GRAPHICAL MODELS FOR INFINITE MEASURES WITH APPLICATIONS TO EXTREMES

BY SEBASTIAN ENGELKE[1,a], JEVGENIJS IVANOVS[2,b] AND KIRSTIN STROKORB[3,c]

[1]*Research Center for Statistics, University of Geneva, 1205 Geneva, Switzerland,* [a]*sebastian.engelke@unige.ch*

[2]*Department of Mathematics, Aarhus University, DK-8000 Aarhus C, Denmark,* [b]*jevgenijs.ivanovs@math.au.dk*

[3]*School of Mathematics, Cardiff University, Cardiff CF24 4AG, United Kingdom,* [c]*strokorbk@cardiff.ac.uk*

Conditional independence and graphical models are well studied for probability distributions on product spaces. We propose a new notion of conditional independence for any measure Λ on the punctured Euclidean space $\mathbb{R}^d \setminus \{0\}$ that explodes at the origin. The importance of such measures stems from their connection to infinitely divisible and max-infinitely divisible distributions, where they appear as Lévy measures and exponent measures, respectively. We characterize independence and conditional independence for Λ in various ways through kernels and factorization of a modified density, including a Hammersley–Clifford type theorem for undirected graphical models. As opposed to the classical conditional independence, our notion is intimately connected to the support of the measure Λ. Our general theory unifies and extends recent approaches to graphical modeling in the fields of extreme value analysis and Lévy processes. Our results for the corresponding undirected and directed graphical models lay the foundation for new statistical methodology in these areas.

CONTENTS

MSC2020 subject classifications: Primary 62H22, 60E07; secondary 60G70, 60G51.

Keywords and phrases: conditional independence, exponent measure, extremes, infinite divisibility, Lévy measure, stable distribution.

1. Introduction. Conditional independence is a central concept in probability theory that enables the definition of graphical models. It is a cornerstone of modern statistical methodology for understanding latent structures in the data, construction of parsimonious models in high dimensions and estimation of causal relationships (Dawid, 1979; Lauritzen, 1996; Maathuis et al., 2019). The classical notion of probabilistic conditional independence is defined for random vectors through factorizations of conditional probabilities. In fact, following an axiomatic approach, conditional independence can be seen as a more general notion of irrelevance with applications beyond random vectors (e.g., Lauritzen, 1996, Chapter 3).

In this paper we define a new notion of conditional independence and graphical models for the class of Borel measures Λ on the punctured d-dimensional Euclidean space $\mathcal{E} = \mathbb{R}^d \setminus \{0\}$, where Λ is finite on all Borel sets that are bounded away from the origin. For disjoint subsets A, B, C of the index set $V = \{1, \ldots, d\}$, we denote this conditional independence of A and B given C for the measure Λ by

$$A \perp B \mid C \; [\Lambda]. \tag{1}$$

The essence of our definition in Section 3.2 is that we require classical conditional independence on the normalized restrictions of Λ to any charged product form set in $\mathcal{E}$ that is bounded away from the origin.

Measures Λ of the above kind are indeed fundamental in various fields of probability and statistics. Notably, they arise as Lévy measures of infinitely divisible probability laws with respect to semi-group operations on $\mathbb{R}^d$ for which the origin is a neutral element (Berg, Christensen and Ressel, 1984). We focus here on the following important examples. A random vector X is (sum-)infinitely divisible or max-infinitely divisible, if for every $n \in \mathbb{N}$ it has the stochastic representation

$$X \stackrel{d}{=} \sum_{i=1}^{n} X^{(n,i)}, \quad \text{or} \quad X \stackrel{d}{=} \bigvee_{i=1}^{n} X^{(n,i)}, \tag{2}$$

respectively, for some independent, identically distributed vectors $X^{(n,1)}, \dots, X^{(n,n)}$, where the sums and maxima are taken in the component-wise sense. In the context of maxima, the term exponent measure is often preferred over Lévy measure for Λ and we adopt the same convention here (Resnick, 2008). Importantly, infinitely divisible distributions appear as limit laws of triangular arrays; for sums this observation dates back to Skorohod (1957) and Feller (1971, Part XVII.11), for instance, and we refer to Balkema and Resnick (1977) and de Haan and Resnick (1977) for maxima. This explains the ubiquity of measures Λ and their appearance in the vast literature on related point processes, random sets or large deviation principles (e.g., Davydov, Molchanov and Zuyev, 2008; Dombry and Ribatet, 2015; Hult and Lindskog, 2007). In most situations, including typical cases of sum- and max-infinitely divisible distributions, the measure Λ explodes at the origin and therefore has infinite mass on $\mathcal{E}$. Therefore, classical probabilistic conditional independence is not meaningful for such measures. Our new notion in (1) resolves this issue in an arguably natural way.

A first fundamental result for our new conditional independence is that, under a mild explosiveness assumption, the four axioms of a semi-graphoid are satisfied (Lauritzen, 1996, Chapter 3). This directly implies various useful properties of the corresponding graphical models. Moreover, we provide several alternative characterizations. If Λ possesses a density λ with respect to some product measure, then the conditional independence (1) is equivalent to the factorization of (a modified version of) this density. Under the mild explosiveness condition, (1) induces certain constraints on the support of the measure Λ. In particular, the independence $A \perp B \, [\Lambda]$ is completely characterized by the condition that Λ can have mass only on certain sub-faces of $\mathcal{E}$. This is in sharp contrast to usual probabilistic independence, which would suggest a factorization of the measure into the product of marginals. Graphical models for measures Λ as above can be defined readily by means of suitable Markov properties. We reveal several fundamental links between properties of the measure Λ and the graph structure. For instance, for an undirected, decomposable graph we prove a Hammersley–Clifford type theorem that states that the global Markov property for Λ is equivalent to a factorization of the modified density of Λ on the cliques of the graph.

The general theory of conditional independence and graphical models for Λ measures translates into new insights in the respective areas of probability theory, where such measures arise naturally. In particular, many of our findings in this manuscript have been motivated by recent developments and open questions from multivariate extreme value theory, which studies the dependence structure in the distributional tail of a random vector. The most fundamental limiting models that appear in this field are max-stable distributions, which are max-infinitely divisible distributions with homogeneous exponent measures Λ, that is,

$$\Lambda(tA) = t^{-\alpha}\Lambda(A), \qquad t > 0, \tag{3}$$

for some $\alpha > 0$ and all Borel sets $A \subset \mathcal{E}$ bounded away from the origin. While classical theory in this field intensively studied stochastic processes with spatial domains (e.g., de Haan, 1984; Kabluchko, Schlather and de Haan, 2009), recent research concentrates on finding sparse structures in multivariate extreme value models; see Engelke and Ivanovs (2021) for an overview. One line of work develops recursive max-linear models on directed acyclic graphs, which correspond to a specific class of spectrally discrete max-stable distributions (Gissibl and Klüppelberg, 2018; Klüppelberg and Lauritzen, 2019). However, if a max-stable distribution admits a positive density, it cannot satisfy non-trivial (classical) conditional independence (Papastathopoulos and Strokorb, 2016). Instead, Engelke and Hitz (2020) define a notion of extremal conditional independence and undirected graphical models on the level of multivariate Pareto distributions (Rootzén and Tajvidi, 2006), which can also be characterized by the homogeneous exponent measure Λ; see Section 8.1 for details. In the case of tree graphs, these models arise as extremal limits of regularly varying Markov trees (Segers, 2020).

These two active research directions have so far co-existed without a clear connection. Since max-linear models and multivariate Pareto distributions are both described by the exponent measure Λ, it seems natural that this object also encodes the graphical properties. Indeed, it turns out that our conditional independence notion in (1) unifies and extends both approaches. We show that for recursive max-linear models, any classical conditional independence implied by its directed acyclic graph is also present on the level of Λ in the sense of (1). Concerning the extremal conditional independence in Engelke and Hitz (2020), we note that it is originally only defined if Λ has a Lebesgue density on $\mathcal{E}$ and no mass on sub-faces. As a consequence, no extremal independence can exist and their extremal graphical models are always connected. Several discussion contributions[1] of their paper raised these points as limitations for statistical modeling. Our conditional independence (1) extends the notion of Engelke and Hitz (2020) to arbitrary exponent measures Λ and thereby overcomes several of these restrictions.

(i) It allows for disconnected extremal graphs with mass on sub-faces and, as an illustration, we provide an extension of Hüsler–Reiss tree models (Asenova, Mazo and Segers, 2021; Engelke and Volgushev, 2022) to forests with richer dependence structures. Since no densities are required, our notion also includes max-linear models.
(ii) It extends to max-infinitely divisible distributions that can exhibit asymptotic independence (Huser, Opitz and Thibaud, 2021), and therefore opens the door to graphical modeling in this large field of research, in which many approaches go back to the conditional approach for extreme value modelling by Heffernan and Tawn (2004).
(iii) The support constraints from the conditional independence of Λ connect our theory to sparsity notions for tails of multivariate distributions and the field of concomitant extremes (e.g., Chiapino and Sabourin, 2017).

We discuss these extensions in Section 8 and provide some basic examples that motivate the potential for the construction of new models. A more in-depth study of these new modelling routes and the development of corresponding statistical methodology is beyond the scope of this paper.

Another promising application of our theory lies in the field of infinitely divisible distributions and Lévy processes (Ivanovs, 2020), where Λ plays the role of the Lévy measure. Much less literature on graphical models and sparsity exists in this area. In the case of α-stable distributions, that is, if the Lévy measure is homogeneous as in (3), Misra and Kuruoglu (2016) introduce recursive stable models that are similar to Gissibl and Klüppelberg (2018). Again, the classical conditional independence is reflected in the corresponding (degenerate) Lévy measure Λ through our conditional independence. For more general Lévy measures, we can define graphical models with respect to our Λ-based notion of conditional independence. We

[1] More precisely, we refer to the contributions of J. Wadsworth, I. Papastathopoulos, K. Strokorb, C. Darne and A. C. Davison, L. Mhalla, P. Wan, Y. Zhang and L. Wang in the discussion part of Engelke and Hitz (2020) (here listed in order of appearance therein).

provide an outlook on how this translates into probabilistic properties of the corresponding Lévy processes in the concluding Section 9; the details of this direction are part of future research.

