinite sequence $\{b_k\}_{k \in \mathbb{N}_0}$ with positive members is said to be completely log-monotone if $(b_0, \ldots, b_d)$ is log-d-monotone for each $d \geq 1$.*

The notion of a log-d-monotone sequence is less intuitive than that of a d-monotone sequence. First notice that, in contrast to the definition of a d-monotone sequence in paragraph 2.1, $\log(b_d) \geq 0$ needs not hold for a log-d-monotone sequence, which is explained by the following useful relationship between $(d-1)$-monotonicity and log-d-monotonicity. It helps to transform statements involving log-d-monotonicity into statements involving only the simpler notion of $(d-1)$-monotonicity[14]:

$$(b_0, \ldots, b_{d-1}) \text{ is } (d-1)\text{-monotone}$$
$$\Leftrightarrow \left(1, e^{-b_0}, e^{-(b_0+b_1)}, \ldots, e^{-\sum_{i=0}^{d-1} b_i}\right) \text{ is log-}d\text{-monotone}. \qquad (4.4)$$

The set of all log-d-monotone sequences starting with $b_0 = 1$ will be denoted by $\mathcal{LM}_d$ in the following. Similarly, $\mathcal{LM}_\infty$ denotes the sets of completely log-monotone sequences starting with $b_0 = 1$. [77, Proposition 4.4] shows that $\{b_k\}_{k \in \mathbb{N}} \in \mathcal{LM}_\infty$ if and only if $\{b_k^t\}_{k \in \mathbb{N}} \in \mathcal{M}_\infty$ for arbitrary $t > 0$. In particular, $\mathcal{LM}_\infty \subset \mathcal{M}_\infty$. Theorem 4.1 below provides a second result, besides Theorem 2.1, showing that whether or not a (log-) d-monotone sequence can be extended to a completely (log-) monotone sequence plays an important role in the context of the present survey.

In order to better understand the following theorem it is helpful to know that the Laplace exponent Ψ of a Lévy subordinator Z is already completely determined by its values on $\mathbb{N}$, i.e. by the sequence $\{\Psi(k)\}_{k \in \mathbb{N}_0}$. Furthermore, the sequence $\{\exp(-\Psi(k))\}_{k \in \mathbb{N}_0}$ equals the moment sequence of the random variable $\exp(-Z_1)$, so lies in $\mathcal{M}_\infty$ by the little moment problem, see paragraph 2.1. Since

---

[14]This statement simply follows from the fact that $\log(1) = 0$ and $\nabla^{d-k-1} b_k = \nabla^{d-k} \log(\tilde{b}_k)$ with $\tilde{b}_k := \exp(-\sum_{i=0}^{k-1} b_i)$ for $k = 0, \ldots, d$, with an empty sum being conveniently defined as zero.

for arbitrary $t > 0$ even the sequence $\{\exp(-t\,\Psi(k))\}_{k\in\mathbb{N}_0}$ lies in $\mathcal{M}_\infty$ as the moment sequence of $\exp(-Z_t)$, the sequence $\{\exp(-\Psi(k))\}_{k\in\mathbb{N}_0}$ even lies in the smaller set $\mathcal{LM}_\infty$ of completely log-monotone sequences. The subset $\mathcal{LM}_\infty \subsetneq \mathcal{M}_\infty$ corresponds to precisely the infinitely divisible laws on $[0,\infty]$, which is the discrete analogue of the well known statement that $\exp(-t\,\Psi)$ is a completely monotone function for arbitrary $t > 0$ if and only if $\Psi^{(1)}$ is completely monotone. With this information and the information of paragraph 4.2 as background the following theorem is now quite intuitive.

Theorem 4.1 solves Problem 1.3 for the property (P) of "satisfying the multivariate lack-of-memory property".

Theorem 4.1 (Lack-of-memory, exchangeability and conditionally iid)**.**

($\mathcal{E}$) *The function (4.1) is a survival function (of some $\boldsymbol{X}$) with support $[0,\infty)^d$ if and only if we have $(b_0,\ldots,b_d) \in \mathcal{LM}_d$. Furthermore, the associated exchangeable Marshall-Olkin distribution admits a stochastic representation that is conditionally iid if there exist $b_{d+1}, b_{d+2}, \ldots$ such that $\{b_k\}_{k\in\mathbb{N}_0} \in \mathcal{LM}_\infty$. To wit, in this case there exists a (possibly killed) Lévy subordinator $Z = \{Z_t\}_{t\geq 0}$, determined in law via*

$$b_k := \mathbb{E}\Big[e^{-k\,Z_1}\Big], \quad k \in \mathbb{N}_0, \tag{4.5}$$

such that $\boldsymbol{X}$ has the same distribution as the vector defined in (1.5).

($\mathcal{G}$) *The function (4.1) is a survival function (of some $\boldsymbol{X}$) with support $\mathbb{N}^d$ if and only if we have $(b_0, b_1, \ldots, b_d) \in \mathcal{M}_d$. Furthermore, the associated exchangeable wide-sense geometric distribution admits a stochastic representation that is conditionally iid if there exist $b_{d+1}, b_{d+2}, \ldots$ such that $\{b_k\}_{k\in\mathbb{N}_0} \in \mathcal{M}_\infty$. To wit, in this case there exists an iid sequence $Y_1, Y_2, \ldots$ of random variables taking values in $[0,\infty]$, determined in law via*

$$b_k := \mathbb{E}\Big[e^{-k\,Y_1}\Big], \quad k \in \mathbb{N}_0, \tag{4.6}$$

such that $\boldsymbol{X}$ has the same distribution as the vector defined in (1.5) when

$$Z_t := Y_1 + Y_2 + \ldots + Y_{\lfloor t\rfloor}, \quad t \geq 0.$$

Proof. Part ($\mathcal{E}$) is due to [71, 72], while part ($\mathcal{G}$) is due to [77].

First, we observe that once the correspondence between $\mathcal{M}_d$ and the wide-sense geometric law is established, the correspondence between $\mathcal{LM}_d$ and the narrow-sense geometric law (or, algebraically equivalent, its continuous counterpart the Marshall-Olkin law) follows from (4.4) together with (4.2) and (4.3). This is because the λ_j in (4.2) are arbitrary non-negative numbers, and the p_i in (4.3) are also arbitrary non-negative up to scaling (i.e. with an additional scale factor $c > 0$ we have that $c\,(p_0, \ldots, p_{d-1})$ and $(\lambda_1, \ldots, \lambda_d)$ both run through all of $[0,\infty)^d \setminus \{(0, \ldots, 0)\}$, noticing that p_d is determined by $p_0, \ldots, p_{d-1}$). Concretely, by the correspondence between $\mathcal{M}_d$ and the wide-sense geometric law, we obtain a correspondence between $\mathcal{M}_d$ and $[0,\infty)^d \setminus \{(0, \ldots, 0)\}$ up to scaling

in (4.3). In particular, the property of being d-monotone is not affected by c. Replacing the λ_j in (4.2) by $c\,p_{j-1}$ and making use of (4.4), we then end up with the correspondence between $\mathcal{LM}_d$ and the Marshall-Olkin law. To establish the correspondence between $\mathcal{M}_d$ and the wide-sense geometric law is really only a tedious algebraic computation, see [77] for details. Essentially, d-monotonicity enters the scene precisely for the same reason as in paragraph 2.1.

Regarding the conditionally iid subfamliy, the crucial insight is that $\mathcal{M}_\infty$ stands in one-to-one relation with the set of probability measures on $[0, \infty]$ via (4.6), which is exactly the well-known statement of the little moment problem, only formulated for the compact interval $[0, \infty]$ instead of the more usual interval $[0, 1]$ via the transformation $-\log$. That the (discrete) random walk construction in part $(\mathcal{G})$ can only be "made continuous" in case Y_1 is infinitely divisible is very intuitive, and the Lévy subordinator in part $(\mathcal{E})$ is simply the continuous analogue of the discrete random walk in that case. □

Since the narrow-sense geometric law of Example 4.2 is a special case of the wide-sense geometric law, it follows that $\mathcal{LM}_d \subsetneq \mathcal{M}_d$, which in fact is not an obvious statement. Furthermore, $\boldsymbol{X}$ in part $(\mathcal{G})$ of Theorem 4.1 happens to be narrow-sense geometric if and only if the random variable Y_1 is infinitely divisible. In fact, the elements of $\mathcal{LM}_\infty$ stand in one-to-one correspondence with the family of infinitely divisible laws on $[0, \infty]$ via (4.5), whereas the elements of the larger set $\mathcal{M}_\infty$ stand in one-to-one correspondence with the family of arbitrary probability laws on $[0, \infty]$ via (4.6), which is just a slight re-formulation of the little moment problem.

Remark 4.1 (Analytical criterion for conditionally iid)*. Given an exchangeable random vector $\boldsymbol{X}$ with lack-of-memory property and parameters $(b_0, \ldots, b_d)$, Theorem 4.1 implies that $\boldsymbol{X}$ has a stochastic representation that is conditionally iid if $(b_0, \ldots, b_d)$ can be extended to a completely (log-) monotone sequence. Using (4.4), an element $(b_0, \ldots, b_d) \in \mathcal{LM}_d$ is extendible to an element in $\mathcal{LM}_\infty$ if and only if the $(d-1)$-monotone sequence $(-\log(b_1/b_0), \ldots, -\log(b_d/b_{d-1}))$ is extendible to a completely monotone sequence. Thus, we can concentrate on the completely monotone case. Deciding whether a d-monotone sequence can be extended to a completely monotone sequence is the truncated Hausdorff moment problem again, see Section 2.1. This means that an effective analytical criterion for extendibility is known.*

The following example demonstrates how a parameter sequence $\{b_k\}_{k \in \mathbb{N}_0}$ for some wide-sense geometric law is conveniently defined via the link to the little moment problem, setting $b_k := \mathbb{E}[X^k]$, $k \in \mathbb{N}_0$, where X is some arbitrary random variable taking values in $[0, 1]$.

Example 4.3 (A two-parametric family based on the Beta distribution)*. Consider a random variable X with density*

$$f_X(x) = \frac{\Gamma(p+q)}{\Gamma(p)\,\Gamma(q)}\,x^{p-1}\,(1-x)^{q-1}, \quad 0 < x < 1,$$

with parameters $p, q > 0$, which is a Beta distribution. The moment sequence is known to be[15]

$$\mathbb{E}[X^k] = \int_0^1 f_X(x)\, x^k \,\mathrm{d}x = \frac{\Gamma(p+k)\,\Gamma(p+q)}{\Gamma(p)\,\Gamma(p+q+k)}, \quad k \in \mathbb{N}_0,$$

so that a two-parametric family of d-variate wide-sense geometric survival functions (for arbitrary $d \geq 1$) is given by

$$\bar{F}_{p,q}(\boldsymbol{n}) = \Big(\frac{\Gamma(p+q)}{\Gamma(p)}\Big)^{n_{[d]}} \prod_{k=1}^{d} \Big(\frac{\Gamma(p+k)}{\Gamma(p+q+k)}\Big)^{n_{[d-k+1]} - n_{[d-k]}}, \quad \boldsymbol{n} \in \mathbb{N}_0^d.$$

The associated probability distribution of Y_1 in Theorem 4.1($\mathcal{G}$) is given by $Y_1 \stackrel{d}{=} -\log(X)$, i.e. the logarithm of the reciprocal of the Beta distribution in concern. Similarly, making use of (4.4), a two-parametric family of d-variate Marshall-Olkin survival functions (for arbitrary $d \geq 1$) is given by

$$\bar{F}_{p,q}(\boldsymbol{x}) = \exp\Big\{ -\frac{\Gamma(p+q)}{\Gamma(p)} \sum_{k=1}^{d}(x_{[d-k+1]} - x_{[d-k]}) \sum_{i=0}^{k-1} \frac{\Gamma(p+i)}{\Gamma(p+q+i)} \Big\}$$

$$= \exp\Big\{ -\frac{\Gamma(p+q)}{\Gamma(p)} \sum_{k=1}^{d} \frac{\Gamma(p+k-1)}{\Gamma(p+q+k-1)} x_{[d-k+1]} \Big\}, \quad \boldsymbol{x} \in [0,\infty)^d.$$

In the special case when $q = 2$, the Lévy subordinator in Theorem 4.1($\mathcal{E}$) is of compound Poisson type with intensity $p + 1$ and jumps that are exponentially distributed with parameter p.