2. Preliminaries.

2.1. *Conditional independence.* Consider a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and let X, Y, Z be random elements with values in some Borel spaces. *Conditional independence* of X and Y given Z with respect to $\mathbb{P}$, denoted by

$$X \perp\!\!\!\perp Y \mid Z,$$

is defined as the factorization of conditional probabilities

$$\mathbb{P}(X \in E_X, Y \in E_Y | Z) = \mathbb{P}(X \in E_X | Z)\mathbb{P}(Y \in E_Y | Z), \qquad \text{a.s.,}$$

for all E_X, E_Y in the respective σ-algebras; we refer to Kallenberg (2002, Chapter 6) for the basic theory and further properties. In terms of the regular conditional probability $\mathbb{P}(\cdot | Z = z)$ it reads

$$\mathbb{P}(X \in E_X, Y \in E_Y | Z = z) = \mathbb{P}(X \in E_X | Z = z)\mathbb{P}(Y \in E_Y | Z = z), \qquad \text{for } \mathbb{P}_Z\text{-almost all } z,$$

where $\mathbb{P}_Z$ denotes the law of Z. By a standard argument (see Kallenberg, 2002, Thm. 6.3), this factorization holds for all such E_X, E_Y and all z outside some fixed $\mathbb{P}_Z$-null set. In other words, X and Y are independent under $\mathbb{P}(\cdot | Z = z)$ for $\mathbb{P}_Z$-almost all z.

When (X, Y, Z) possesses a joint density $f(x, y, z)$ with respect to some product measure $\mu = \mu_X \otimes \mu_Y \otimes \mu_Z$, where each component is a σ-finite measure, the conditional independence $X \perp\!\!\!\perp Y \mid Z$ is equivalent to the following density decomposition:

$$f(x, y, z) f_Z(z) = f_{XZ}(x, z) f_{YZ}(y, z) \qquad \text{for } \mu\text{-almost all } (x, y, z).$$

Here, we used the naturally induced notation for the marginal densities; see for instance Dawid (1979) or Lauritzen (1996, Chapter 3). In particular, if (X, Y, Z) is discrete, then its probability mass function factorizes for all (x, y, z). If μ is Lebesgue and all the involved densities are continuous, then this equality must be true for all (x, y, z). Note that continuity of f does not imply continuity of the marginal densities, which explains the formulation of the last sentence.

2.2. *Infinitely divisible distributions for sums and maxima.* For $d \geq 1$ consider the domain $\mathcal{E} = \mathbb{R}^d \setminus \{0\}$, which is a d-dimensional space of real numbers punctured at the origin. Let Λ be a Borel measure on this domain, such that the following basic boundedness condition is satisfied:

$$\text{(B)} \qquad \Lambda(A) < \infty \qquad \text{for any Borel set } A \subset \mathcal{E} \text{ bounded away from the origin.}$$

As discussed in the introduction, such measures appear naturally in various areas of probability and statistics. In particular, they arise in the limits of sums and maxima of random variables in the theory of infinitely divisible and max-infinitely divisible distributions as in (2).

In the max-infinitely divisible case we assume that each component X_j is non-negative with lower endpoint 0, that is, each X_j has some mass arbitrarily close to 0:

$$\inf\{x : \mathbb{P}(X_j \leq x) > 0\} = 0. \tag{4}$$

Then the distribution function of X is given by

$$\mathbb{P}(X \leq x) = \exp\big[-\Lambda(\mathcal{E}_+ \setminus [0, x])\big], \qquad x \geq 0, \tag{5}$$

where Λ is null outside of the domain $\mathcal{E}_+ = [0,\infty)^d \setminus \{0\}$ and satisfies (B), while $x \geq 0$ is meant componentwise; see Resnick (2008, Prop. 5.8). Conversely, every such measure Λ, called an *exponent measure*, leads to a max-infinitely divisible distribution. The standard assumption $\Lambda(\mathcal{E}_+) = \infty$ corresponds to no point mass of X at the origin. One may also define a Poisson point process with intensity measure $\mathrm{d}t \otimes \Lambda(\mathrm{d}x)$ and retrieve X as the maximum over the space-points with $t \leq 1$.

In the infinitely divisible case the characteristic function of X is given by

$$\mathbb{E}e^{\mathrm{i}\langle\theta,X\rangle} = \exp\left[-\frac{1}{2}\langle\theta,\Sigma\theta\rangle + \mathrm{i}\langle\gamma,\theta\rangle + \int_{\mathcal{E}}\left(e^{\mathrm{i}\langle\theta,x\rangle} - 1 - \mathrm{i}\langle\theta,x\rangle 1_{\{\|x\|\leq 1\}}\right)\Lambda(\mathrm{d}x)\right],\ \theta \in \mathbb{R}^d,$$

where $\gamma \in \mathbb{R}^d$ is a drift parameter, Σ is the symmetric, non-negative definite $d \times d$ covariance matrix of the Gaussian component, and Λ is the so-called *Lévy measure*, which satisfies (B) together with $\int_{\|x\|\in(0,1)} \|x\|^2 \Lambda(\mathrm{d}x) < \infty$, where $\|\cdot\|$ is an arbitrary norm on $\mathbb{R}^d$; see Sato (2013, Thm. 8.1). The Lévy measure is used to define a Poisson point process of jumps with intensity measure $\mathrm{d}t \otimes \Lambda(\mathrm{d}x)$, the compensated sum of which yields the non-Gaussian part of X in the limit sense. Conversely, every such characteristic triplet $(\gamma, \Sigma, \Lambda)$ leads to an infinitely divisible law.

EXAMPLE 2.1. Let $\Phi_\rho : \mathbb{R}^2 \to [0,1]$ be the distribution function of a bivariate normal distribution with correlation coefficient $\rho \in [0,1]$ and standard normal margins. When $\rho > 0$, the bivariate normal distribution is max-infinitely divisible (Resnick, 2008, Section 5.2), a property inherited by the corresponding log-normal distribution $\Phi_\rho \circ \log$. Thus, by (5)

$$\Lambda_\rho(\mathcal{E}_+ \setminus [0,x]) = -\log \Phi_\rho(\log x), \quad x \geq 0,$$

is a valid measure satisfying (B). In addition, it explodes at the origin.

Various connections between the infinitely divisible and max-infinitely divisible laws have been explored in Kabluchko (2009) and Wang and Stoev (2010). Davydov, Molchanov and Zuyev (2008) and Molchanov (2009) draw further attention to multivariate ℓ_p-infinitely divisible laws, or more generally, infinitely divisible laws with respect to semi-group operations. All of these works focus on the important special case where the involved Lévy measures (or exponent measures) Λ are $-\alpha$-*homogeneous*, that is, the property (3) holds for some $\alpha > 0$. A max-infinitely divisible X with (standard) α-Fréchet marginals is *max-stable* if and only if Λ is $-\alpha$-homogeneous (Resnick, 2008, Prop. 5.11). An infinitely divisible X is α-*stable* with $\alpha \in (0,2)$ if and only if $\Sigma = 0$ and Λ is $-\alpha$-homogeneous (Sato, 2013, Thm. 14.3(ii)).

EXAMPLE 2.2. The exponent measure corresponding to the bivariate max-stable Hüsler–Reiss distributions (Hüsler and Reiss, 1989) with parameter $\Gamma > 0$ is

$$\Lambda_\Gamma(\mathcal{E}_+ \setminus [0,x]) = -\frac{1}{x_1}\Phi\left(\frac{\sqrt{\Gamma}}{2} + \frac{\log(x_2/x_1)}{\sqrt{\Gamma}}\right) - \frac{1}{x_2}\Phi\left(\frac{\sqrt{\Gamma}}{2} + \frac{\log(x_1/x_2)}{\sqrt{\Gamma}}\right), \quad x \geq 0,$$

where Φ denotes the univariate standard normal cumulative distribution function. This measure also satisfies (B). Since it is -1-homogeneous, it also explodes at the origin.

3. Definition of conditional independence for Λ and first properties.

3.1. *Setup and notation.* As previously, let $V=\{1,\dots,d\}$ be a finite index set and Λ a measure on $\mathcal{E}=\mathbb{R}^V\setminus\{0_V\}$ satisfying (B). For a subset $A\subset\mathcal{E}$, we say that a property holds for Λ-almost all $y\in A$ if the subset of A where it does not hold is a Λ-null set. As test sets for conditional independence we consider the class $\mathcal{R}(\Lambda)$ of charged product-form sets in $\mathcal{E}$ not containing the origin 0_V in their closure

$$\mathcal{R}(\Lambda)=\Big\{R=\bigtimes_{v\in V}R_v\,:\,R_v\in\mathcal{B}(\mathbb{R}),\,\Lambda(R)>0,\,0_V\notin\overline{R}\Big\}. \tag{6}$$

By definition, we have $\Lambda(R)\in(0,\infty)$ for any $R\in\mathcal{R}(\Lambda)$, which allows us to define a probability measure $\mathbb{P}_R$ on R via

$$\mathbb{P}_R(\mathrm{d}y)=\frac{\Lambda(\mathrm{d}y)}{\Lambda(R)}. \tag{7}$$

We write $Y\sim\mathbb{P}_R$ for a d-dimensional random vector Y distributed according to $\mathbb{P}_R$.

For a non-empty subset D of V, the measure Λ induces two natural measures on the lower dimensional domain

$$\mathcal{E}^D=\mathbb{R}^D\setminus\{0_D\}, \tag{8}$$

namely the marginal measure Λ_D and the restricted measure Λ_D^0 as given by

$$\Lambda_D(\cdot)=\Lambda(y_D\in\cdot),\qquad\qquad \Lambda_D^0(\cdot)=\Lambda(y_D\in\cdot\,,\,y_{V\setminus D}=0_{V\setminus D}), \tag{9}$$

respectively. Both measures also satisfy the basic assumption (B) in their common domain $\mathcal{E}^D$.

The class $\mathcal{R}(\Lambda_D)$ is understood to contain product form sets $R_D=\bigtimes_{d\in D}R_d$ (satisfying the other conditions as in (6) with respect to $\mathbb{R}^D$ instead of $\mathbb{R}^V$) and the respective probability measures are denoted by $\mathbb{P}_{R_D}$. The product set R will sometimes be written as $R_D\times R_{V\setminus D}$ irrespective of the indices in D. Note that $R_D\in\mathcal{R}(\Lambda_D)$ implies $R_D\times\mathbb{R}^{V\setminus D}\in\mathcal{R}(\Lambda)$. Some basic consistency properties are stated in Appendix A.1.

3.2. *Definition of conditional independence for* Λ. We are now ready to give our definition of conditional independence for the measure Λ based on the above test sets.

DEFINITION 3.1. For disjoint sets $A,B,C\subset V$ that form a partition of V, we say that Λ admits *conditional independence* of A and B given C, denoted by

$$A\perp B\mid C\;[\Lambda],$$

if we have the classical conditional independence

$$Y_A\perp\!\!\!\perp Y_B\mid Y_C\qquad\text{for}\qquad Y\sim\mathbb{P}_R\qquad\text{for all}\qquad R\in\mathcal{R}(\Lambda). \tag{10}$$

This is trivially true for A or B being empty, and for $C=\emptyset$ we say that Λ admits *independence* of A and B, and write

$$A\perp B\;[\Lambda].$$

If the sets A,B and C are not a partition of V, then the above definition remains the same with the test class in (10) replaced by $\mathcal{R}(\Lambda_{A\cup B\cup C})$.

REMARK 3.2. If Λ is a zero measure, the test classes $\mathcal{R}(\Lambda_{A\cup B\cup C})$ are empty, hence any such (conditional) independence statement with respect to Λ is true. Moreover, for strictly monotone, continuous transformations $T_v:\mathbb{R}\to\mathbb{R}$ with $T_v(0)=0$, $v\in V$, the conditional independence of Λ in Definition 3.1 is equivalent to the corresponding statement for the pushforward measure $\Lambda\circ T^{-1}$ for $T(x)=(T_v(x_v))_{v\in V}$. Definition 3.1 is similar to the notion of conditional inner independence as suggested by Zhang and Wang (2020) for a random vector supported by a non-product-form space.

It is useful to observe that for $A\cup B\cup C\subset D\subset V$, the two statements

$$A\perp B\mid C\,[\Lambda] \qquad \text{and} \qquad A\perp B\mid C\,[\Lambda_D] \tag{11}$$

are equivalent. However, some caution is needed. We would like to stress that the conditional independence $A\perp B\mid C\,[\Lambda]$ does not(!) imply

$$Y_A \perp\!\!\!\perp Y_B\mid Y_C \qquad \text{for} \quad Y\sim\mathbb{P}_R \quad \text{for all} \quad R\in\mathcal{R}(\Lambda)$$

unless (A,B,C) is a partition of V, because restriction according to the additional dimensions in $V\setminus(A\cup B\cup C)$ amounts to conditioning on the respective components; see Lemma A.2 in Appendix A.1.

A second simple observation, which will accompany us throughout this work, is the following:

LEMMA 3.3. *$A\perp B\mid C\,[\Lambda]$ implies $A\perp B\,[\Lambda^0_{A\cup B}]$.*

PROOF. If $\Lambda^0_{A\cup B}$ is null, there is nothing to prove. Else consider $R_{A\cup B}\in\mathcal{R}(\Lambda^0_{A\cup B})$ and the respective $Y_{A\cup B}$. Set $Y_C=0_C$ and note that the law of such a vector Y alternatively arises from the measure Λ and the set $R_{A\cup B}\times\{0_C\}\in\mathcal{R}(\Lambda)$. Now $A\perp B\mid C\,[\Lambda]$ implies $Y_A\perp\!\!\!\perp Y_B\mid Y_C$ and thus also $Y_A\perp\!\!\!\perp Y_B$, because $Y_C=0_C$ a.s. ∎

REMARK 3.4. Slightly more general, the same reasoning shows that $A\perp B\mid C\,[\Lambda]$ implies $A\perp B\mid C'\,[\Lambda^0_{A\cup B\cup C'}]$ for any subset $C'\subset C$. When $C=\emptyset$, the statement in Lemma 3.3 is a tautology.

3.3. *Semi-graphoid properties.* A semi-graphoid is an abstract independence model that satisfies a set of algebraic properties (Lauritzen, 1996; Maathuis et al., 2019; Pearl, 1988). For our conditional independence defined in Section 3.2 for a measure Λ, if A,B,C,D are disjoint subsets of V, these properties read as follows.

(L1) If $A\perp B\mid C\,[\Lambda]$, then $B\perp A\mid C\,[\Lambda]$. (Symmetry)

(L2) If $A\perp (B\cup D)\mid C\,[\Lambda]$, then $A\perp B\mid C\,[\Lambda]$. (Decomposition)

(L3) If $A\perp (B\cup D)\mid C\,[\Lambda]$, then $A\perp B\mid C\cup D\,[\Lambda]$. (Weak union)

(L4) If $A\perp B\mid C\,[\Lambda]$ and $A\perp D\mid B\cup C\,[\Lambda]$, then $A\perp (B\cup D)\mid C\,[\Lambda]$. (Contraction)

These conditions are crucial in the study of graphical models as they directly imply certain equivalences or implications among fundamental Markov properties usually considered in directed and undirected graphical models; see Sections 6 and 7.

If we restrict our attention to homogeneous measures Λ that admit a Lebesgue density on $\mathcal{E}_+$ as considered in Engelke and Hitz (2020) in the context of extreme value theory, Steffen Lauritzen argues already in his discussion contribution that their extremal conditional independence is a semi-graphoid. We will confirm this finding below (Section 3.3) and even

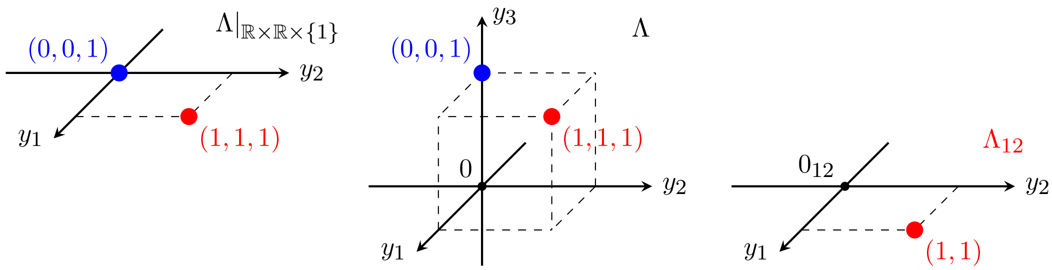

FIG 1. *Illustration of Example 3.6. The measure Λ consists of two point masses with different y_2-coordinate, so that $\{1\} \perp \{3\} \mid \{2\}$ $[\Lambda]$ (center). Its marginal measure Λ_{12} is a point mass (right), hence $\{1\} \perp \{2\}$ $[\Lambda]$. However, the restriction of Λ to the admissible product set $\mathbb{R} \times \mathbb{R} \times \{1\}$ does not factorize (left); the L4 conclusion $\{1\} \perp \{2, 3\}$ $[\Lambda]$ is violated.*

show that homogeneous measures Λ with a positive continuous Lebesgue-density on $\mathcal{E}_+$ (and independent concatenations thereof) give rise to a *graphoid* (Appendix E).

However, our general setting here goes far beyond this situation in several ways. Our Λ is not necessarily homogeneous, it does not necessarily have a density, its mass is not restricted to the positive upper orthant and it may even have mass on sub-faces of $\mathcal{E}$. This leads to a situation, where only the properties (L1)–(L3) always hold. The contraction condition (L4) is more involved.

PROPOSITION 3.5. *Conditional independence with respect to Λ as defined in Definition 3.1 satisfies (L1), (L2) and (L3).*

PROOF. Properties (L1) and (L3) follow directly from their respective counterparts for probability laws, whereas (L2) follows similarly if one takes into account the compatibility with marginal measures as stated in Lemma A.2. ∎

It will be shown in Theorem 5.3 below that also property (L4) is true under the explosiveness assumption (E1). This property is not true, however, for general finite measures Λ, in which case our Definition 3.1 might not be natural. The problem arises from the exclusion of the origin in the domain $\mathcal{E} = \mathbb{R}^V \setminus \{0_V\}$, which is neither necessary nor natural in the finite case. The following simple example illustrates this.

EXAMPLE 3.6 (Violation of (L4) for a finite measure). Consider $A = \{1\}, B = \{2\}, C = \emptyset, D = \{3\}$ and Λ putting mass 1 at each of the points $(0, 0, 1)$ and $(1, 1, 1)$; see Figure 1 for an illustration. Observe that $\Lambda_{A\cup B}$ is a point mass at $(1, 1)$ and so we trivially have $A \perp B$ $[\Lambda]$. Furthermore, we also have $A \perp D \mid B$ $[\Lambda]$, because the two point masses have a different B coordinate and so any admissible $\mathbb{P}_R$ results in the corresponding conditional independence statement. Finally, the statement $A \perp (B \cup D)$ $[\Lambda]$ is not true, because there exists a product set, namely $R = \mathbb{R} \times \mathbb{R} \times \{1\}$, for which we get that $Y \sim \mathbb{P}_R$ can only assume the values $(0, 0, 1)$ or $(1, 1, 1)$, each with probability $1/2$, and thus, Y_1 and (Y_2, Y_3) are not independent.

Even stronger properties than those of a semi-graphoid are represented by the notion of a graphoid (Lauritzen, 1996). The Λ-induced conditional independence defines a *graphoid*, when not only (L1)–(L4) are satisfied, but additionally the axiom

(L5) If $A \perp B \mid C \cup D\,[\Lambda]$ and $A \perp C \mid B \cup D\,[\Lambda]$, then $A \perp B \cup C \mid D\,[\Lambda]$ (Intersection)

In Appendix E, we document situations when the graphoid axioms are satisfied, and both trivial and subtle cases under which (L1)–(L4) hold, but (L5) is violated.

4. Alternative characterizations of Λ-conditional independence. We establish three alternative characterizations of the conditional independence $A \perp B \mid C\ [\Lambda]$ in this section, each of which further demonstrates that our notion is both natural and intuitive. The first characterization, Theorem 4.1, is a reduction to test sets of a simple form. The second, Theorem 4.4, is in terms of the underlying probability (Markov) kernel $\nu(y_C, \cdot)$ from $\mathcal{E}^C$ to $\mathbb{R}^{A\cup B}$. It allows to view our notion of conditional independence as the classical independence for a family of probability laws. The third, Theorem 4.6, is in terms of the factorization of the density of Λ with respect to some dominating measure, assuming it exists.

Without loss of generality, we focus on the case, where the sets A, B and C form a partition of V. The independence case $C = \emptyset$ is excluded, unless mentioned otherwise, but we return to it in Section 5. The more technical proofs and auxiliary results are given in Appendix A.

4.1. *Reduction of the test class.* Our first result reduces the test class $\mathcal{R} = \mathcal{R}(\Lambda)$ to sets of the form

$$R_{v,\varepsilon} = \{y \in \mathbb{R}^d : |y_v| \geq \varepsilon\} = (\mathbb{R} \setminus (-\varepsilon, \varepsilon))^{\{v\}} \times \mathbb{R}^{V\setminus\{v\}}, \qquad v \in V,\ \varepsilon > 0. \tag{12}$$

When writing $Y \sim \mathbb{P}_{R_{v,\varepsilon}}$, we implicitly assume $\Lambda(R_{v,\varepsilon}) > 0$ and hence $R_{v,\varepsilon} \in \mathcal{R}(\Lambda)$. The proof is postponed to Appendix A.