5. Max-/ min-stable laws and extreme-value copulas

Throughout this paragraph, for the sake of a more compact notation we implicitly make excessive use of the abbreviations $f(0) := \lim_{x \searrow 0} f(x)$ and $f(\infty) := \lim_{x \to \infty} f(x)$ for functions $f : (0,\infty) \to (0,\infty)$, provided the respective limits exist in $[0,\infty]$.

5.1. Max-/ min-stability and multivariate extreme-value theory

Definition 5.1 (Max- and min-stability)**.** *We denote by F (resp. $\bar{F}$) the d-variate distribution function (resp. survival function) of some d-dimensional random vector $\boldsymbol{Y} = (Y_1, \ldots, Y_d)$ (resp. $\boldsymbol{X} = (X_1, \ldots, X_d)$).*

(a) *(The probability law of) $\boldsymbol{Y}$ is said to be* max-stable *if for arbitrary $t > 0$ there are $\alpha_i(t) > 0$, $\beta_i(t) \in \mathbb{R}$ such that*

$$F(\boldsymbol{x})^t = F\big(\alpha_1(t)\, x_1 + \beta_1(t), \ldots, \alpha_d(t)\, x_d + \beta_d(t)\big).$$

---

[15]See, e.g., [29, p. 35].

In this case, we also say that F is max-stable. *In words, F^t is again a distribution function and equals F modulo a linear transformation of its arguments.*

(b) *(The probability law of) $\boldsymbol{X}$ is said to be* min-stable *if for arbitrary $t > 0$ there are $\alpha_i(t) > 0$, $\beta_i(t) \in \mathbb{R}$ such that*

$$\bar{F}(\boldsymbol{x})^t = \bar{F}\big(\alpha_1(t)\, x_1 + \beta_1(t), \dots, \alpha_d(t)\, x_d + \beta_d(t)\big).$$

In this case, we also say that $\bar{F}$ is min-stable. *In words, $\bar{F}^t$ is again a survival function and equals $\bar{F}$ modulo a linear transformation of its arguments.*

If $\boldsymbol{Y}$ is max-stable and $\boldsymbol{Y}^{(i)}$ are independent copies of $\boldsymbol{Y}$, then for arbitrary $n \in \mathbb{N}$ we observe

$$\boldsymbol{Y} \stackrel{d}{=} \Big(\frac{\max_{i=1}^n \{Y_1^{(i)}\} - \beta_1(1/n)}{\alpha_1(1/n)}, \dots, \frac{\max_{i=1}^n \{Y_d^{(i)}\} - \beta_d(1/n)}{\alpha_d(1/n)} \Big).$$

Similarly, if $\boldsymbol{X}$ is min-stable this means

$$\boldsymbol{X} \stackrel{d}{=} \Big(\frac{\min_{i=1}^n \{X_1^{(i)}\} - \beta_1(1/n)}{\alpha_1(1/n)}, \dots, \frac{\min_{i=1}^n \{X_d^{(i)}\} - \beta_d(1/n)}{\alpha_d(1/n)} \Big).$$

In words, the component-wise re-scaled maxima of iid copies of $\boldsymbol{Y}$ (resp. minima of iid copies of $\boldsymbol{X}$) have the same distribution as $\boldsymbol{Y}$ (resp. $\boldsymbol{X}$).

Max and min-stability play a central role in multivariate extreme-value theory, as will briefly be explained. If $\boldsymbol{V}^{(i)}$ are independent copies of some random vector $\boldsymbol{V} = (V_1, \dots, V_d)$, one is interested in the probability law of the vectors of component-wise maxima, that is

$$\big(\max_{i=1}^n \{V_1^{(i)}\}, \dots, \max_{i=1}^n \{V_d^{(i)}\} \big), \quad n \in \mathbb{N}.$$

If one can find sequences $\alpha_1(n), \dots, \alpha_d(n) > 0$ and $\beta_1(n), \dots, \beta_d(n) \in \mathbb{R}$ such that the re-scaled vector

$$\Big(\frac{\max_{i=1}^n \{V_1^{(i)}\} - \beta_1(n)}{\alpha_1(n)}, \dots, \frac{\max_{i=1}^n \{V_d^{(i)}\} - \beta_d(n)}{\alpha_d(n)} \Big)$$

converges in distribution to some $\boldsymbol{Y} = (Y_1, \dots, Y_d)$ as $n \to \infty$, then one says that $\boldsymbol{Y}$ has a *multivariate extreme-value distribution*. A classical result in multivariate extreme-value theory states that $\boldsymbol{Y}$ has a multivariate extreme-value distribution if and only if $\boldsymbol{Y}$ is max-stable, see, e.g., [50, pp. 172-174].

Since $\boldsymbol{Y}$ is max-stable if and only if $-\boldsymbol{Y}$ is min-stable (obviously), max- and min-stability can be studied jointly by focusing on one of the two concepts. Classical extreme-value theory textbooks typically focus on max-stability and further subdivide the study of the probability law of max-stable $\boldsymbol{Y}$ into two sub-studies:

(i) By the Fisher-Tippett-Gnedenko Theorem, the univariate distribution function F_k of each component Y_k necessarily belongs to either the Gumbel, the Fréchet or the Weibull family, see [7, Chapter 2, p. 45 ff] for background.

(ii) Having understood the univariate distribution functions $F_1, \ldots, F_d$ of the components according to (i), the distribution function F of $\boldsymbol{Y}$ necessarily takes the form

$$F(\boldsymbol{x}) = C\big(F_1(x_1), \ldots, F_d(x_d)\big),$$

for a copula $C : [0,1]^d \to [0,1]$ with the characterizing property that $C(\boldsymbol{u})^t = C(u_1^t, \ldots, u_d^t)$ for each $t > 0$, a so-called *extreme-value copula*.

In order to focus on a deeper understanding of extreme-value copulas it is convenient to normalize the margins $F_1, \ldots, F_d$. In classical extreme-value theory, it is standard to normalize to standardized Fréchet distributions, i.e. $F_k(x) = \exp(-\lambda_k/x)\,1_{\{x>0\}}$ for some $\lambda_k > 0$. Furthermore, we observe that $\boldsymbol{X} := (1/Y_1, \ldots, 1/Y_d)$ is well-defined, X_k is exponential with rate λ_k, and $\boldsymbol{X}$ is min-stable (since $x \mapsto 1/x$ is strictly decreasing, so max-stability of $\boldsymbol{Y}$ is flipped to min-stability of $\boldsymbol{X}$). The vector $\boldsymbol{X}$ is thus called *min-stable multivariate exponential* and has survival function

$$\bar{F}(\boldsymbol{x}) = \mathbb{P}(\boldsymbol{X} > \boldsymbol{x}) = \mathbb{P}(\boldsymbol{Y} < 1/\boldsymbol{x}) = C\big(e^{-\lambda_1 x_1}, \ldots, e^{-\lambda_d x_d}\big),$$

with extreme-value copula C. The survival function $\bar{F}$ is min-stable, satisfying

$$\bar{F}(\boldsymbol{x})^t = \bar{F}(t\,\boldsymbol{x}), \quad t > 0. \tag{5.1}$$

The analytical property (5.1) characterizes the concept of min-stable multivariate exponentiality on the level of survival functions, and serves as a convenient starting point to study the conditionally iid subfamily (of extreme-value copulas, resp. min-stable multivariate exponential distributions). For a given extreme-value copula C it further turns out convenient to consider its so-called *stable tail dependence function*

$$\ell(\boldsymbol{x}) := -\log\Big(C\big(e^{-x_1}, \ldots, e^{-x_d}\big)\Big), \quad \boldsymbol{x} \in [0, \infty)^d,$$

which satisfies $\ell(t\,\boldsymbol{x}) = t\,\ell(\boldsymbol{x})$. Clearly, ℓ determines C and C determines ℓ, so that investigating ℓ instead of C is just a matter of convenience. Wrapping up, a min-stable multivariate exponential distribution is fully determined by the rates $(\lambda_1, \ldots, \lambda_d)$ specifying the one-dimensional exponential margins, and by a stable tail dependence function ℓ which stands in a one-to-one relationship with the associated extreme-value copula C.

5.2. *Analytical characterization of conditionally iid*

In the sequel, we are interested in the question: when is a min-stable multivariate exponential vector $\boldsymbol{X}$, i.e. one whose survival function satisfies (5.1), conditionally iid? We start with two important examples.

Example 5.1 (Independent exponentials)**.** *If the components $X_1, \ldots, X_d$ of $\boldsymbol{X}$ are iid, then we only need to consider the law of X_1. By definition, X_1 must have an exponential law, so there is some $\lambda > 0$ such that for arbitrary $t > 0$ we have*

$$\bar{F}(\boldsymbol{x})^t = \Big(\prod_{k=1}^d e^{-\lambda x_k} \Big)^t = e^{-t\lambda \sum_{k=1}^d x_k} = \prod_{k=1}^d e^{-\lambda t x_k} = \bar{F}(t\,\boldsymbol{x}).$$

Consequently, $\boldsymbol{X}$ is min-stable multivariate exponential. The associated stable tail dependence function is $\ell(\boldsymbol{x}) = x_1 + \ldots + x_d = \|\boldsymbol{x}\|_1$.

For arbitrary $c \geq 0$ we introduce the notation $\mathfrak{H}_{+,c} \subset \mathfrak{H}_+$ for distribution functions of non-negative random variables with mean equal to c. For $G \in \mathfrak{H}_+$ we further denote by $M_G := \int_0^\infty 1 - G(x)\,\mathrm{d}x \in [0, \infty]$ its mean.