THEOREM 4.1 (CI via test classes). *Let (A, B, C) be a partition of V. Then $A \perp B \mid C\ [\Lambda]$ is equivalent to any of the following statements:*

(i) *$Y_A \perp\!\!\!\perp Y_B \mid Y_C$ with $Y \sim \mathbb{P}_{R_{v,\varepsilon}}$ for all $v \in V$, $\varepsilon > 0$,*
(ii) *$Y_A \perp\!\!\!\perp Y_B \mid Y_C$ with $Y \sim \mathbb{P}_{R_{c,\varepsilon}}$ for all $c \in C$, $\varepsilon > 0$, and additionally, $A \perp B\ [\Lambda^0_{A\cup B}]$.*

Moreover, $A \perp B\ [\Lambda]$ is equivalent to (i) with $C = \emptyset$.

REMARK 4.2. The characterizations (i) and (ii) in Theorem 4.1 can be modified in various ways. For instance in item (i) a single $v \in V$ is sufficient when $\Lambda^0_{V\setminus\{v\}}$ is a zero measure; see Lemma A.7 below. If the measure Λ is homogeneous, it suffices to consider a fixed $\varepsilon > 0$ instead of all $\varepsilon > 0$, e.g., $\varepsilon = 1$.

4.2. *Probability kernel.* We verify first that the measure Λ can be represented in terms of a probability kernel (Kallenberg, 2002, p. 20) when restricted to the set $\{y \in \mathcal{E} : y_C \neq 0\}$. To this end we use the notation that a set R is of *product form* if it can be represented as $R = \bigtimes_{v\in V} R_v = R_{A\cup B} \times R_C$ with $R_v \in \mathcal{B}(\mathbb{R})$ and $0_C \notin R_C$. We postpone the technical proof to Appendix A.

LEMMA 4.3 (Probability kernel representation). *There exists a Λ_C-unique probability kernel ν_C from $\mathcal{E}^C$ to $\mathbb{R}^{A\cup B}$ such that for all R of product form we have*

$$\Lambda(R) = \int_{y_C \in R_C} \nu_C(y_C, R_{A\cup B}) \Lambda_C(\mathrm{d}y_C) \in [0, \infty], \qquad \text{if } 0_C \notin R_C.$$

Moreover, for any $c \in C$ and $\varepsilon > 0$ such that $\Lambda(R_{c,\varepsilon}) > 0$ there is the identity

$$\nu_C(y_C, R_{A\cup B}) = \mathbb{P}(Y_{A\cup B} \in R_{A\cup B} | Y_C = y_C), \qquad Y \sim \mathbb{P}_{R_{c,\varepsilon}}$$

for Λ_C-almost all y_C with $|y_c| \geq \varepsilon$.

We can now characterize conditional independence as given by Definition 3.1 in terms of this kernel representation.

THEOREM 4.4 (CI via kernel factorization). *Let (A,B,C) be a partition of V. Then $A \perp B \mid C\;[\Lambda]$ if and only if the probability kernel factorizes,*

$$\nu_C(y_C, R_{A\cup B}) = \nu_C(y_C, R_A \times \mathbb{R}^B)\nu_C(y_C, R_B \times \mathbb{R}^A)$$

for Λ_C-almost all $y_C \neq 0_C$ and any $R_{A\cup B}$ of product form, and additionally $A \perp B\;[\Lambda^0_{A\cup B}]$.

PROOF. According to Theorem 4.1 (ii) it is sufficient to show that $Y_A \perp\!\!\!\perp Y_B \mid Y_C$ with $Y \sim \mathbb{P}_{R_{c,\varepsilon}}$ for all $c \in C$, $\varepsilon > 0$ if and only if the kernel factorizes for Λ_C-almost all $y_C \neq 0_C$. Fix $c \in C$, $\varepsilon > 0$ and note that $Y_A \perp\!\!\!\perp Y_B \mid Y_C$ is equivalent to

$$\mathbb{P}(Y_{A\cup B} \in R_{A\cup B} | Y_C = y_C) = \mathbb{P}(Y_A \in R_A | Y_C = y_C)\mathbb{P}(Y_B \in R_B | Y_C = y_C)$$

for $\mathbb{P}_{Y_C}$-almost all y_C and any product form $R_{A\cup B}$. These are Λ_C-almost all y_C satisfying $|y_c| \geq \varepsilon$ and the factorization can be stated in terms of the kernel ν_C according to its representation in Lemma 4.3. This factorization readily extends to Λ_C-almost all $y_C \neq 0_C$, since $c \in C$, $\varepsilon > 0$ are arbitrary, which completes the proof. ■

4.3. *Density factorization.* An important case concerns the measure Λ having density λ on $\mathcal{E}$ with respect to some product measure

$$\mu = \bigotimes_{v\in V} \mu_v,$$

where each μ_v is a σ-finite measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. For any non-empty $D \subset V$, the marginal density λ_D, the density of Λ_D in (9), with respect to $\mu_D = \bigotimes_{v\in D} \mu_v$ is

$$\lambda_D(y_D) = \int \lambda(y_D, y_{V\setminus D})\mu_{V\setminus D}(\mathrm{d}y_{V\setminus D}), \qquad y_D \neq 0_D,$$

which must be finite for μ_D-almost all $y_D \neq 0_D$. We also note that the density λ^0_D of Λ^0_D with respect to μ_D is given by

$$\lambda^0_D(y_D) = \lambda(y_D, 0_{V\setminus D})\mu_{V\setminus D}(\{0_{V\setminus D}\}), \qquad y_D \neq 0_D.$$

LEMMA 4.5 (Kernel density). *Assume that Λ has a μ-density λ. Then the probability kernel $\nu(y_C, \cdot)$ has density*

$$f_{y_C}(y_{A\cup B}) = \frac{\lambda(y)}{\lambda_C(y_C)}$$

with respect to $\mu_{A\cup B}$ for Λ_C-almost all $y_C \neq 0_C$.

PROOF. Take any R of product form such that $0_C \notin R_C$, and observe that

$$\Lambda(R) = \int_{y_C \in R_C} \Big(\int_{y_{A\cup B} \in R_{A\cup B}} \frac{\lambda(y)}{\lambda_C(y_C)} \mu_{A\cup B}(\mathrm{d}y_{A\cup B}) \Big) \Lambda_C(\mathrm{d}y_C).$$

Hence the expression in brackets coincides with $\nu(y_C, R_{A\cup B})$ for all $R_{A\cup B}$ and Λ_C-all $y_C \neq 0_C$, see Lemma 4.3. ■

This readily yields another characterization result of our conditional independence (Definition 3.1 with $C \neq \emptyset$).

THEOREM 4.6 (CI via density factorization). *Let (A,B,C) be a partition of V and assume that Λ has a μ-density λ. Then $A \perp B \mid C\;[\Lambda]$ if and only if the factorization*

$$\lambda(y)\lambda_C(y_C) = \lambda_{A\cup C}(y_{A\cup C})\lambda_{B\cup C}(y_{B\cup C})$$

holds for μ-almost all $y \in \mathcal{E}$, $y_C \neq 0_C$, and additionally, $A \perp B\;[\Lambda^0_{A\cup B}]$.

PROOF. According to Theorem 4.4 it will suffice to establish equivalence to the kernel factorization. By Lemma 4.5 kernel factorization is equivalent to $\mu_{A\cup B}$-almost everywhere factorization of the density f_{y_C} for Λ_C-almost all $y_C \neq 0_C$. The marginal density of f_{y_C} with respect to A is given by $\lambda_{A\cup C}(y_{A\cup C})/\lambda_C(y_C)$ and the analogous result holds for the set B. Thus, we have

$$\frac{\lambda(y)}{\lambda_C(y_C)} = \frac{\lambda_{A\cup C}(y_{A\cup C})}{\lambda_C(y_C)}\frac{\lambda_{B\cup C}(y_{B\cup C})}{\lambda_C(y_C)}$$

for $\mu_{A\cup B}\otimes\Lambda_C$-almost all $(y_{A\cup B}, y_C)$ with $y_C \neq 0_C$. It is left to note that the set $\{y : \lambda_C(y_C) = 0, \lambda_{A\cup C}(y_{A\cup C}) > 0\}$ is μ-negligible. ■

Importantly, we do not claim that the independence statement $A \perp B\,[\Lambda]$ implies the factorization $\lambda(y) = \lambda_A(y_A)\lambda_B(y_B)$. The marginals λ_A and λ_B are not even defined for $y_A = 0_A$ or $y_B = 0_B$, respectively, and considering a restriction to $y_A \neq 0_A$, $y_B \neq 0_B$ does not resolve the matter either; see Section 5.3 below. Instead, as will be also shown in Section 5.3, one needs to consider a modified density in order to achieve an equivalence to a (modified) density factorization.

5. Independence and the support of Λ.

5.1. *Independence characterization via the support of* Λ. Throughout this section we make the following explosiveness assumption

(E0) $$\Lambda\big(y_v \neq 0\big) = \Lambda_{\{v\}}\big(\mathcal{E}^{\{v\}}\big) = \Lambda\big(\mathcal{E}^{\{v\}}\times\mathbb{R}^{V\setminus\{v\}}\big) \in \{0,\infty\} \qquad \text{for all } v \in V,$$

which means that all one-dimensional marginal measures $\Lambda_{\{v\}}$ of Λ are either infinite or zero measures. It is much less restrictive than imposing homogeneity on the measure Λ; however note that (E0) is always met for a $-\alpha$-homogeneous Λ, since $\mathcal{E}^{\{v\}}\times\mathbb{R}^{V\setminus\{v\}}$ is a scale-invariant set.

PROPOSITION 5.1 (Independence). *Consider a partition* (A,B) *of* V *and assume* (E0). *Then*

$$A \perp B\,[\Lambda] \qquad \Longleftrightarrow \qquad \Lambda(y_A \neq 0_A \text{ and } y_B \neq 0_B) = 0.$$

PROOF. The if statement follows trivially from the definition, since any $Y \sim \mathbb{P}_R$ for an admissible $R \in \mathcal{R}(\Lambda)$ is such that either $Y_A = 0_A$ a.s. or $Y_B = 0_B$ a.s. (the assumption (E0) is not needed for this direction).

Now assume $A \perp B\,[\Lambda]$ and that the support of Λ contains some y with $y_a \neq 0$ and $y_b \neq 0$ for some $a \in A$, $b \in B$. Note that $\Lambda_{\{a\}}$ is infinite by (E0), since it can not be a zero measure. Furthermore, we may take $\delta > 0$ small such that $\Lambda(R_{a,\delta}\cap R_{b,\delta}) > 0$. Letting $Y \sim \mathbb{P}_{R_{a,\varepsilon}}$ with $\varepsilon \in (0,\delta)$ we observe that

$$\frac{\Lambda(R_{a,\delta}\cap R_{b,\delta})}{\Lambda(R_{a,\varepsilon})} = \mathbb{P}(|Y_a|\geq\delta, |Y_b|\geq\delta) = \mathbb{P}(|Y_a|\geq\delta)\mathbb{P}(|Y_b|\geq\delta) = \frac{\Lambda(R_{a,\delta})}{\Lambda(R_{a,\varepsilon})}\frac{\Lambda(R_{b,\delta}\cap R_{a,\varepsilon})}{\Lambda(R_{a,\varepsilon})}.$$

Multiply both sides by $\Lambda(R_{a,\varepsilon}) \in (0,\infty)$ and observe that $\Lambda(R_{b,\delta}\cap R_{a,\varepsilon})$ stays bounded, whereas $\Lambda(R_{a,\varepsilon}) \to \Lambda(y_a \neq 0) = \infty$ as $\varepsilon \downarrow 0$. Hence $\Lambda(R_{a,\delta}\cap R_{b,\delta}) = 0$, a contradiction. ■

Proposition 5.1 is closely related to the concept of asymptotic independence in extremes, which we discuss in Section 8. We may rephrase the condition in Proposition 5.1 in a number of ways, for instance in terms of the sets

(13) $$\mathcal{H}_D = \{y \in \mathcal{E} : y_{V\setminus D} = 0_{V\setminus D}\} = \bigcup_{\emptyset\neq A\subset D} \mathcal{E}_A,$$

where

(14) $$\mathcal{E}_A = \{y \in \mathcal{E} : y_i \neq 0\ \forall i \in A \text{ and } y_i = 0 \forall i \in V \setminus A\} = (\mathbb{R} \setminus \{0\})^A \times \{0_{V \setminus A}\}$$

is the sub-face of $\mathcal{E}$ corresponding to the subset $A \subset V$; it should not be confused with $\mathcal{E}^A = \mathbb{R}^A \setminus \{0_A\}$. For convenience we summarize a few of these reformulations as follows. Figure 2 provides an illustration.

COROLLARY 5.2. *Consider a partition (A, B) of V and assume (E0). The following statements are equivalent.*

(i) $A \perp B\,[\Lambda]$
(ii) $\Lambda(\mathcal{E}^A \times \mathcal{E}^B) = \Lambda(y_A \neq 0_A \textit{ and } y_B \neq 0_B) = 0$
(iii) $\mathrm{supp}(\Lambda) \subset \mathcal{H}_A \cup \mathcal{H}_B$
(iv) $\Lambda_A = \Lambda_A^0$
(v) $\Lambda_B = \Lambda_B^0$

The independence characterization of Proposition 5.1 (or its extended reformulation Corollary 5.2) only treats the case where (A, B) is a partition of V. Suppose that this is not the case and denote by (A, B, C) a partition of V. We recall that by (11), $A \perp B\,[\Lambda]$ is equivalent to $A \perp B\,[\Lambda_{A\cup B}]$ and that (E0) for Λ implies (E0) for $\Lambda_{A\cup B}$. Thus, applying Proposition 5.1 to the measure $\Lambda_{A\cup B}$, we obtain

(15) $$A \perp B\,[\Lambda] \iff \Lambda(y_A \neq 0_A,\, y_B \neq 0_B) = 0 \iff \mathrm{supp}(\Lambda) \subset \mathcal{H}_{A\cup C} \cup \mathcal{H}_{B\cup C}.$$

Due to Lemma 3.3, we will often work with the statement $A \perp B\,[\Lambda^0_{A\cup B}]$. If (E0) holds for the reduced measure $\Lambda^0_{A\cup B}$, Proposition 5.1 translates into the equivalence

$$\begin{aligned} A \perp B\,[\Lambda^0_{A\cup B}] \quad &\iff \quad \Lambda(y_A \neq 0_A,\, y_B \neq 0_B,\, y_C = 0_C) = 0 \\ &\iff \quad \mathrm{supp}(\Lambda) \subset \mathcal{H}_A \cup \mathcal{H}_B \cup (\mathcal{E} \setminus \mathcal{H}_{A\cup B}). \end{aligned}$$

When translating items (iv) and (v) from Corollary 5.2 to these situations, where A and B are not a partition of V, but only disjoint, it is useful to take Lemma A.1 from Appendix A.1 into account.

5.2. *Contraction semi-graphoid property (L4).* We are now in a position to establish the remaining contraction semi-graphoid property (L4). In order to do so, we need to strengthen the explosiveness assumption (E0), so that it also holds for all underlying restricted measures:

(E1) $$\begin{aligned} \Lambda_D^0(y_d \neq 0) &= (\Lambda_D^0)_{\{d\}}(\mathcal{E}^{\{d\}}) \\ &= \Lambda(\mathcal{E}^{\{d\}} \times \mathbb{R}^{D\setminus\{d\}} \times \{0_{V\setminus D}\}) \in \{0, \infty\} \qquad \text{for all } d \in D \subset V. \end{aligned}$$

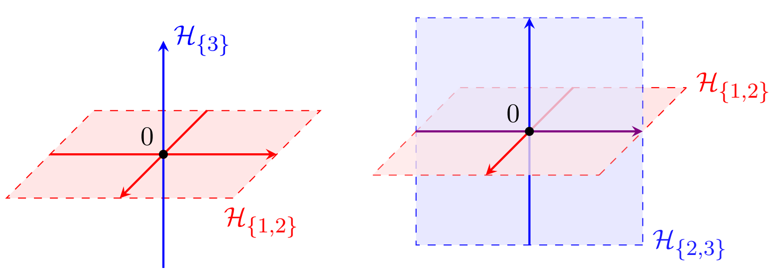

FIG 2. *Illustration of the sets $\mathcal{H}_D$ for $V = \{1, 2, 3\}$. No mass of Λ outside of $\mathcal{H}_{\{1,2\}}$ and $\mathcal{H}_{\{3\}}$ corresponds to $\{1, 2\} \perp \{3\}\,[\Lambda]$, see Proposition 5.1. No mass of Λ outside of $\mathcal{H}_{\{1,2\}}$ and $\mathcal{H}_{\{2,3\}}$ corresponds to $\{1\} \perp \{3\}\,[\Lambda]$, see (15).*

Assumption (E1) means that all one-dimensional marginals $(\Lambda_D^0)_{\{d\}}$ of all restricted measures Λ_D^0 (including $\Lambda = \Lambda_V^0$) are either infinite or zero measures. This necessarily entails that also all one-dimensional marginals $(\Lambda_D)_{\{d\}} = \Lambda_{\{d\}}$ of all marginal measures Λ_D are infinite or zero measures, hence (E0).

THEOREM 5.3. *Under Assumption* (E1) *conditional independence with respect to* Λ *as defined in Definition 3.1 satisfies all four semi-graphoid properties (L1)–(L4).*

PROOF. Properties (L1)–(L3) are already established in Proposition 3.5. To prove (L4), it is sufficient to assume that the union of A, B, C, D is V; see Lemma A.2.

First we establish (L4) for $C = \emptyset$. That is, we show that for a partition (A, B, D) of V the statements $A \perp B\,[\Lambda]$ and $A \perp D \mid B\,[\Lambda]$ imply $A \perp B \cup D\,[\Lambda]$. According to Proposition 5.1 we need to show that Λ has no mass outside of $\mathcal{H}_A \cup \mathcal{H}_{B\cup D}$. Note that $A \perp B\,[\Lambda]$ implies that there is no mass outside of $\mathcal{H}_{A\cup D} \cup \mathcal{H}_{B\cup D}$, whereas $A \perp D \mid B\,[\Lambda]$ implies $A \perp D\,[\Lambda^0_{A\cup D}]$, see Lemma 3.3, and thus all the mass of Λ on $\mathcal{H}_{A\cup D}$ must be contained in $\mathcal{H}_A \cup \mathcal{H}_D$. Summarizing, $\mathrm{supp}(\Lambda) \subset (\mathcal{H}_A \cup \mathcal{H}_D) \cup \mathcal{H}_{B\cup D} = \mathcal{H}_A \cup \mathcal{H}_{B\cup D}$, completing the first part of the proof. Here we used the explosiveness assumption (E0) for the measures $\Lambda_{A\cup B}$ and $\Lambda^0_{A\cup D}$, which follows from (E1).

Finally we prove the case $C \neq \emptyset$. As a consequence of Lemma 3.3 and Lemma A.1 we see that $A \perp B \mid C\,[\Lambda]$ implies $A \perp B\,[\Lambda^0_{A\cup B\cup D}]$, whereas $A \perp D \,|\, B \cup C\,[\Lambda]$ implies $A \perp D \,|\, B\,[\Lambda^0_{A\cup B\cup D}]$ (see Remark 3.4). Hence from the first step we readily get $A \perp (B \cup D)\,[\Lambda^0_{A\cup B\cup D}]$. According to Theorem 4.1 (ii) it is left to consider $Y \sim \mathbb{P}_{R_{c,\varepsilon}}$ for $c \in C$, $\varepsilon > 0$, and to establish $Y_A \perp\!\!\!\perp Y_{B\cup D} \mid Y_C$. But the assumption in (L4) implies $Y_A \perp\!\!\!\perp Y_B \,|\, Y_C$ (see Lemma A.2) and $Y_A \perp\!\!\!\perp Y_D \mid Y_{B\cup C}$, and we apply the classical analogue of (L4) to complete the proof. ∎

REMARK 5.4. At first sight, the explosiveness assumption (E1) may seem restrictive. However, it is a very natural condition for exponent measures and Lévy measures in the context of infinitely divisible distributions; see Section 2.2. Importantly, for a $-\alpha$-homogeneous Λ as in (3), all measures Λ_D^0 are $-\alpha$-homogeneous and so Assumption (E1) is satisfied. In addition, the necessity of this assumption (see Example 3.6) underlines the special role of the origin in our setting, which is not apparent through the first three properties (L1)–(L3).

5.3. *A modified density.* Here we revisit the case when Λ has a density λ with respect to a product measure $\mu = \bigotimes_{v\in V} \mu_v$. Let us first review the independence situation. Under assumption (E0) the independence $A \perp B\,[\Lambda]$ is equivalent to $\Lambda(y_A \neq 0_A, y_B \neq 0_B) = 0$. In this case

$$\lambda_D(y_D) = \lambda_D^0(y_D), \qquad \text{for } \mu_D\text{-almost all } y_D \in \mathcal{E}^D \text{ for } D \in \{A, B\},$$

whereas $\lambda(y) = 0$ for μ-almost all $y = (y_A, y_B) \in \mathcal{E}^A \times \mathcal{E}^B$. So, unless we have the degenerate situation that the full measure Λ is entirely concentrated either on $\mathcal{H}_A$ only or on $\mathcal{H}_B$ only, a factorization of the form $\lambda(y) = \lambda_A(y_A)\lambda_B(y_B)$ for μ-almost all $y \in \mathcal{E}^A \times \mathcal{E}^B$ cannot hold true. Hence the density does not factorize on the product set where the marginals are defined.