Example 5.2 (An important semi-parametric family)**.** *Let $G \in \mathfrak{H}_+$ with $0 < M_G < \infty$. With an iid sequence of unit exponentials $\eta_1, \eta_2, \ldots$ we consider the stochastic process*

$$Z_t := \sum_{n \geq 1} -\log \Big\{ G\Big(\frac{\eta_1 + \ldots + \eta_n}{t} - \Big) \Big\}, \quad t \geq 0,$$

taking values in $[0, \infty]$. It is not difficult to see that $H := 1 - \exp(-Z)$ takes values in $\mathfrak{H}_+$. Consequently, we may define a conditionally iid random vector $\boldsymbol{X}$ via the canonical stochastic model (1.5) from this process H. Conditioned on H, the components of $\boldsymbol{X}$ are iid with distribution function H. It turns out that $\boldsymbol{X}$ is min-stable multivariate exponential. To see this, we recall that the increasing sequence $\{\eta_1 + \ldots + \eta_n\}_{n \geq 1}$ equals the enumeration of the points of a Poisson random measure on $[0, \infty)$ with intensity measure equal to the Lebesgue measure. This implies with the help of [91, Proposition 3.6] in $(*)$ below that the survival function $\bar{F}$ of $\boldsymbol{X}$ is given by

$$\bar{F}(\boldsymbol{x}) = \mathbb{P}(Z_{x_1} \leq \epsilon_1, \ldots, Z_{x_d} \leq \epsilon_d) = \mathbb{E}\Big[e^{-\sum_{k=1}^d Z_{x_k}} \Big]$$

$$= \mathbb{E}\Big[\exp\Big\{ -\sum_{n \geq 1} -\log \Big\{ \prod_{k=1}^d G\Big(\frac{\eta_1 + \ldots + \eta_n}{x_k} - \Big) \Big\} \Big\} \Big]$$

$$\overset{(*)}{=} \exp\Big(-\int_0^\infty 1 - \prod_{k=1}^d G\Big(\frac{u}{x_k}\Big)\,\mathrm{d}u \Big).$$

We introduce the notation

$$\ell_G(\boldsymbol{x}) := -\log(\bar{F}(\boldsymbol{x}/M_G)) = \frac{1}{M_G} \int_0^\infty 1 - \prod_{k=1}^d G\Big(\frac{u}{x_k}\Big)\,\mathrm{d}u,$$

and we observe by substitution that $t\,\ell_G(\boldsymbol{x}) = \ell_G(t\,\boldsymbol{x})$ for arbitrary $t > 0$. This implies $\bar{F}(\boldsymbol{x})^t = \bar{F}(t\,\boldsymbol{x})$, so $\boldsymbol{X}$ is min-stable multivariate exponential. The function ℓ_G is the stable tail dependence function of $\boldsymbol{X}$. The constant M_G equals the exponential rate of the exponential random variables $X_1, \ldots, X_d$.

The main theorem in this section states that Examples 5.1 and 5.2 are general enough to understand the structure of the set of all infinite exchangeable sequences $\{X_k\}_{k\in\mathbb{N}}$ whose finite-dimensional margins are both min-stable multivariate exponential and conditionally iid. Concretely, Theorem 5.1 solves Problem 1.3 for the property (P) of "having a min-stable multivariate exponential distribution (in some dimension)". In analytical terms, it states that the stable tail dependence function associated with the extreme-value copula of a conditionally iid min-stable multivariate exponential random vector is a convex mixture of stable tail dependence functions having the structural form as presented in Examples 5.1 and 5.2.

Theorem 5.1 (Which min-stable laws are conditionally iid?). *Let $\{X_k\}_{k\in\mathbb{N}}$ be an infinite exchangeable sequence of positive random variables such that $\boldsymbol{X} = (X_1,\ldots,X_d)$ is min-stable multivariate exponential for all $d \in \mathbb{N}$. Assume that $\{X_k\}_{k\in\mathbb{N}}$ is not iid, i.e. not given as in Example 5.1. Then there exists a unique triplet (b,c,γ) of two constants $b \geq 0$, $c > 0$ and a probability measure γ on $\mathfrak{H}_{+,1}$, such that X_k is exponential with rate $b + c$ for each $k \in \mathbb{N}$ and the stable tail dependence function of $\boldsymbol{X}$ equals*

$$\ell(\boldsymbol{x}) := -\log\left\{\bar{F}\left(\frac{\boldsymbol{x}}{b+c}\right)\right\} = \frac{b}{b+c}\,\|\boldsymbol{x}\|_1 + \frac{c}{b+c}\int_{\mathfrak{H}_{+,1}}\ell_G(\boldsymbol{x})\,\gamma(\mathrm{d}G).$$

In probabilistic terms, the random distribution function H, defined as the limit of empirical distribution functions of the $\{X_k\}_{k\in\mathbb{N}}$ as in Lemma 1.8, necessarily satisfies $H \stackrel{d}{=} 1 - \exp(-Z)$ with

$$Z_t = b\,t + c\sum_{n\geq 1} -\log\left\{G^{(n)}_{\frac{\eta_1+\ldots+\eta_n}{t}-}\right\}, \quad t \geq 0, \tag{5.2}$$

where $G^{(k)}$ is an iid sequence drawn from the probability measure γ, independent of the iid unit exponentials $\eta_1, \eta_2, \ldots$.

Proof. A proof consists of three steps, which have been accomplished in the three references [74, 59, 66], respectively, and which are sketched in the sequel.

(i) For $Z = -\log(1 - H)$ with H as defined in Lemma 1.8 from the sequence $\{X_k\}_{k\in\mathbb{N}}$, [74, Theorem 5.3] shows that

$$Z \stackrel{d}{=} \left\{\sum_{i=1}^n Z^{(i)}_{\frac{t}{n}}\right\}_{t\geq 0}, \quad n \in \mathbb{N}, \tag{5.3}$$

where $Z^{(i)}$ are independent copies of Z. Conversely, it is shown that if Z is non-decreasing and satisfies (5.3), then $1 - \exp(-Z)$ is an element of $\Theta_d^{-1}(\mathfrak{M}_{**})$, when $\mathfrak{M}_{**}$ is as in Problem 1.3 and (P) is the property of "having a min-stable multivariate exponential distribution".

(ii) [59] show that a non-negative stochastic process Z satisfying (5.3) admits a series representation of the form

$$\{Z_t\}_{t\geq 0} \stackrel{d}{=} \left\{b\,t + \sum_{n\geq 1} f^{(n)}_{\frac{t}{\eta_1+\ldots+\eta_n}}\right\}_{t\geq 0},$$

where $f^{(n)}$ are iid copies of some càdlàg stochastic process f with $f_0 = 0$ satisfying some integrability condition, and $b \in \mathbb{R}$.

(iii) [66] proves that in the series representation in (ii) necessarily $b \geq 0$ and f is almost surely non-decreasing. Furthermore, the integrability condition on f can be re-phrased to say that $t \mapsto \tilde{G}_t := \exp(-\lim_{s \downarrow t} f_{1/s})$ defines almost surely the distribution function of some random variable with finite mean $M_{\tilde{G}} = \int_0^\infty 1 - \tilde{G}_t \, dt > 0$. Finally, the distribution function $t \mapsto G_t := \tilde{G}_{M_{\tilde{G}} t}$ has unit mean, and the claimed representation for ℓ is obtained when $c := \mathbb{E}[M_{\tilde{G}}]$ and γ is defined as the probability law of G after an appropriate measure change. That (b, c, γ) is unique follows from the normalization to unit mean of G (for each single realization). $\qquad \square$

Stochastic processes with property (5.3) are said to be *strongly infinitely divisible with respect to time (strong IDT)*. Particular examples of strong IDT processes have been studied in [78, 27, 43], with an emphasis on the associated multivariate min-stable laws also in [74, 9, 64, 76].

Every Lévy process is strong IDT, but the converse needs not hold. For instance, if $Z = \{Z_t\}_{t \geq 0}$ is a non-trivial Lévy subordinator and $a > b > 0$, then the stochastic process $\{Z_{a t} + Z_{b t}\}_{t \geq 0}$ is strong IDT, but not a Lévy subordinator. The probability law γ in Theorem 5.1 in case of a Lévy subordinator is specified as the probability law of

$$G_t = e^{-M} + \left(1 - e^{-M}\right) 1_{\{1 - e^{-M} \geq 1/t\}}, \quad t \geq 0, \tag{5.4}$$

with an arbitrary random variable M taking values in $(0, \infty]$. The Lévy measure of Z and the probability law of M stand in one-to-one relation. We know from the preceding section that if Z is a Lévy subordinator, the associated element in $\mathfrak{M}_{**}$ is a d-variate Marshall-Olkin distribution. Indeed, the Marshall-Olkin distribution is one of the most important examples of min-stable multivariate exponential distributions. Two further examples are presented in the sequel.

Example 5.3 (The (negative) logistic model). *If we reconsider Example 5.2 with the Fréchet distribution function $G(x) = \exp(-\{\Gamma(1 - \theta) \, x\}^{-1/\theta})$ for $\theta \in (0, 1)$, then we observe*

$$\ell_G(\boldsymbol{x}) = \left(\sum_{k=1}^d x_k^{\frac{1}{\theta}} \right)^\theta = \|\boldsymbol{x}\|_{\frac{1}{\theta}}.$$

This is the so-called logistic model. *It is particularly convenient to be looked at from the perspective of conditionally iid models, since the associated strong IDT process Z takes a very simple form, to wit*

$$Z \overset{d}{=} \left\{ S \, t^{\frac{1}{\theta}} \right\}_{t \geq 0}, \quad S \text{ a } \theta\text{-stable random variable, i.e. } \mathbb{E}\left[e^{-x S} \right] = e^{-x^\theta}.$$

In particular, the resulting extreme-value copula, named Gumbel copula *after [41, 42], is also an Archimedean copula, see Remark 3.4. In fact, it is the only*

copula that is both Archimedean and of extreme-value kind, a result first discovered in [39].

A related example is obtained, if we choose the Weibull distribution function $G(x) = 1 - \exp(-\{\Gamma(\theta + 1)\, x\}^{1/\theta})$, *which implies*

$$
\ell_G(\boldsymbol{x}) = \sum_{j=1}^{d} (-1)^{j+1} \sum_{1 \leq i_1 < \ldots < i_j \leq d} \Big(\sum_{k=1}^{j} x_{i_k}^{-\theta} \Big)^{-\frac{1}{\theta}}.
$$

This is the so-called negative logistic model. *The associated extreme-value copula is named* Galambos copula *after [36]. There exist many analogies between logistic and negative logistic models, the interested reader is referred to [37] for background. In particular, the Galambos copula is the most popular representative of the family of so-called* reciprocal Archimedean copulas *as introduced in [38], see also paragraph 7.1 below.*

Example 5.4 (A rich parametric family). *For $G \in \mathfrak{H}_{+,1}$ the function $\Psi_G(z) := \int_0^\infty 1 - G(t)^z\, \mathrm{d}t$ defines a Bernstein function with $\Psi_G(1) = 1$, see [64, Lemma 3]. This implies for $z \in (0, \infty)$ that $G_z \in \mathfrak{H}_{+,1}$, where $G_z(x) := G(x\,\Psi_G(z))^z$. Consequently, if M is a positive random variable, we may define $\gamma \in M_+^1(\mathfrak{H}_{+,1})$ as the law of G_M. The associated stable tail dependence function equals $\ell(\boldsymbol{x}) := \mathbb{E}[\ell_{G_M}(\boldsymbol{x})]$. Many parametric models from the literature are comprised by this construction. In particular, Example 5.2 corresponds to the case $M \equiv 1$, and if $G(x) = \exp(-1) + (1 - \exp(-1))\, 1_{\{1 - \exp(-1) \geq 1/x\}}$ we observe that G_M equals the random distribution function (5.4) corresponding to the Marshall-Olkin subfamily. See [76] for a detailed investigation and applications of this parametric family.*

Remark 5.1 (Extension to laws with exponential minima). *We have seen that the Marshall-Olkin distribution is a subfamily of min-stable multivariate exponential laws. The seminal reference [26] treats both families as multivariate extensions of the univariate exponential law and in the process introduces the even larger family of laws with exponential minima. A random vector $\boldsymbol{X}$ is said to have* exponential minima *if $\min\{X_{i_1}, \ldots, X_{i_k}\}$ has a univariate exponential law for arbitrary $1 \leq i_1 < \ldots i_k \leq d$. Obviously, a min-stable multivariate exponential law has exponential minima, but the converse needs not hold in general. It is shown in [74] that if $Z = \{Z_t\}_{t \geq 0}$ is a right-continuous, non-decreasing process such that $\mathbb{E}[\exp(-x\, Z_t)] = \exp(-t\, \Psi(x))$ for some Bernstein function Ψ, then $\boldsymbol{X}$ as defined in (1.5) has exponential minima. The process Z is said to be* weakly infinitely divisible with respect to time (weak IDT), *and – as the nomenclature suggests – every strong IDT process is also weak IDT. However, there exist weak IDT processes which are not strong IDT. Notice in particular that a Lévy subordinator is uniquely determined in law by the law of Z_1 (or equivalently the Bernstein function Ψ), but neither strong nor weak IDT processes are determined in law by the law of Z_1. If one takes two independent, but different, strong IDT processes $Z^{(1)}, Z^{(2)}$ subject to $Z_1^{(1)} \overset{d}{=} Z_1^{(2)}$, then the*

stochastic process

$$Z_t := \begin{cases} Z_t^{(1)} & \text{if } B = 1, \\ Z_t^{(2)} & \text{if } B = 0, \end{cases}, \quad B \text{ independent Bernoulli}\left(\frac{1}{2}\right)\text{-variate}, \quad t \geq 0,$$

is weak IDT, but not strong IDT. On the level of $\boldsymbol{X}$ this means that the mixture of two min-stable multivariate exponential random vectors always has exponential minima, but needs not be min-stable anymore.