However, with a slight modification of the density, we can achieve a factorization, as long it is tested only on the correct subsets of $\mathbb{R}^V$. To this end, we will assume $\mu_v(\{0_v\}) > 0$ for all $v \in V$, and we define for a non-empty $D \subset V$ the modified densities

$$\bar{\lambda}_D(y_D) = \begin{cases} \lambda_D(y_D) & \text{for } y_D \in \mathcal{E}^D = \mathbb{R}^D \setminus \{0_D\}, \\ 1/\mu_D(\{0_D\}) & \text{for } y_D = 0_D, \end{cases} \tag{16}$$

whereas $\overline{\lambda}_\emptyset = 1$. In particular, $\lambda(y) = \overline{\lambda}(y)$ for all $y \in \mathcal{E} = \mathbb{R}^V \setminus \{0_V\}$. As a special case of the following result we will see that $A \perp B\ [\Lambda]$ is equivalent to the factorization

$$\overline{\lambda}(y) = \overline{\lambda}_A(y_A)\overline{\lambda}_B(y_B)$$

for μ-almost all $y \in \mathcal{E}$ with either $y_A = 0_A$ or $y_B = 0_B$.

PROPOSITION 5.5. *Assume (E1) and that Λ has a μ-density λ, where $\mu_v(\{0_v\}) > 0$ for all $v \in V$. Then for a partition (A, B, C) of V with possibly empty C the conditional independence $A \perp B \mid C\ [\Lambda]$ is equivalent to the factorization*

$$\overline{\lambda}(y)\overline{\lambda}_C(y_C) = \overline{\lambda}_{A\cup C}(y_{A\cup C})\overline{\lambda}_{B\cup C}(y_{B\cup C}) \tag{17}$$

for μ-almost all $y \in \mathbb{R}^V \setminus \{y : y_A \neq 0_A, y_B \neq 0_B, y_C = 0_C\}$.

PROOF. For $y = 0_V$ the statement (17) is trivial; it follows simply from the factorization of $\mu(\{0_V\})$ into the factors $\mu_v(\{0_v\})$, $v \in V$. On the set $\{y_C \neq 0_C\}$ the factorization (17) is that of λ as in Theorem 4.6. In view of Theorem 4.6 it is therefore sufficient to check that $A \perp B\ [\Lambda^0_{A\cup B}]$ is equivalent to the factorization of $\overline{\lambda}$ for μ-almost all $y \neq 0_V$ with $y_{A\cup C} = 0_{A\cup C}$ or $y_{B\cup C} = 0_{B\cup C}$.

Indeed, according to Corollary 5.2 the independence $A \perp B\ [\Lambda^0_{A\cup B}]$ is equivalent to $\Lambda(y_A \neq 0_A, y_B \neq 0_B, y_C = 0_C) = 0$, which holds if and only if

$$\lambda_{A\cup C}(y_{A\cup C}) = \lambda^0_A(y_A)/\mu_C(\{0_C\}) = \lambda(y)\mu_B(\{0_B\}) \tag{18}$$

for μ-almost all $y \neq 0_V$ with $y_{B\cup C} = 0_{B\cup C}$, and the analogous statement with exchanged roles of A and B holds. The equality (18) can be rewritten as

$$\lambda(y)/\mu_C(\{0_C\}) = \lambda_{A\cup C}(y_{A\cup C})/\mu_{B\cup C}(\{0_{B\cup C}\}),$$

which is the claimed factorization of $\overline{\lambda}$ on $\{y \in \mathcal{E} : y_{B\cup C} = 0_{B\cup C}\}$. The same argument applies with roles of A and B interchanged, and the proof is complete. ∎

REMARK 5.6. If $A \perp B \mid C\ [\Lambda]$, then $\Lambda(\{y : y_A \neq 0_A, y_B \neq 0_B, y_C = 0_C\}) = 0$. Proposition 5.5 does not need the entire extended explosiveness assumption (E1). It is sufficient to assume (E0) for the reduced measure $\Lambda^0_{A\cup B}$ instead, which is guaranteed by (E1).

5.4. *A generic construction.* To conclude this section we give a constructive example of a measure Λ in $d = 3$ dimensions with index set $V = \{1, 2, 3\}$ such that the conditional independence $\{1\} \perp \{3\} \mid \{2\}\ [\Lambda]$ holds. As a dominating product measure μ we take

$$\mu(\mathrm{d}x_1, \ldots, \mathrm{d}x_d) = \bigotimes_{i=1}^d \mu_i(\mathrm{d}x_i) = \bigotimes_{i=1}^d (\mathrm{d}x_i + \delta_0(\mathrm{d}x_i)), \tag{19}$$

where δ_0 is the point mass at 0. As building blocks, we employ two bivariate densities κ_{ij} for $(i, j) \in \{(1, 2), (2, 3)\}$ defined on $(0, \infty)^2$ with identical univariate marginal densities $\kappa_i(y_i) = m(y_i)$, $y_i \in (0, \infty)$, for all $i \in V$. As in (B) we further assume that the marginal measure on $(0, \infty)$ defined by the density m puts finite mass on sets bounded away from 0. For each of these bivariate densities κ_{ij} we can add mass on the axes to define a density with respect to the measure $\mu_{ij} = \mu_i \otimes \mu_j$ on $\mathcal{E}^{ij}_+ = [0, \infty)^{\{i,j\}} \setminus \{0_{\{i,j\}}\}$ by

$$\lambda_{ij}(y_{ij}) = p_{ij}\mathbf{1}_{\{y_i, y_j \neq 0\}}\kappa_{ij}(y_{ij}) + q_{ij}\mathbf{1}_{\{y_j = 0\}}m(y_i) + q_{ij}\mathbf{1}_{\{y_i = 0\}}m(y_j), \quad y_{ij} \in \mathcal{E}^{ij}_+, \tag{20}$$

where $p_{ij} \in [0, 1]$ is a mixture probability and $q_{ij} = 1 - p_{ij}$; the case $p_{ij} = 0$ implies independence between components i and j in the sense of Proposition 5.1.

By construction, the marginal densities of λ_{ij} also satisfy $\lambda_i(y_i) = m(y_i)$ for $i \in V$. We can combine these two bivariate densities using the modified densities in (16) into a trivariate density

$$\lambda(y) = \overline{\lambda}_{12}(y_{12})\overline{\lambda}_{23}(y_{23})/\overline{\lambda}_2(y_2), \tag{21}$$

for any $y \in \mathcal{E}_+ \setminus \{y : y_1 \neq 0, y_3 \neq 0, y_2 = 0\}$, and $\lambda(y) = 0$ otherwise. To be more precise, we can write this out in terms of the densities λ_{ij} and obtain for any $y \in \mathcal{E}_+$

$$\begin{aligned} \lambda(y) &= \lambda_{12}(y_{12})\lambda_{23}(y_{23})/m(y_2), && y_2 \neq 0, \\ \lambda(y_1, 0, 0) &= \lambda_{12}(y_1, 0), && y_1 \neq 0, \\ \lambda(0, 0, y_3) &= \lambda_{23}(0, y_3), && y_3 \neq 0, \\ \lambda(y_1, 0, y_3) &= 0, && y_1 \neq 0,\ y_3 \neq 0. \end{aligned}$$

The corresponding measure is determined for any Borel set $A \subset \mathcal{E}_+$, bounded away from the origin, by

$$\Lambda(A) = \int_A \lambda(y)\mu(\mathrm{d}y).$$

By construction, Λ has identical marginals $\Lambda(y_i > u) = \int_u^\infty m(y)\mathrm{d}y$, $u > 0$, $i \in V$, and it satisfies the finiteness condition (B). Indeed, for any set $A \subset \mathcal{E}_+$ bounded away from the origin there exists $\epsilon > 0$ such that $A \subset L_\epsilon = \mathcal{E}_+ \setminus [0, \epsilon]^d$, and therefore

$$\Lambda(A) \leq \Lambda(L_\epsilon) \leq \sum_{i \in V} \Lambda(y_i > \epsilon) = d \int_\epsilon^\infty m(y)\mathrm{d}y < \infty.$$

The measure Λ also satisfies the conditional independence $\{1\} \perp \{3\} \mid \{2\}\ [\Lambda]$ according to Proposition 5.5. If $0 < p_{12}, p_{23} < 1$ then the density of Λ puts mass on every sub-face of $\mathcal{E}_+ = [0, \infty)^3 \setminus \{0\}$ except for $\{y : y_1 > 0, y_2 = 0, y_3 > 0\}$. Indeed, the latter is the only sub-face that cannot contain mass in this situation if Λ satisfies the explosiveness assumption (E1); see Lemma 3.3 and Proposition 5.1. Further details and an illustration of this construction are provided in Appendix B. Concrete examples of this construction involving either the homogeneous measure from Example 2.2 or the non-homogeneous measure from Example 2.1 will be given in Sections 8.1 and 8.2, respectively.

6. Undirected graphical models.

6.1. *Fundamentals.* For the fundamental definitions we follow again the axiomatic approach to conditional independence as in Lauritzen (1996), from which we also adopt relevant notions that characterize relations in a graph. Let our index set V coincide with the vertex set (node set) of an undirected graph $\mathcal{G} = (V, E)$ with set of edges E. The subset C is said to *separate* A from B if all paths from any of the nodes of A to any of the nodes of B necessarily intersect with C.

DEFINITION 6.1. The measure Λ satisfies the *global Markov property* with respect to an undirected graph $\mathcal{G}$ if $A \perp B \mid C\ [\Lambda]$ for any disjoint subsets A, B, C where (possibly empty) C separates A from B. In this case we say that Λ is an undirected graphical model on $\mathcal{G}$.

REMARK 6.2. In Definition 6.1, it is sufficient to consider only partitions (A, B, C) of V. If (A, B, C) is instead a triplet of disjoint subsets of V not forming a partition, one can enlarge the subsets A and B so that separation still holds, i.e. there exist subsets A', B', such that (A', B', C) forms a partition of V with $A \subset A'$, $B \subset B'$ and C separates A' and B'; and according to the semi-graphoid property (L2) $A' \perp B' \mid C\ [\Lambda]$ implies $A \perp B \mid C\ [\Lambda]$.

Throughout this section we will further assume that the measure Λ satisfies our fundamental explosiveness assumption (E1). This allows us to apply the independence characterization in Proposition 5.1 to any restricted measure Λ^0_D for $\emptyset \neq D \subset V$ and is important as the condition $A \perp B\,[\Lambda^0_{A\cup B}]$ appears always when checking conditional independence statements; see Section 4. It has critical implications on the support of Λ. The following simple result is a direct consequence of Lemma 3.3 and Proposition 5.1 and helps understanding where Λ is allowed to have mass.

COROLLARY 6.3. *Assume* (E1) *and that* Λ *is globally Markov with respect to an undirected graph* $\mathcal{G}$. *Then* $\Lambda(y_A \neq 0_A,\, y_B \neq 0_B,\, y_C = 0_C) = 0$ *for any disjoint subsets* A, B, C *where (possibly empty)* C *separates* A *from* B.

Assumption (E1) ensures the validity of all four semi-graphoid properties (L1)–(L4); see Theorem 5.3. This implies that various standard results for conditional independence in the probabilistic setting carry over automatically to the Λ-based conditional independence, as they only rely on these properties. For example, the global Markov property implies the *local Markov property*, which in turn implies the *pairwise Markov property* (see Lauritzen, 1996, Prop. 3.4).

The theory of conditional independence and graphical models in the case where Λ has no mass on any of the lower-dimensional sets $\mathcal{H}_D$ defined in (13) is similar to the classical theory of probabilistic graphical models, and issues arising from the special role of the origin can then be avoided. Our main focus in this section is therefore on the general case with mass on several lower-dimensional sub-faces, which is highly desirable in applications for instance in extremes; see Section 8.

6.2. *Decomposable graphs.* Decomposable (or chordal) graphs are an important subclass of general graphical structures. A triple of disjoint subsets (A, B, S) is a *decomposition* of $\mathcal{G}$ if $A \cup B \cup S = V$ and a fully connected S separates A from B. A graph is then called *decomposable* if it can be reduced to a set of cliques $\mathcal{C}$ (i.e., maximal fully connected subgraphs) via decompositions. Importantly, a graph is decomposable if and only if it has the following *running intersection property* (e.g. Lauritzen (1996) Prop. 2.17 or Maathuis et al. (2019) Section 1.6): there exists an ordering of its cliques $C_1, \dots, C_n$ such that for all $j < n$:

- $S_j := C_{j+1} \cap D_j \subset C_i$ for some $i \leq j$, where $D_j = \bigcup_{i \leq j} C_i$;
- the multiset $\mathcal{S} = \{S_1, \dots, S_{n-1}\}$ is independent of the ordering, and it is comprised of all minimal fully connected separators of $\mathcal{G}$.

We note that $\mathcal{S}$ is a multiset since it may contain the same set several times, and it contains the empty set if $\mathcal{G}$ is disconnected. Figures 3 and 4 illustrate two examples of such graphs with their cliques and separators.

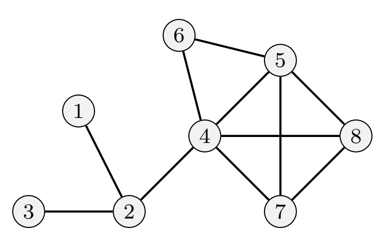

FIG 3. *A decomposable graph with (maximal) cliques* $\mathcal{C} = \{\{1,2\},\{2,3\},\{2,4\},\{4,5,6\},\{4,5,7,8\}\}$ *They are separated by the (minimal) separators* $\mathcal{S} = \{\{2\},\{2\},\{4\},\{4,5\}\}$. *Note that* $\mathcal{S}$ *is a multiset.*

For classical probabilistic graphical models, the Hammersley–Clifford theorem (Lauritzen, 1996, Theorem 3.9) states that the global Markov property on a decomposable graph is equivalent to the factorization of the density of the model into marginal densities on the cliques. Considering instead the measure Λ with density λ, we obtain a similar result. If all mass of Λ is concentrated on $\mathcal{E}_V = \{y : y_i \neq 0 \,\forall i \in V\}$, then the global Markov property in Definition 6.1 is equivalent to

$$\lambda(y) \prod_{S \in \mathcal{S}} \lambda_S(y_S) = \prod_{C \in \mathcal{C}} \lambda_C(y_C), \tag{22}$$

for μ-almost all $y \in \mathcal{E}$ with $y_i \neq 0$ for all $i \in V$. This statement is a special case of the much more general Theorem 6.4 below; see also Engelke and Hitz (2020, Theorem 1) who showed such as factorization for homogeneous measures Λ in the context of extreme value theory.

As we turn to the general case with possible mass on several lower-dimensional subsets $\mathcal{H}_D$, $D \subset V$, an interesting twist of the Λ-conditional independence theory appears compared to the classical probabilistic theory. The factorization (22) for λ is in general no longer equivalent to the global Markov property and the modified densities $\overline{\lambda}$ defined in (16) become instrumental. Our main theorem of this section is a characterization of the global Markov property of Λ in the spirit the classical Hammersley–Clifford theorem. For its proof, we need an auxiliary result, Proposition C.1 in Appendix C.1, which allows us to derive the global Markov property from a decomposition into a globally Markov sub-graph and a clique. Its proof is based on the four semi-graphoid properties (L1)–(L4), and therefore Assumption (E1) is critical again. We define for an undirected graph $\mathcal{G} = (V, E)$ the set

$$\mathcal{Z} = \mathcal{Z}(\mathcal{G}) = \bigcup_{\substack{a,b \in V,\, S \in \mathcal{S}: \\ S \text{ separates } a \text{ from } b}} \{y \in \mathbb{R}^V : y_a \neq 0, y_b \neq 0, y_S = 0_S\},$$

and observe that Λ being globally Markov for $\mathcal{G}$ implies $\Lambda(\mathcal{Z}) = 0$ by Corollary 6.3.

THEOREM 6.4. *Assume that Λ satisfying* (E1) *has a density λ with respect to a product measure μ with $\mu_v(\{0_v\}) > 0$, and the graph $\mathcal{G}$ is decomposable. Then Λ is globally Markov if and only if*

$$\overline{\lambda}(y) \prod_{S \in \mathcal{S}} \overline{\lambda}_S(y_S) = \prod_{C \in \mathcal{C}} \overline{\lambda}_C(y_C) \tag{23}$$

for μ-almost all $y \notin \mathcal{Z}(\mathcal{G})$, and $\overline{\lambda}_\emptyset = 1$ by convention.

The proof of this result is based on an inductive argument using the running intersection property for decomposable graphs and given in Appendix C.2.

REMARK 6.5. In the setting of Theorem 6.4, if Λ is globally Markov for $\mathcal{G}$, we have that $\Lambda(\mathcal{Z}) = 0$ by Corollary 6.3. Hence $\lambda(y) = 0$ for μ-almost all $y \in \mathcal{Z}$ in this situation. However, it is important to note that this is not(!) true for the right hand side of (23). A simple example of this phenomenon becomes apparent from the construction of Section 5.4; see Table 2 (the sixth row) in Appendix B.

REMARK 6.6. Alternatively to the factorization of $\overline{\lambda}$ in Theorem 6.4, one may be tempted to consider instead a characterization of the global Markov property via the factorization of λ as in (22) on the set where all the relevant marginals are defined together with respective support assumptions as in Corollary 6.3. However, first, such a factorization cannot hold for a globally Markov Λ on a disconnected graph as discussed in Section 5.3. Second, although the factorization of λ does hold for a connected graph when Λ is globally Markov, one needs to be cautious with the converse implication; one may not deduce the global Markov property from the factorization of λ alone. In Appendix C.3 we give an example that illustrates this caveat.

Collectively, this shows that factorization of λ is not a suitable characterization of Markov properties of Λ on an undirected graph, and that the modified densities $\overline{\lambda}$ are the more natural object for this purpose.

7. Directed graphical models.

7.1. *Fundamentals.* In this section we consider a *directed acyclic graph (DAG)* $\mathcal{G}$ with vertex set V. The classical definition of the directed Markov properties require the following graph notions; see Lauritzen (1996) for details. Let $\mathrm{pa}(v)$, $\mathrm{an}(v)$ and $\mathrm{de}(v)$ denote *parents*, *ancestors* and *descendants* of $v \in V$, respectively. The *ancestral set* of a subset $A \subset V$ in a DAG $\mathcal{G}$ is defined as A together with all the ancestors of every vertex in A. The *moral graph* of $\mathcal{G}$ is an undirected graph obtained by connecting the parents of every vertex in $\mathcal{G}$ and dropping the directions of the original edges. The following directed Markov properties are again in line with the axiomatic approach to conditional independence.

DEFINITION 7.1 (Directed Markov properties). **(DL)** We say that Λ satisfies the *directed local Markov property* with respect to the DAG $\mathcal{G}$ if for every vertex $v \in V$

$$v \perp V \setminus (\{v\} \cup \mathrm{de}(v) \cup \mathrm{pa}(v)) \mid \mathrm{pa}(v) \; [\Lambda].$$

(DG) We say that Λ satisfies the *directed global Markov property* with respect to $\mathcal{G}$ if for every triplet of disjoint subsets (A, B, S) of V it holds that $A \perp B \mid S \; [\Lambda]$ if A and B are separated by S in the moral graph of the ancestral set of $A \cup B \cup S$.

An alternative equivalent characterization of the global Markov property via the concept of d-*separation* is also possible (e.g., Lauritzen, 1996). Further, it is shown in Lauritzen et al. (1990, Prop. 4) that in the classical setting the two properties (DL) and (DG) are equivalent. The proof only requires properties (L1)–(L4) of the classical conditional independence, and so due to Theorem 5.3 this is also true in our setting as long as Λ satisfies our basic explosiveness assumption (E1).

COROLLARY 7.2. *Assume Λ satisfies (E1) and V is the vertex set of a DAG $\mathcal{G}$. Then Λ satisfies (DG) with respect to $\mathcal{G}$ if and only if Λ satisfies (DL) with respect to $\mathcal{G}$.*

7.2. *Recursive max-linear models.* Recursive max-linear models introduced by Gissibl and Klüppelberg (2018) received a lot of attention in recent years (e.g., Améndola et al., 2022; Gissibl, Klüppelberg and Otto, 2018; Gissibl, Klüppelberg and Lauritzen, 2021). They are defined on a DAG $\mathcal{G}$ by

$$X_i = \bigvee_{j \in \mathrm{pa}(i)} \beta_{ij} \, X_j \vee \beta_{ii} \, \varepsilon_i, \qquad i \in V, \tag{24}$$

where $\beta_{ij} > 0$ and $\varepsilon_i \sim F_i$ are independent non-negative random variables. It is a special case of the classical recursive structural equation model and as such X satisfies the classical probabilistic (DL) property (cf., Pearl, 2009, Thm. 1.4.1). Note that β_{ij} corresponds to the arc $j \to i$ in the graph $\mathcal{G}$, which is a common notation in the literature.

The recursive equation (24) can be rewritten as

$$X_i = \bigvee_{j\in V} \gamma_{ij}\,\varepsilon_j, \quad i\in V, \qquad \text{with} \qquad \gamma_{ij} = \bigvee_{p\in \mathrm{paths}_{ji}} \prod_{(k\to l)\in p} \beta_{lk}, \quad i,j\in V, \tag{25}$$

where paths_{ji} is the set of paths from j to i with the additional convention that each path starts with the $(j\to j)$ transition, and $\gamma_{ij}=0$ for empty paths_{ji}.