Remark 5.2 (Archimax copulas)**.** *The study of min-stable multivariate exponentials is analogous to the study of extreme-value copulas. From this perspective, Theorem 5.1 gives us a canonical stochastic model for all conditionally iid extreme-value copulas. Another family of copulas for which we understand the conditionally iid subfamily pretty well is Archimedean copulas, related to ℓ_1-norm symmetric distributions and mentioned in Remark 3.4. The family of so-called* Archimedean copulas *is a superclass of both extreme-value and Archimedean copulas. It has been studied in [14, 15] with the intention to create a rich copula family that comprises well-known subfamilies. An extreme-value copula C is conveniently described in terms of its stable tail dependence function. Recall that Theorem 5.1 is formulated in terms of the stable tail dependence function and gives an analytical criterion for C to be conditionally iid. An Archimax copula C is a multivariate distribution function of the functional form*

$$C_{\ell,\varphi}(u_1, \ldots, u_d) = \varphi\Big(\ell\big(\varphi^{-1}(u_1), \ldots, \varphi^{-1}(u_d)\big)\Big).$$

It is recognized that if $\ell(x_1, \ldots, x_d) = \|\boldsymbol{x}\|_1$ then $C_{\ell,\varphi}$ is an Archimedean copula, and if $\varphi(x) = \exp(-x)$, then $C_{\ell,\varphi}$ is an extreme-value copula. By combining our knowledge from Theorems 3.2 and 5.1 about Archimedean and extreme-value copulas, it is immediate to show that

$$\Big\{\tilde{\gamma} \in M_+^1(\mathfrak{H}_+) \,:\, \big\{1 - e^{-Z_M t}\big\}_{t \geq 0} \sim \tilde{\gamma}, \, M > 0 \text{ a positive random variable, and}$$

$$\{Z_t\}_{t \geq 0} \text{ non-decreasing strong IDT}\Big\} \subset \Theta_d^{-1}\big(\mathfrak{M}_{**}\big), \tag{5.5}$$

when $\mathfrak{M}$ denotes the family of all probability laws with the property (P) of "having a survival function of the functional form $\varphi \circ \ell$ with ℓ some stable tail dependence function". In this case, the function φ equals the Laplace transform of M and ℓ is given in terms of a triplet (b, c, γ) such as in Theorem 5.1, associated with the strong IDT process Z, and $b + c = 1$. Notice that each stable tail dependence function ℓ equals the restriction of an orthant-monotonic norm to $[0, \infty)^d$, see [85], so that survival functions of the form $\varphi \circ \ell$ are precisely the survival functions that are symmetric with respect to the norm ℓ.

6. Exogenous shock models

The present section studies a family $\mathfrak{M}$ of multivariate distribution functions that have a stochastic representation according to the following exogenous shock

model: We consider some system consisting of d components and interpret the k-th component of our random vector $\boldsymbol{X} = (X_1, \ldots, X_d)$ with law in $\mathfrak{M}$ as the lifetime of the k-th component in our system. A component lives until it is affected by an exogenous shock, and the arrival times of these exogenous shocks are modeled stochastically. For each non-empty subset $I \subset \{1, \ldots, d\}$ of components, we denote by E_I a non-negative random variable. We assume that all E_I are independent and interpret E_I as the arrival time of an exogenous shock affecting all components of our random vector which are indexed by I. This means that we define

$$X_k := \min\{E_I \,:\, k \in I\}, \quad k = 1, \ldots, d. \tag{6.1}$$

Such exogenous shock models are popular in reliability theory, insurance risk, and portfolio credit risk. Recall from Example $4.1(\mathcal{E})$ that this model is a generalization of the Marshall-Olkin distribution, which arises as special case if all the E_I are exponentially distributed, see also Example 6.1 below. For the sake of clarity, we formally introduce the following definition.

Definition 6.1 (Exogenous shock model). *A probability measure $\mu \in M_+^1(\mathbb{R}^d)$ is said to* define an exogenous shock model *if on some probability space there exists a random vector $\boldsymbol{X}$ with stochastic representation (6.1) such that $\boldsymbol{X} \sim \mu$.*

6.1. Exchangeability and the extendibility problem

We are interested in a solution of Problem 1.3 for the property (P) of "defining an exogenous shock model". By Lemma 1.1 exchangeability is a necessary requirement on $\boldsymbol{X}$, and we observe immediately from (6.1) that this implies that the distribution function of E_I is allowed to depend on the subset I only through its cardinality $|I|$. Some simple algebraic manipulations, see the proof of Theorem 6.1 below, reveal that the survival function of $\boldsymbol{X}$ necessarily must be given as the product of its arguments after being ordered and idiosyncratically distorted. Already the characterization of the exchangeable subfamily in analytical terms is an interesting problem, the interested reader is referred to [68] for its solution.

The conditionally iid subfamily $\mathfrak{M}_{**}$ is also investigated in [68]. One major finding is that when the increments of the factor process Z in the canonical construction (1.5) are independent, then one ends up with an exogenous shock model. Recall that a càdlàg stochastic process $Z = \{Z_t\}_{t \geq 0}$ with independent increments is called *additive*, see [94] for a textbook treatment. For our purpose, it is sufficient to be aware that the probability law of a non-decreasing additive process $Z = \{Z_t\}_{t \geq 0}$ with $Z_0 = 0$ can be described uniquely in terms of a family $\{\Psi_t\}_{t \geq 0}$ of Bernstein functions defined by $\Psi_t(x) := -\log(\mathbb{E}[\exp(-x \, Z_t)])$, $x \geq 0$, i.e. Ψ_t equals the Laplace exponent of the infinitely divisible random variable Z_t. The independent increment property implies for $0 \leq s \leq t$ that $\Psi_t - \Psi_s$ is also a Bernstein function and equals the Laplace exponent of the infinitely divisible random variable $Z_t - Z_s$. The easiest example for a non-decreasing additive process is a Lévy subordinator, in which case $\Psi_t = t \, \Psi_1$,

i.e. the probability law is described completely in terms of just one Bernstein function Ψ_1 (due to the defining property that the increments are not only independent but also identically distributed). Two further compelling examples of (non-Lévy) additive processes are presented in subsequent paragraphs.

Theorem 6.1 (Additive subordinators and exogenous shock models)**.** *Let $\mathfrak{M}$ denote the family of probability laws with the property* (P) *of "defining an exogenous shock model". A random vector $\boldsymbol{X}$ has law in $\mathfrak{M}$ and is exchangeable if and only if it admits a survival copula of the functional form*

$$\hat{C}(u_1, \ldots, u_d) = u_{[1]} \prod_{k=2}^{d} g_k(u_{[k]}), \quad u_1, \ldots, u_d \in [0, 1], \tag{6.2}$$

with certain functions $g_k : [0, 1] \to [0, 1]$. Furthermore,

$$\left\{ \gamma \in M_+^1(\mathfrak{H}_+) \, : \, \left\{ 1 - e^{-Z_t} \right\}_{t \geq 0} \sim \gamma, \, \{Z_t\}_{t \geq 0} \text{ additive process} \right\} = \Theta_d^{-1}(\mathfrak{M}_{**}).$$

Proof. A proof for the inclusion "$\supset$" has been accomplished only recently and can be found in [102]. A proof sketch for the inclusion "$\subset$" works as follows, see [68] for details. The survival function of the random vector $\boldsymbol{X}$ defined by (6.1) can be written in terms of the one-dimensional survival functions of the E_I as

$$\mathbb{P}(\boldsymbol{X} > \boldsymbol{x}) = \prod_{\emptyset \neq I} \mathbb{P}(E_I > \max\{x_k \, : \, k \in I\}).$$

Exchangeability of $\boldsymbol{X}$ implies that the probability law of E_I depends on I only via its cardinality $|I| \in \{1, \ldots, d\}$. If we denote the survival function of E_I with $|I| = m$ by $\bar{H}_m$, we observe that

$$\mathbb{P}(\boldsymbol{X} > \boldsymbol{x}) = \prod_{m=1}^{d} \prod_{I \, : \, |I| = m} \bar{H}_m(\max\{x_k \, : \, k \in I\})$$

$$= \prod_{m=1}^{d} \prod_{k=1}^{d-m+1} \left(\bar{H}_m(x_{[d-k+1]}) \right)^{\binom{d-k}{m-1}} = \prod_{k=1}^{d} \prod_{m=1}^{d-k+1} \left(\bar{H}_m(x_{[d-k+1]}) \right)^{\binom{d-k}{m-1}}. \tag{6.3}$$

Noting for $\boldsymbol{x} = (x, 0, \ldots, 0)$ that $x_{[d]} = x$ and $x_{[1]} = \ldots = x_{[d-1]} = 0$, we observe that the one-dimensional margins are

$$\mathbb{P}(X_k > x) = \prod_{m=1}^{d} \left(\bar{H}_m(x) \right)^{\binom{d-1}{m-1}} =: \bar{F}_1(x), \quad k = 1, \ldots, d.$$

That (6.3) can be written as $\hat{C}\big(\mathbb{P}(X_1 > x_1), \ldots, \mathbb{P}(X_d > x_d)\big)$ with $\hat{C}$ as in (6.2) follows by a tedious yet straightforward computation with the g_k defined as

$$g_k := \prod_{m=1}^{d-k+1} \left(\bar{H}_m \circ \bar{F}_1^{-1} \right)^{\binom{d-k}{m-1}}, \quad k = 2, \ldots, d,$$

where $\bar{F}_1^{-1}$ denotes the generalized inverse of the non-increasing function $\bar{F}_1$, which is defined analogous to the generalized inverse of a distribution function as

$$\bar{F}_1^{-1}(x) := \inf\{t > 0 \,:\, \bar{F}(t) \leq x\}.$$

Now assume Z is an additive process with associated family of Bernstein functions $\{\Psi_t\}_{t \geq 0}$. The survival copula of the random vector $\boldsymbol{X}$ of Equation (1.5) can be computed in closed form using the independent increment property of Z. It is easily shown to be of the structural form (6.2), when

$$g_k(u) := \exp\big(-\Psi_{\bar{F}_1^{-1}(u)}(k) + \Psi_{\bar{F}_1^{-1}(u)}(k-1)\big), \quad k = 2, \ldots, d,$$

with $\bar{F}_1(x) := \exp(-\Psi_x(1))$, $x \geq 0$. $\qquad\qquad\qquad\square$

Remark 6.1 (Related literature). *Interestingly, there exists quite some literature that proposes the use of the random distribution function $H = 1 - \exp(-Z)$ with Z additive as a prior distribution when estimating the probability law of observed samples $X_1, \ldots, X_d$ – without noticing the relation to exogenous shock models. H is sometimes called a* neutral-to-the-right prior *in these references. The use of additive Z in this particular application is mainly explained by analytical convenience, because it implies that prior and posterior distributions have a similar algebraic structure. To provide some examples, the interested reader is referred to the papers [51, 48, 61] and the references mentioned therein. Furthermore, [55] calls the random probability measures defined via the random distribution function H* completely random measures, *and characterizes them by the property that the values they take on disjoint subsets are independent.*

There are some interesting subfamilies of exogenous shock models that are worth mentioning with respect to their conditionally iid substructure. The first of them is a well known friend from previous sections, re-visited once again in the following example.

Example 6.1 (The Marshall–Olkin law revisited). *If all random variables E_I in the exogenous shock model construction (6.1) are exponentially distributed, we are in the special situation of Example 4.1($\mathcal{E}$). Indeed, it has already been shown in the original reference [79] that every Marshall–Olkin distribution can be constructed like this. Hence, we already know from Theorem 4.1($\mathcal{E}$) that an exogenous shock model with exponential arrival times is obtained via the canonical conditionally iid model (1.5) if the associated stochastic process $H = \{H_t\}_{t \geq 0} \sim \gamma \in M_+^1(\mathfrak{H}_+)$ is such that $Z_t := -\log(1 - H_t)$, $t \geq 0$, defines a Lévy subordinator, which is a special additive subordinator.*

Example 6.2 (A simple global shock model). *A special case of copulas of the form (6.2) is considered in [22], namely $g_2 = \ldots = g_d$, which we briefly put in context with the additive process construction. To this end, let g_2 be a strictly increasing and continuous distribution function of some random variable taking values in $[0, 1]$, assuming $x \mapsto g_2(x)/x$ is non-increasing on $(0, 1]$. The function $F_M(x) := x/g_2(x)$ then is a distribution function on $[0, 1]$, and we let M be a*

random variable with this distribution function. Independently, let $W_1, W_2, \ldots$ be an iid sequence drawn from g_2. We consider the infinite exchangeable sequence $\{X_k\}_{k \in \mathbb{N}}$ with $X_k := \min\{E_{\{k\}}, E_{\{1,2,\ldots\}}\}$, $k \in \mathbb{N}$, where

$$E_{\{k\}} := -\log\big(F_M(W_k)\big),\, k \in \mathbb{N}, \quad E_{\{1,2,\ldots\}} := -\log\big(F_M(M)\big).$$

By definition, each finite d-margin has an exogenous shock model representation, and the survival copula $\hat{C}$ is easily seen to be of the form (6.2) with $g_2 = \ldots = g_d$, for arbitrary $d \geq 2$. The conditional distribution function is static in the sense of Section 3, and given by $H = 1 - \exp(-Z)$ with

$$Z_t := \begin{cases} -\log\Big(g_2\big\{F_M^{-1}(e^{-t})\big\}\Big) & , \text{ if } t < E_{\{1,2,\ldots\}} \\ \infty & , \text{ if } t \geq E_{\{1,2,\ldots\}} \end{cases}, \quad t \geq 0.$$

The random variable $E_{\{1,2,\ldots\}}$ is unit exponential, and Z is additive with associated family of Bernstein functions $\Psi_t(x) = -\log(\mathbb{E}[\exp(-x Z_t)])$ given by

$$\Psi_t(x) = t \, 1_{\{x>0\}} + x \Big(-\log\big\{g_2\{F_M^{-1}(e^{-t})\}\big\}\Big), \quad x, t \geq 0.$$

Notice that for each fixed $t > 0$ this corresponds to an infinitely divisible distribution of Z_t that is concentrated on the set

$$\Big\{ -\log\big(g_2\{F_M^{-1}(e^{-t})\}\big), \infty \Big\}.$$

The case $g_2(x) = x^\alpha$ with $\alpha \in [0,1]$ implies that Z is a killed Lévy subordinator that grows linearly before it jumps to infinity, $\boldsymbol{X}$ has a Marshall-Olkin law, and the E_I are exponential. In the general case, Z needs not grow linearly before it gets killed.

Two further examples are studied in greater detail in the following two paragraphs, since they give rise to nice characterization results.

6.2. The Dirichlet prior and radial symmetry

In the two landmark papers [33, 34], T.S. Ferguson introduces the so-called Dirichlet prior and shows that it can be constructed by means of an additive process. More clearly, let $c > 0$ be a model parameter and let $G \in \mathfrak{H}_+$, continuous and strictly increasing. Consider a non-decreasing additive process $Z = \{Z_t\}_{t \in [G^{-1}(0), G^{-1}(1)]}$ whose probability law is determined by a family of Bernstein functions $\{\Psi_t\}_{t \in (G^{-1}(0), G^{-1}(1))}$, which are given by

$$\Psi_t(x) = \int_0^\infty \big(1 - e^{-x u}\big) \frac{e^{-u c (1-G(t))} - e^{-u c}}{u (1 - e^{-u})} \, du, \quad x \geq 0.$$

The random distribution function $H = \{H_t\}_{t \in [G^{-1}(0), G^{-1}(1)]}$ defined by $H_t := 1 - \exp(-Z_t)$ satisfies the following property: For arbitrary $G^{-1}(0) < t_1 < \ldots < t_d < G^{-1}(1)$ the random vector

$$\big(H_{t_1}, H_{t_2} - H_{t_1}, \ldots, H_{t_d} - H_{t_{d-1}}, 1 - H_{t_d}\big)$$

has a Dirichlet distribution[16] with parameters

$$c\big(G(t_1), G(t_2) - G(t_1), \ldots, G(t_d) - G(t_{d-1}), 1 - G(t_d)\big),$$

and H is called *Dirichlet prior* with parameters (c, G), denoted $DP(c, G)$ in the sequel. The probability distribution of $(X_1, \ldots, X_d)$ in (1.5), when $H = DP(c, G)$ for some G with support $[G^{-1}(0), G^{-1}(1)] := [0, \infty]$, is given by

$$\mathbb{P}(X_1 > x_1, \ldots, X_d > x_d) = \hat{C}_c\big(1 - G(x_1), \ldots, 1 - G(x_d)\big), \text{ with}$$

$$\hat{C}_c(u_1, \ldots, u_d) = u_{[1]} \prod_{k=2}^{d} \frac{c\, u_{[k]} + k - 1}{c + k - 1}. \tag{6.4}$$

It is insightful to remark that for $c \searrow 0$ the copula $\hat{C}_c$ converges to the so-called *upper-Fréchet Hoeffding copula* $\hat{C}_0(\boldsymbol{u}) = u_{[1]}$, and for $c \nearrow \infty$ to the copula $\hat{C}_\infty(\boldsymbol{u}) = \prod_{k=1}^{d} u_k$ associated with independence. The intuition of the Dirichlet prior model is that all components of $\boldsymbol{X}$ have distribution function G, but one is uncertain whether G is really the correct distribution function. So the parameter c models an uncertainty about G in the sense that the process H must be viewed as a "distortion" of G. For $c \nearrow \infty$ we obtain $H = G$, while for $c \searrow 0$ the process H is maximally chaotic (in some sense) and does not resemble G at all.

Interestingly, if the probability law dG is symmetric about its median $\mu := G^{-1}(0.5)$, then the random vector $(X_1, \ldots, X_d)$ is radially symmetric, which can be verified using Lemma 1.7. One can furthermore show that there exists no other conditionally iid exogenous shock model satisfying this property, see the following lemma. To this end, recall that a copula C is called *radially symmetric* if $C = \hat{C}$, i.e. it equals its own survival copula, which means that $\boldsymbol{U} = (U_1, \ldots, U_d) \overset{d}{=} (1 - U_1, \ldots, 1 - U_d)$ for $\boldsymbol{U} \sim C$.

Lemma 6.1 (Radial symmetry in exchangeable exogenous shock models). *A copula of the structural form (6.2) is radially symmetric if and only if the functions g_k are linear, $k = 2, \ldots, d$, which is the case if and only if there is a $c \in [0, \infty]$ such that the copula takes the form (6.4).*

Proof. This is [69, Theorem 3.5]. In order to prove necessity, the principle of inclusion and exclusion can be used to express the survival copula of $\hat{C}$ as an alternating sum of lower-dimensional margins of $\hat{C}$. By radial symmetry, this expression equals $\hat{C}$, and on both sides of the equation one may now take the derivatives with respect to all d arguments. A lengthy but tedious computation then shows that the g_k must all be linear, which implies the claim. Sufficiency is proved using the Dirichlet prior construction. The defining properties of the Dirichlet prior imply that the assumptions of Lemma 1.7 are satisfied, which implies the claim. □

---

[16]Recall from Remark 3.5 that $\boldsymbol{S} = (S_1, \ldots, S_d)$ has a Dirichlet distribution with parameters $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_d)$ if $\boldsymbol{S} \overset{d}{=} \boldsymbol{G}/\|\boldsymbol{G}\|_1$ for a vector $\boldsymbol{G}$ of independent unit-scale Gamma-distributed random variables.

6.3. The Sato-frailty model and self-decomposability

A real-valued random variable X is called *self-decomposable* if for arbitrary $c \in (0,1)$ there exists an independent random variable Y such that $X \stackrel{d}{=} c\,X + Y$. It can be shown that a self-decomposable X is infinitely divisible, so self-decomposable laws are special cases of infinitely divisible laws. In particular, if X takes values in $(0,\infty)$ and is infinitely divisible with Laplace exponent given by the Bernstein function Ψ, then X is self-decomposable if and only if the function $x \mapsto x\,\Psi^{(1)}(x)$ is again a Bernstein function, see [104, Theorem 2.6, p. 227]. Now let Ψ be the Bernstein function associated with a self-decomposable law on $(0,\infty)$, and consider a family of Bernstein functions defined by $\Psi_t(x) := \Psi(x\,t)$, $t \geq 0$. One can show that there exists an additive subordinator $Z = \{Z_t\}_{t \geq 0}$ which is uniquely determined in law by $\{\Psi_t\}_{t \geq 0}$ via $\Psi_t(x) = -\log(\mathbb{E}[\exp(-x\,Z_t)])$, $x, t \geq 0$, called *Sato subordinator*. If we use this process in (1.5), the conditionally iid random vector $\boldsymbol{X}$ obtained by this construction has survival function given by

$$\mathbb{P}(\boldsymbol{X} > \boldsymbol{x}) = \exp\left\{-\sum_{k=1}^{d} \Psi\big((d-k+1)\,x_{[k]}\big) - \Psi\big((d-k)\,x_{[k]}\big)\right\}, \quad \boldsymbol{x} \in [0,\infty)^d.$$
(6.5)

The following lemma characterizes self-decomposability analytically in terms of multivariate probability laws given by (6.5).

Lemma 6.2 (Characterization of self-decomposable Bernstein functions)**.** *Let $\Psi : [0,\infty) \to [0,\infty)$ be some function. The d-variate function (6.5) defines a proper survival function on $[0,\infty)^d$ for all $d \geq 2$ if and only if Ψ equals the Bernstein function of a self-decomposable probability law on $(0,\infty)$.*

Proof. Sufficiency is an instance of the general Theorem 6.1, as demonstrated above. Necessity, i.e. that self-decomposability can actually be characterized in terms of the multivariate survival functions (6.5), is shown in [70] and relies on some purely analytical, technical computations. $\qquad\square$

Example 6.3 (A one-parametric, multivariate Pareto distribution)**.** *Let $\Psi(x) = \alpha\,\log(1+x)$ be the Bernstein function associated with a Gamma distribution[17] with parameter $\alpha > 0$. The Gamma distribution is self-decomposable and the survival function (6.5) takes the explicit, one-parametric form*

$$\mathbb{P}(\boldsymbol{X} > \boldsymbol{x}) = \left(\prod_{k=1}^{d} \frac{(d-k)\,x_{[k]} + 1}{(d-k+1)\,x_{[k]} + 1}\right)^{\alpha}.$$

The one-dimensional marginal survival functions are given by $\bar{F}_1(x) = (1+x)^{-\alpha}$. Notice that this equals the survival function of $Y - 1$, when Y has a Pareto distribution with scale parameter (aka left-end point of support) equal to one

---

[17]This is precisely the Gamma distribution with density (3.7) for $\alpha = \alpha_k$.

and tail index α. Thus, the random vector $\boldsymbol{X} + \boldsymbol{1} := (X_1 + 1, \ldots, X_d + 1)$ might be viewed as a multivariate extension of the Pareto distribution with scale parameter equal to one and tail index α.

7. Related open problems

7.1. Extendibility-problem for further families

The present article surveys solutions to Problems 1.1 and 1.3 for several families $\mathfrak{M}$ of interest. One goal of the survey is to encourage others to solve the problem also for other families. We provide examples that we find compelling:

(i) The family of min-stable laws in Section 5 can be generalized to min-infinitely divisible laws. Generalizing (5.1), a multivariate survival function $\bar{F}$ is called min-infinitely divisible if for each $t > 0$ there is a survival function $\bar{F}_t$ such that $\bar{F}(\boldsymbol{x})^t = \bar{F}_t(\boldsymbol{x})$. Like min-stability is analogous to max-stability, the concept of min-infinite divisibility is equivalent to the concept of max-infinite divisibility, on which [91] provides a textbook treatment. It is pretty obvious that non-decreasing infinitely divisible processes occupy a commanding role with regards to the conditionally iid subfamily, but to work out a convenient analytical treatment of these in relation with the associated min-infinitely divisible laws appears to be a promising direction for further research. Notice that the family of reciprocal Archimedean copulas, introduced in [38], is one particular special case of max-infinitely divisible distribution functions, and in this special case the conditionally iid subfamily is determined similarly as in the case of Archimedean copulas, see [38, Section 7]. This might serve as a good motivating example for the aforementioned generalization.

(ii) Theorem 3.3 studies d-variate densities of the form $g_d(x_{[d]})$, and [40] also considers a generalization to densities of the form $g(x_{[1]}, x_{[d]})$, depending on $x_{[1]}$ and $x_{[d]}$. From a purely algebraic viewpoint it is tempting to investigate whether exchangeable densities of the structural form $\prod_{k=1}^{d} g_k(x_{[k]})$ allow for a nice theory as well. When are these conditionally iid? This generalization of the ℓ_∞-norm symmetric case is motivated by a relation to non-homogeneous pure birth processes, as already explained in Remark 3.7. Such processes are of interest in reliability theory, as explained in [99], see also [11] and [67, Section 4] for related investigations.

(iii) On page 721 it was mentioned that the Marshall-Olkin distribution is characterized by the property that for all subsets of components the respective "survival indicator process" is a continuous-time Markov chain. This property may naturally be weakened to the situation when only the survival indicator process $Z_t := (1_{\{X_1 > t\}}, \ldots, 1_{\{X_d > t\}})$ of all components is a continuous-time Markov chain. On the level of multivariate distributions, one generalizes the Marshall-Olkin distribution to a more general family of multivariate laws that has been shown to be interesting in mathematical finance in [46]. Furthermore, it is a subfamily of the even larger

family of so-called multivariate phase-type distributions, see [5]. Which members of theses families of distributions are conditionally iid? Presumably, this research direction requires to generalize the Lévy subordinator in the Marshall–Olkin case to more general non-decreasing Markov processes.

7.2. Testing for conditional independence

If a specific d-variate law in some family $\mathfrak{M}$ is given, do we have a practically useful, analytical criterion to decide whether or not this law is in $\mathfrak{M}_*$, resp. $\mathfrak{M}_{**}$, or not? According to Theorem 1.2, in general this requires to check whether a supremum over bounded measurable functions is bounded from above, which in practice is rather inconvenient – at least on first glimpse. For certain families, however, there is hope to find more useful criteria. For instance, for Marshall-Olkin distributions the link to the truncated moment problem in Remark 4.1 is helpful in this regard, like it is for binary sequences. For Archimedean copulas (resp. ℓ_1-norm symmetric survival functions) this boils down to checking whether a d-monotone function is actually completely monotone, i.e. a Laplace transform. However, it is an open problem for the family of extreme-value copulas. Of course, Theorem 5.1 tells us which stable tail dependence functions correspond to conditionally iid laws. But given some specific stable tail dependence function, how can we tell effectively whether or not this given function has the desired form? Even for dimension $d = 2$, in which case the problem is presumably easier due to the fact that the 2-dimensional unit simplex is one-dimensional, this problem is non-trivial and open. Given we find such effective analytical criterion for some family $\mathfrak{M}$, is it even possible to build a useful statistical test based on it, i.e. can we test the hypothesis that the law is conditionally iid?

7.3. Combination of one-factor models to multi-factor models

This is probably the most obvious application of the presented concepts. The idea works as follows. According to our notation, the dependence-inducing latent factor in a conditionally iid model is H. Depending on the stochastic properties of $H \sim \gamma \in M_+^1(\mathfrak{H})$, it may be possible to construct H from a pair $(H^{(1)}, H^{(2)}) \sim \gamma_1 \otimes \gamma_2 \in M_+^1(\mathfrak{H}) \times M_+^1(\mathfrak{H})$ of two independent processes of the same structural form, say $H = f(H^{(1)}, H^{(2)})$. For example, if $H^{(1)}$ and $H^{(2)}$ are two strong IDT processes, see Section 5, then so is their sum $H = H^{(1)} + H^{(2)}$. In this situation, we may define *dependent* processes $H^{(1,1)}, \ldots, H^{(1,J)}$ from $J + 1$ *independent* processes $H^{(0)}, \ldots, H^{(J)}$ as $H^{(1,j)} = f(H^{(0)}, H^{(j)})$. The conditionally iid vectors $\boldsymbol{X}^{(1)}, \ldots, \boldsymbol{X}^{(J)}$ defined via (1.4) from $H^{(1,1)}, \ldots, H^{(1,J)}$ are then dependent, so that the combined vector $\boldsymbol{X} = (\boldsymbol{X}^{(1)}, \ldots, \boldsymbol{X}^{(J)})$ has a hierarchical dependence structure. Such structures break out of the – sometimes undesired and limited – exchangeable cosmos and have the appealing property that the lowest-level groups are conditionally iid, so the whole structure can be

sized up, i.e. is dimension-free to some degree. Of particular interest is the situation when the random vector $(X_1^{(1)}, \ldots, X_1^{(J)})$ composed of one component from each of the J different groups is conditionally iid and its latent factor process equals $H^{(0)}$ in distribution. In this particular situation, an understanding of the whole dependence structure of the hierarchical model $\boldsymbol{X}$ is retrieved from an understanding of the conditionally iid sub-models based on the $H^{(j)}$. In other words, the conditionally iid model can be *nested* to construct highly tractable, non-exchangeable, multi-factor dependence models from simple building blocks. For instance, hierarchical elliptical laws, Archimedean copulas[18], and min-stable laws can be constructed based on the presented one-factor building blocks, see [73] for an overview. For these and other families, the design, estimation, and efficient simulation of such hierarchical structures is an active area of research or even an unsolved problem.

7.4. Parameter estimation with uncertainty

The classical statistical parameter estimation problem is to estimate the (true) parameter m of a one-parametric distribution function F_m from iid observations $X_1, \ldots, X_d \sim F_m$. A parameter estimate is then a function $\hat{m} = \hat{m}(X_1, \ldots, X_d)$ of the observations into the set of admissible parameters. This classical problem relies on the hypothesis that there is a "true" parameter m, from which the observations are drawn. But what if we are uncertain whether or not the observations are actually drawn from some F_m? The Dirichlet prior has been introduced in [33, 34] with the motivation to model uncertainty about the hypothesis that observations are drawn from some F_m. Instead, it is assumed that they are drawn from $DP(c, F_m)$ with an uncertainty parameter $c > 0$. On a high level, this amounts to observing one sample $\boldsymbol{X} = (X_1, \ldots, X_d)$, with large d, from a parametric conditionally iid model. Optimal estimates for m based on the observations can then be derived due to the convenient Dirichlet prior setting, see [33, 34] for details. But this question can clearly also be posed for other conditionally iid models. Let us provide a second motivation that appears to be natural: let $X_1, \ldots, X_d$ be observed time points of company bankruptcy filings within the last 10 years. An iid assumption for $X_1, \ldots, X_d$ is well known to be inappropriate. Instead, a popular model for such time points is a Marshall-Olkin distribution, see [28]. If we assume in addition – now for mathematical convenience – that $\boldsymbol{X} = (X_1, \ldots, X_d)$ is conditionally iid, we know from Theorem 4.1($\mathcal{E}$) and Lemma 1.8 that the empirical distribution function of $X_1, \ldots, X_d$ is approximately equal to $1 - \exp(-Z)$ for a Lévy subordinator Z. Depending on a specific parametric model for Z, it is well possible that we can estimate the parameters based on the observed empirical distribution function. For example, if Z is a compound Poisson process with constant jump size m, then huge (small) jumps in the empirical distribution function apparently indicate a large (small) value of m. Such parameter estimation problems based on one (large) sample $\boldsymbol{X} = (X_1, \ldots, X_d)$ from a conditionally iid model appear to be very

---

[18]See also the many references in Remark 3.4.

model-specific and thus possibly interesting, and the two motivating examples above indicate that one might find natural motivations for these.

7.5. Quantification of diversity of possible extensions

All of the presented theorems solve Problem 1.3, but only in some cases[19] the solution set $\mathfrak{M}_{**}$ is shown to coincide with the in general larger solution set $\mathfrak{M}_*$ in Problem 1.1. Can one show that $\mathfrak{M}_* = \mathfrak{M}_{**}$ in the other presented solutions of Problem 1.3? To provide one concrete example, from Theorem 4.1($\mathcal{G}$) we know that $(b_0, b_1, b_2) \in \mathcal{M}_2$ determines a three-dimensional, exchangeable wide-sense geometric law. However, this exchangeable probability distribution is only in $\mathfrak{M}_{**}$ if there exist $b_3, b_4, \ldots$ such that $\{b_k\}_{k \in \mathbb{N}_0} \in \mathcal{M}_\infty$. Could it be that the last extension property does fail, but the three-dimensional, exchangeable wide-sense geometric law associated with (b_0, b_1, b_2) is still conditionally iid? If so, then necessarily there is some $n > 3$ and an infinite exchangeable sequence $\{X_k\}_{k \in \mathbb{N}}$ such that (X_1, X_2, X_3) has the given wide-sense geometric law but $(X_1, \ldots, X_n)$ is not wide-sense geometric.

A related question concerns only elements in $\mathfrak{M}_{**}$. There might be two infinite exchangeable sequences $\{X_k^{(1)}\}_{k \in \mathbb{N}}$ and $\{X_k^{(2)}\}_{k \in \mathbb{N}}$ with $\{X_k^{(1)}\}_{k \in \mathbb{N}} \overset{d}{\neq} \{X_k^{(2)}\}_{k \in \mathbb{N}}$ but $(X_1^{(1)}, \ldots, X_d^{(1)}) \overset{d}{=} (X_1^{(2)}, \ldots, X_d^{(2)})$ for some $d \in \mathbb{N}$. To provide an example, related to Theorems 2.1 and 4.1, the vector $(1, b_1)$ with $b_1 \in [0, 1]$ can always be extended to a sequence $\{b_k\}_{k \in \mathbb{N}}$ that is completely monotone, for example set $b_k = b_1^k$. In case of Theorem 4.1($\mathcal{G}$), all the different possible extensions imply different exchangeable sequences $\{X_k\}_{k \in \mathbb{N}}$ such that 2-margins follow the associated wide-sense geometric law with parameters $(1, b_1)$. But all these extensions in Theorem 4.1($\mathcal{G}$) have in common that arbitrary d-margins are always wide-sense geometric. Can one quantify *how* different such extensions are allowed to be? A similar question is: Is the "$\subset$" in (5.5) actually a "$=$"? Notice that the proof ideas in [20, 90], who study such issues in the case of some static laws, might help to approach such questions.

7.6. Characterization of stochastic objects via multivariate
probability laws

As a general rule, for an infinite exchangeable sequence $\{X_k\}_{k \in \mathbb{N}}$ defined via (1.4) the probability law of the random distribution function H is uniquely determined by its mixed moments

$$\mathbb{E}[H_{t_1} \cdots H_{t_d}] = \mathbb{P}(X_1 \leq t_1, \ldots, X_d \leq t_d), \quad d \in \mathbb{N}, \, t_1, \ldots, t_d \in \mathbb{R}.$$

This often implies interesting analytical characterizations of the stochastic object H in terms of the multivariate distribution functions $\boldsymbol{t} \mapsto \mathbb{P}(\boldsymbol{X} \leq \boldsymbol{t})$. In

---

[19]To wit, Example 1.1, Schoenberg's Theorem 3.1 and Theorem 3.3.

particular, if H is of the form $H = 1 - \exp(-Z)$ like in (1.5), then the mixed moments above become

$$\mathbb{E}\left[e^{-\sum_{k=1}^{d} Z_{t_k}}\right] = \mathbb{P}(X_1 > t_1, \ldots, X_d > t_d), \quad d \in \mathbb{N}, \, t_1, \ldots, t_d \geq 0,$$

that is the survival functions $\boldsymbol{t} \mapsto \mathbb{P}(X_1 > t_1, \ldots, X_d > t_d)$ stand in one-to-one relation with the Laplace transforms of finite-dimensional margins of the non-decreasing process Z. This general relationship explains the close connection between conditionally iid probability laws and moment problems/ Laplace transforms encountered several times in this survey. For instance, Theorem 3.2 shows that φ is a Laplace transform if and only if $\boldsymbol{x} \mapsto \varphi(\|\boldsymbol{x}\|_1)$ is a survival function for all $d \geq 1$, or Theorem 4.1 characterizes Lévy subordinators in terms of multivariate survival functions, or Lemma 6.2 characterizes self-decomposable Bernstein functions via multivariate survival functions. Can further characterizations be found? Is there a compelling application for such characterizations in terms of multivariate survival functions?

References

[1] ALDOUS, D.J. (1985). *Exchangeability and related topics*. Springer, École d'Été de Probabilités de Saint-Flour XIII-1983, Lecture Notes in Mathematics 1117, 1–198. MR0883646

[2] ALDOUS, D.J. (1985). *More uses of exchangeability: representations of complex random structures*. in Probability and Methematical Genetics – papers in honour of Sir John Kingman, Cambridge University Press 35–63. MR2744234

[3] ALFSEN, E.M. (1971). *Compact convex sets and boundary integrals*. Springer, Berlin. MR0445271

[4] ARNOLD, B.C. (1975). A characterization of the exponential distribution by multivariate geometric compounding. *Sankhyā: The Indian Journal of Statistics* **37:1** 164–173. MR0440792

[5] ASSAF, D. and LANGBERG, N.A. and SAVITS, T.H. and SHAKED, M. (1984). Multivariate phase-type distributions. *Operations Research* **32:3** 688–702. MR0756014

[6] BARLOW, R.E. and PROSCHAN, F. (1975). *Statistical theory of reliability and life testing*. Rinehart and Winston, New York. MR0438625

[7] BEIRLANT, J. and GOEGEBEUR, Y. and TEUGELS, J. and SEGERS, J. (2004). *Statistics of extremes: theory and applications*. John Wiley & Sons, Chichester. MR2108013

[8] BERG, C. and CHRISTENSEN, J.P.R. and RESSEL, P. (1984). *Harmonic analysis on semigroups*. Springer, Berlin. MR0747302

[9] BERNHART, G. and MAI, J.-F. and SCHERER, M. (2015). On the construction of low-parametric families of min-stable multivariate exponential distributions in large dimensions. *Dependence Modeling* **3** 29–46. MR3418655

[10] BERNSTEIN, S. (1929). Sur les fonctions absolument monotones. *Acta Mathematica* **52** 1–66. MR1555269

[11] BEZGINA, E. and BURKSCHAT, M. (2019). On total positivity of exchangeable random variables obtained by symmetrization, with applications to failure-dependent lifetimes. *Journal of Multivariate Analysis* **169** 95–109. MR3875589

[12] BILLINGSLEY, P. (1995). *Probability and measure.* Wiley Series in Probability and Statistics, Wiley, New York. MR1324786

[13] BRIGO, D. and MAI, J.-F. and SCHERER, M. (2016). Markov multivariate survival indicators for default simulation as a new characterization of the Marshall–Olkin law *Statistics and Probability Letters* **114** 60–66. MR3491973

[14] CAPÉRAÀ, P. and FOUGÈRES, A.-L. and GENEST, C. (2000). Bivariate distributions with given extreme value attractor. *Journal of Multivariate Analysis* **72** 30–49. MR1747422

[15] CHARPENTIER, A. and FOUGÈRES, A.-L. and GENEST, C. and NEŠLEHOVÁ, J.G. (2014). Multivariate Archimax copulas. *Journal of Multivariate Analysis* **126** 118–136. MR3173086

[16] COSSETTE, H. and GADOURY, S.-P. and MARCEAUAND, E. and MTALAI, I. (2017). Hierarchical Archimedean copulas through multivariate compound distributions. *Insurance: Mathematics and Economics* **76** 1–13. MR3698183

[17] DABONI, L. (1982). *Exchangeability and completely monotone functions.* In: Exchangeability in Probability and Statistics, edited by G. Koch and F. Spizzichino, North-Holland Publishing Company 39–45. MR0675963

[18] DE FINETTI, B. (1931). Funzione caratteristica di un fenomeno aleatorio. *Atti della R. Academia Nazionale dei Lincei, Serie 6. Memorie, Classe di Scienze Fisiche, Mathematica e Naturale* **4** 251–299.

[19] DE FINETTI, B. (1937). La prévision: ses lois logiques, ses sources subjectives. *Annales de l'Institut Henri Poincaré* **7** 1–68. MR1508036

[20] DIACONIS, P. and FREEDMAN, D. (1987). A dozen de Finetti-style results in search of a theory. *Annales de l'Institute Henri Poincaré* **23** 397–423. MR0898502

[21] DICKINSON, P.J.C. and GIJBEN, L. (2014). On the computational complexity of membership problems for the completely positive cone and its dual. *Computational Optimization and Applications* **57:2** 403–415. MR3165055

[22] DURANTE, F. and QUESADA-MOLINA, J.J. and ÚBEDA-FLORES, M. (2007). *A method for constructing multivariate copulas.* In: New Dimensions in Fuzzy Logic and Related Technologies – Proceedings of the 5th EUSFLAT Conference, volume 1, edited by M. Štěpnička et al. 191–195.

[23] DURRETT, R. (2010). *Probability: theory and examples, 4th edition.* Cambridge University Press, Cambridge. MR2722836

[24] DYKSTRA, R.L. and HEWETT, J.E. and THOMPSON, JR., W.A. (1973). Events which are almost independent. *Annals of Statistics* **1:4** 674–681. MR0397815

[25] EMBRECHTS, P. and HOFERT, M. (2013). A note on generalized inverses. *Mathematical Methods of Operations Research* **77** 423–432. MR3072795

[26] ESARY, J.D. and MARSHALL, A.W. (1974). Multivariate distributions with exponential minimums. *Annals of Statistics* **2** 84–98. MR0362704

[27] ES-SEBAIY, K. and OUKNINE, Y. (2008). How rich is the class of processes which are infinitely divisible with respect to time. *Statistics and Probability Letters* **78** 537–547. MR2400867

[28] GIESECKE, K. (2003). A simple exponential model for dependent defaults. *Journal of Fixed Income* **13:3** 74–83.

[29] GUPTA, A.K. and NADARAJAH, S. (2004). *Handbook of beta distributions and its applications.* Marcel Dekker, New York. MR2079703

[30] HEWITT, E. and SAVAGE, L.J. (1955). Symmetric measures on Cartesian products. *Transactions of the American Mathematical Society* **80** 470–501. MR0076206

[31] FANG, K.-T. and KOTZ, S. and NG, K.-W. (1990). *Symmetric multivariate and related distributions.* Chapman and Hall, London. MR1071174

[32] FELLER, W. (1966). *An introduction to probability theory and its applications, volume II, 2nd edition.* John Wiley and Sons, Inc., Hoboken.

[33] FERGUSON, T.S. (1973). A Bayesian analysis of some nonparametric problems. *Annals of Statistics* **1** 209–230. MR0350949

[34] FERGUSON, T.S. (1974). Prior distributions on spaces of probability measures. *Annals of Statistics* **2** 615–629. MR0438568

[35] FRANK, M.J. (1979). On the simultaneous associativity of $F(x,y)$ and $x + y - F(x,y)$. *Aequationes Mathematicae* **19** 194–226. MR0556722

[36] GALAMBOS, J. (1975). Order statistics of samples from multivariate distributions. *Journal of the American Statistical Association* **70** 674–680. MR0405714

[37] GENEST, C. and NEŠLEHOVÁ, J.G. (2017). *When Gumbel met Galambos.* In: Copulas and Dependence Models With Applications: Contributions in Honor of Roger B. Nelsen (M. Úbeda Flores, E. de Amo Artero, F. Durante, J. Fernández Sánchez, Eds.), Springer, 83–93. MR3822198

[38] GENEST, C. and NEŠLEHOVÁ, J.G. and RIVEST, L.-P. (2018). The class of multivariate max-id copulas with ℓ_1-norm symmetric exponent measure. *Bernoulli* **24** 3751–3790. MR3788188

[39] GENEST, C. and RIVEST, L.-P. (1989). Characterization of Gumbel's family of extreme value distributions. *Statistics and Probability Letters* **8** 207–211. MR1024029

[40] GNEDIN, A.V. (1995). On a class of exchangeable sequences. *Statistics and Probability Letters* **25** 351–355. MR1363235

[41] GUMBEL, E.J. (1960). Bivariate exponential distributions. *Journal of the American Statistical Association* **55** 698–707. MR0116403

[42] GUMBEL, E.J. (1961). Bivariate logistic distributions. *Journal of the American Statistical Association* **56** 335–349. MR0158451

[43] HAKASSOU, A. and OUKNINE, Y. (2013). IDT processes and associated Lévy processes with explicit constructions. *Stochastics* **85:6** 1073–1111. MR3176501

[44] HAUSDORFF, F. (1921). Summationsmethoden und Momentfolgen I. *Mathematische Zeitschrift* **9:3-4** 74–109. MR1544453

[45] HAUSDORFF, F. (1923). Momentenproblem für ein endliches Intervall. *Mathematische Zeitschrift* **16** 220–248. MR1544592

[46] HERBERTSSON, A. and ROOTZÉN, H. (2008). Pricing kth-to-default swaps under default contagion: the matrix-analytic approach. *Journal of Computational Finance* **12** 49–72. MR2504900

[47] HERING, C. and HOFERT, M. and MAI, J.-F. and SCHERER, M. (2010). Constructing hierarchical Archimedean copulas with Lévy subordinators. *Journal of Multivariate Analysis* **101** 1428–1433. MR2609503

[48] HJORT, N.L. (1990). Nonparametric Bayes estimators based on beta processes in models for life history data. *Annals of Statistics* **18:3** 1259–1294. MR1062708

[49] HOFERT, M. and SCHERER, M. (2011). CDO pricing with nested Archimedean copulas. *Quantitative Finance* **11** 775–787. MR2800641

[50] H. Joe (1997). *Multivariate models and dependence concepts.* Chapman & Hall/CRC, Boca Raton. MR1462613

[51] KALBFLEISCH, J.D. (1978). Non-parametric Bayesian analysis of survival time data. *Journal of the Royal Statistical Society Series B* **40:2** 214–221. MR0517442

[52] KALLENBERG, O. (1982). *A dynamical approach to exchangeability.* In: Exchangeability in Probability and Statistics, edited by G. Koch and F. Spizzichino, North-Holland Publishing Company, 87–96. MR0675967

[53] KARLIN, S. and SHAPLEY, L.S. (1953). Geometry of moment spaces. *Memoirs of the American Mathematical Society* **12:93**. MR0059329

[54] KIMBERLING, C.H. (1974). A probabilistic interpretation of complete monotonicity. *Aequationes Mathematicae* **10** 152–164. MR0353416

[55] KINGMAN, J.F.C. (1967). Completely random measures. *Pacific Journal of Mathematics* **21:1** 59–78. MR0210185

[56] KINGMAN, J.F.C. (1972). On random sequences with spherical symmetry. *Biometrika* **59** 492–494. MR0343420

[57] KINGMAN, J.F.C. (1978). Uses of exchangeability. *Annals of Probability* **6:2** 183–197. MR0494344

[58] KONSTANTOPOULOS, T. and YUAN, L. (2019). On the extendibility of finitely exchangeable probability measures. *Transactions of the American Mathematical Society* **371** 7067–7092. MR3939570

[59] KOPP, C. and MOLCHANOV, I. (2018). Series representations of time-stable stochastic processes. *Probability and Mathematical Statistics* **38:2** 299–315. MR3896713

[60] LIGGETT, T.M. and STEIFF, J.E. and TÓTH, B. (2007). Statistical mechanical systems on complete graphs, infinite exchangeability, finite extensions and a discrete finite moment problem. *Annals of Probability* **35:3** 867–914. MR2319710

[61] LIJOI, A. and PRÜNSTER, I. and WALKER, S.G. (2008). Posterior analysis for some classes of nonparametric models. *Journal of Nonparametric Statistics* **20:5** 447–457. MR2424252

[62] Lindskog, F. and McNeil, A.J. (2003). Common Poisson shock models: applications to insurance and credit risk modelling. *ASTIN Bulletin* **33:2** 209–238. MR2035051

[63] Lukacs, E. (1955). A characterization of the gamma distribution. *Annals of Mathematical Statistics* **26** 319–324. MR0069408

[64] Mai, J.-F. (2018). Extreme-value copulas associated with the expected scaled maximum of independent random variables. *Journal of Multivariate Analysis* **166** 50–61. MR3799634

[65] Mai, J.-F. (2019). Simulation of hierarchical Archimedean copulas beyond the completely monotone case. *Dependence Modeling* **7** 202–214. MR3977499

[66] Mai, J.-F. (2020). Canonical spectral representation for exchangeable max-stable sequences. *Extremes* **23** 151–169. MR4064608

[67] Mai, J.-F. (2020). The de Finetti structure behind some norm-symmetric multivariate densities with exponential decay. *Dependence Modeling* **8** 210–220. MR4156799

[68] Mai, J.-F. and Schenk, S. and Scherer, M. (2016). Exchangeable exogenous shock models. *Bernoulli* **22** 1278–1299. MR3449814

[69] Mai, J.-F. and Schenk, S. and Scherer, M. (2016). Analyzing model robustness via a distortion of the stochastic root: a Dirichlet prior approach. *Statistics and Risk Modeling* **32** 177–195. MR3507979

[70] Mai, J.-F. and Schenk, S. and Scherer, M. (2017). Two novel characterizations of self-decomposability on the positive half-axis. *Journal of Theoretical Probability* **30** 365–383. MR3615092

[71] Mai, J.-F. and Scherer, M. (2009). Lévy-frailty copulas. *Journal of Multivariate Analysis* **100** 1567–1585. MR2514148

[72] Mai, J.-F. and Scherer, M. (2011). Reparameterizing Marshall–Olkin copulas with applications to sampling. *Journal of Statistical Computation and Simulation* **81** 59–78. MR2747378

[73] Mai, J.-F. and Scherer, M. (2012). H-extendible copulas. *Journal of Multivariate Analysis* **110** 151–160. MR2927515

[74] Mai, J.-F. and Scherer, M. (2014). Characterization of extendible distributions with exponential minima via processes that are infinitely divisible with respect to time. *Extremes* **17** 77–95. MR3179971

[75] Mai, J.-F. and Scherer, M. (2017). *Simulating copulas, 2nd edition.* World Scientific Publishing, Singapore. MR3729417

[76] Mai, J.-F. and Scherer, M. (2019). Subordinators which are infinitely divisible w.r.t. time: construction, properties, and simulation of max-stable sequences and infinitely divisible laws. *ALEA: Latin American Journal of Probability and Mathematical Statistics* **16:2** 977–1005. MR3999795

[77] Mai, J.-F. and Scherer, M. and Shenkman, N. (2013). Multivariate geometric laws, (logarithmically) monotone sequences, and infinitely divisible laws. *Journal of Multivariate Analysis* **115** 457–480. MR3004570

[78] Mansuy, R. (2005). On processes which are infinitely divisible with respect to time. *Working paper, arXiv:math/0504408.*

[79] Marshall, A.W. and Olkin, I. (1967). A multivariate exponential distribution. *Journal of the American Statistical Association* **62** 30–44. MR0215400

[80] Marshall, A.W. and Olkin, I. (1979). *Inequalities: theory of majorization and its applications.* Academic Press, New York. MR0552278

[81] McNeil, A.J. and Frey, R. and Embrechts, P. (2005). *Quantitative risk management.* Princeton University Press, Princeton. MR2175089

[82] McNeil, A.J. (2008). Sampling nested Archimedean copulas. *Journal of Statistical Computation and Simulation* **78** 567–581. MR2516827

[83] McNeil, A.J. and Nešlehová, J. (2009). Multivariate Archimedean copulas, d-monotone functions and l_1-norm symmetric distributions. *Annals of Statistics* **37:5B** 3059–3097. MR2541455

[84] McNeil, A.J. and Nešlehová, J. (2010). From Archimedean to Liouville copulas. *Journal of Multivariate Analysis* **101** 1772–1790. MR2651954

[85] Molchanov, I. (2008). Convex geometry of max-stable distributions. *Extremes* **11:3** 235–259. MR2429906

[86] Müller, A. and Stoyan, D. (2002). *Comparison methods for stochastic models and risks.* John Wiley and Sons, Chichester (2002). MR1889865

[87] Papangelou, F. (1989). On the Gaussian fluctuations of the critical Curie-Weiss model in statistical mechanics. *Probability Theory and Related Fields* **83** 265–278. MR1012501

[88] Pestman, W.R. (2009). *Mathematical statistics, 2nd edition.* De Gruyter, Berlin. MR2516478

[89] Puccetti, G. and Wang, R. (2015). Extremal dependence concepts. *Statistical Science* **30:4** 485–517. MR3432838

[90] Rachev, S.T. and Rüschendorf, L. (1991). Approximate independence of distributions on spheres and their stability properties. *Annals of Probability* **19** 1311–1337. MR1112418

[91] Resnick, S.I. (1987). *Extreme values, regular variation and point processes.* Springer-Verlag, Berlin. MR0900810

[92] Ressel, P. (1985). de Finetti type theorems: an analytical approach. *Annals of Probability* **13** 898–922. MR0799427

[93] Ryll-Nardzewski, C. (1957). On stationary sequences of random variables and the de Finetti equivalence. *Colloquium Mathematicum* **4** 149–156. MR0088823

[94] Sato, K.-I. (1999). *Lévy processes and infinitely divisible distributions.* Cambridge University Press, Cambridge. MR1739520

[95] Scarsini, M. (1985). Lower bounds for the distribution function of a k-dimensional n-extendible exchangeable process. *Statistics and Probability Letters* **3** 57–62. MR0792789

[96] Schilling, R. and Song, R. and Vondracek, Z. (2010). *Bernstein functions.* De Gruyter, Berlin. MR2978140

[97] Schoenberg, I.J. (1938). Metric spaces and positive definite functions. *Transactions of the American Mathematical Society* **44** 522–536. MR1501980

[98] SHAKED, M. (1977). A concept of positive dependence for exchangeable random variables. *Annals of Statistics* **5** 505–515. MR0436414

[99] SHAKED, M. and SPIZZICHINO, F. and SUTER, F. (2002). Nonhomogeneous birth processes and ℓ_∞-spherical densities, with applications in reliability theory. *Probability in the Engineering and Informational Sciences* **16** 271–288. MR1914427

[100] SIBLEY, D.A. (1971). A metric for weak convergence of distribution functions. *Rocky Mountain Journal of Mathematics* **1:3** 427–430. MR0314089

[101] SKLAR, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris* **8** 229–231. MR0125600

[102] SLOOT, H. (2020). The deFinetti representation of generalised Marshall–Olkin sequences. *Dependence Modeling* **8:1** 107–118. MR4121354

[103] SPIZZICHINO, F. (1982). *Extendibility of symmetric probability distributions and related bounds.* In: Exchangeability in Probability and Statistics, edited by G. Koch and F. Spizzichino, North-Holland Publishing Company, 313–320. MR0675986

[104] STEUTEL, F.W. and VAN HARN, K. (2003). *Infinite divisibility of probability distributions on the real line.* CRC Press, Boca Raton. MR2011862

[105] TALEB, N.N. (2020). *Statistical consequences of fat tails.* STEM Academic Press.

[106] WILLIAMSON, R.E. (1956). Multiply monotone functions and their Laplace transforms. *Duke Mathematical Journal* **23** 189–207. MR0077581

[107] ZHU, W. and WANG, C.-W. and TAN, K.S. (2016). Structure and estimation of Lévy subordinated hierarchical Archimedean copulas (LSHAC): theory and empirical tests. *Journal of Banking and Finance* **69** 20–36.