Such a model without a restriction on the non-negative γ_{ij} is called *max-linear*. It is immediate that max-linear models are always max-infinitely divisible since every univariate random variable ε_j is max-infinitely divisible. We additionally assume that

$$\inf\{x : F_j(x)>0\}=0, \qquad j\in V, \tag{26}$$

which guarantees that X satisfies (4) and so the exponent measure Λ is finite away from the origin. Let us further recall that the support of this measure consists of $|V|$ rays, each spanned by a column $(\gamma_{\cdot j})$ (e.g., Yuen and Stoev, 2014).

LEMMA 7.3. *Let Λ be the exponent measure of the recursive max-linear model (24) satisfying (26). Then Λ is given by*

$$\Lambda(E) = \sum_{j\in V} \eta_j\big(\{t>0 \,:\, t\gamma_{\cdot j}\in E\}\big)$$

for measurable $E\subset\mathcal{E}_+$, where $\gamma_{\cdot j}$ denotes the j-th column of the $|V|\times|V|$ matrix (γ_{ij}) and η_j denotes the exponent measure of ε_j.

PROOF. From (5), it is immediate that

$$\Lambda(\mathcal{E}_+\setminus[0,x]) = \sum_{j\in V\,:\,\gamma_{\cdot j}\neq 0_V} \eta_j((a_j,\infty)) \quad \text{with} \quad a_j = \bigwedge_{i\in V\,:\,\gamma_{ij}>0} x_i/\gamma_{ij}.$$

It suffices to note that the case $\gamma_{\cdot j}=0_V$ cannot appear due to $\gamma_{jj}>0$ and that for $t>0$ we have that $t>a_j$ if and only if $t\gamma_{\cdot j}\in\mathcal{E}_+\setminus[0,x]$, while sets of the form $\mathcal{E}_+\setminus[0,x]$ form a separating measure-determining class. ∎

The restricted measures Λ_D^0 for $\emptyset\neq D\subset V$ are also readily obtained from Lemma 7.3 and given by

$$\Lambda_D^0(E) = \sum_{\substack{j\in V\,:\,\gamma_{Dj}\neq 0_D\\ \gamma_{(V\setminus D)j}=0_{V\setminus D}}} \eta_j\big(\{t>0 \,:\, t\gamma_{Dj}\in E\}\big)$$

for measurable $E\subset\mathcal{E}_+^D$. The measure Λ_D^0 is supported on the rays spanned by the columns $(\gamma_{\cdot j})$ (restricted to the D-components), whose complementary $V\setminus D$-components are all equal to zero, if such columns exist; otherwise $\Lambda_D^0=0$. In particular, if not null, Λ_D^0 corresponds itself to a max-linear model for a subset of the innovations (ε_j).

LEMMA 7.4. *Let Λ be the exponent measure of the recursive max-linear model (24) satisfying (26). If F_j has no mass at 0 for all $j\in V$, then Λ satisfies (E1).*

PROOF. Since each of the restricted measures Λ_D^0 corresponds itself to another max-linear model for a subset of (ε_j) if not null, it is sufficient to prove (E0). Now for any $v\in V, t>0$

$$\Lambda(y_v>t) = \sum_{j\in V\,:\,\gamma_{vj}>0} \eta_j((t/\gamma_{vj},\infty)) = -\sum_{j\in V\,:\,\gamma_{vj}>0} \log F_j(t/\gamma_{vj}),$$

which either converges to ∞ as $t\downarrow 0$ (because $F_j(0+)=0$ by assumption) or remains 0 when all $\gamma_{vj}=0$. ∎

The main result of this section shows that the exponent measure Λ of a recursive max-linear model (24) satisfies also the properties of a directed graphical model with respect to its underlying graph $\mathcal{G}$ in the sense of Definition 7.1 (in addition to X being a classical probabilistic graphical model with respect to $\mathcal{G}$).

THEOREM 7.5. *Consider the recursive max-linear model in* (24) *defined on a DAG* $\mathcal{G}$ *and satisfying* (26)*. Then the corresponding exponent measure* Λ *satisfies the directed local Markov property (DL) on* $\mathcal{G}$*. It also satisfies the directed global Markov property (DG) on* $\mathcal{G}$ *if* $\varepsilon_j > 0$ *for all* $j \in V$.

PROOF. In view of Lemma 7.4 and Corollary 7.2, it is sufficient to show (DL). In other words, we need to establish $A \perp B \mid S\, [\Lambda]$ for an arbitrary vertex $v \in V$ and $A = \{v\}$, $S = \mathrm{pa}(v)$, $B = V \setminus (\{v\} \cup \mathrm{de}(v) \cup \mathrm{pa}(v))$. This is the same as conditional independence for the marginal measure $\Lambda_{A\cup B\cup S} = \Lambda_{V\setminus \mathrm{de}(v)}$, which is the exponent measure of the original max-linear model (24) with (X_i, ε_i), $i \in \mathrm{de}(v)$ removed. Thus, we may assume that v is a terminal node (it has no children), and hence (A, B, S) is a partition of V.

Let us first consider the measure $\Lambda^0_{A\cup B} = \Lambda^0_{V\setminus\mathrm{pa}(v)}$, which is supported by the rays corresponding to the columns $\gamma_{\cdot j}$ with the index j such that $\gamma_{pj} = 0$ for all parent nodes $p \in \mathrm{pa}(v)$. Now either $\gamma_{vj} = 0$ or $v = j$, and in the latter case $\gamma_{iv} = 0$ for all $i \neq v$ since v is terminal, in particular for $i \in B$. Thus, we obtain $A \perp B\, [\Lambda^0_{A\cup B}]$.

Next, suppose that there is a pair of indices (i, j) such that

$$\gamma_{Si} = c\gamma_{Sj} \neq 0_S$$

for some $c > 0$. Then, since $S = \mathrm{pa}(v)$, both i and j are ancestors of v and it holds that

$$\gamma_{vi} = \bigvee_{p\in pa(v)} \beta_{vp}\gamma_{pi} = c \bigvee_{p\in pa(v)} \beta_{vp}\gamma_{pj} = c\gamma_{vj}.$$

Hence the two rays (corresponding to $\gamma_{\cdot i}$ and $\gamma_{\cdot j}$) projected on $\mathrm{pa}(v) \cup \{v\}$ result in the same ray. Choose any $p \in \mathrm{pa}(v)$ and consider the random vector $Y \sim \mathbb{P}_{R_{p,\varepsilon}}$ with $R_{p,\varepsilon}$ as in (12). It is clear from the above proportionality that conditioning on Y_S renders Y_v a constant. Hence Y_v is independent of Y_B given Y_S and the proof is complete in view of Theorem 4.1 (ii). ■

REMARK 7.6. Beyond establishing the (DL) property for Λ, the proof of Theorem 7.5 also reveals a certain degeneracy of the recursive max-linear model. That is, conditioning on the parents identifies their common child. More precisely, the kernel $\nu_{\mathrm{pa}(v)}(y_{\mathrm{pa}(v)}, \cdot)$ has a deterministic v-th component for $\Lambda_{\mathrm{pa}(v)}$-all values of $y_{\mathrm{pa}(v)} \neq 0_{\mathrm{pa}(v)}$.

The recursive max-linear models are not *faithful* as a classical directed graphical model (see Améndola et al., 2022). That is, X typically will satisfy additional conditional independence properties which can not be inferred from the graph. This continues to be true for the respective Λ-graphical model as can be verified for instance by using the same diamond graph example as in Améndola et al. (2022).

7.3. *Recursive infinitely divisible models.* By replacing the maximum operation with the sum operation we obtain another basic example of a structural equation model:

$$X_i = \sum_{j\in\mathrm{pa}(i)} \beta_{ij} X_j + \beta_{ii}\varepsilon_i = \sum_{j\in V} \gamma_{ij}\varepsilon_j, \qquad i \in V, \tag{27}$$

where $\beta_{ij} \neq 0$ and γ_{ij} is given by (25) but with maximum replaced by the sum over the paths from j to i. This is again a directed graphical model with respect to the underlying graph (Pearl, 2009, Thm. 1.4.1). Misra and Kuruoglu (2016) assume that ε_i are mutually independent α-stable random variables, making X a multivariate α-stable vector. More generally, we may consider ε_j to be independent infinitely divisible random variables, and so X is infinitely divisible.

We remark that, while every univariate distribution is max-infinitely divisible, it is not necessarily infinitely divisible. This explains the need for the extra assumption on the innovation terms ε_j in order to employ the framework of infinitely divisible vectors and respective Lévy measures.

LEMMA 7.7. *If all ε_j are infinitely divisible, then the Lévy measure of the infinitely divisible X defined in (27) is supported by the rays spanned by $\gamma_{\cdot j}$.*

PROOF. Consider the respective Lévy processes $\varepsilon_j(t)$ and $X(t)$, and recall that their independence implies that they do not jump simultaneously a.s. (Sato, 2013). Thus, every jump of $X(t)$ arises from a jump $\Delta\varepsilon_j(t)$ of a single $\varepsilon_j(t)$, and therefore it has a form $\gamma_{\cdot j}\Delta\varepsilon_j(t)$. These jumps belong to the ray spanned by $\gamma_{\cdot j}$ and the proof is complete. ■

Thus, the Lévy measure Λ has the same form as the exponent measure Λ corresponding to the recursive max-linear model, and so we arrive at the analogue of Theorem 7.5.

COROLLARY 7.8. *Consider a recursive linear model in (27) defined on a DAG $\mathcal{G}$ such that all ε_j are infinitely divisible. Then the Lévy measure Λ corresponding to infinitely divisible vector X satisfies the directed local Markov property (DL) with respect to $\mathcal{G}$. It also satisfies the directed global Markov property (DG) with respect to $\mathcal{G}$ when the Lévy measures of ε_j explode at 0 for all $j \in V$.*

PROOF. In light of Lemma 7.7, the proof of Theorem 7.5 can be repeated almost verbatim for the present situation (replace maximum by sum when expressing γ_{vi}) to establish (DL). Note that the arguments do not depend on Λ being supported by the positive orthant. Furthermore, we may establish (E1) similarly to the proof of Lemma 7.4, where we note that

$$\Lambda(|y_v| > t) = \sum_{j \in V} \eta_j(|y\gamma_{vj}| > t),$$

where η_j is the Lévy measure of ε_j. If all η_j are exploding at 0, the limit (as $t \to 0$) is either 0 or ∞. ■

8. Relation to conditional independence in extremes. In this section we discuss the implications of our results on several aspects in the study of sparsity in multivariate extreme value models. When studying limit theorems in this field, the sub-class of exponent measures Λ on the space $\mathcal{E}_+ = [0,\infty)^d \setminus \{0\}$ plays a central role. To any such measure we can associate a max-infinitely divisible random vector $Z = (Z_i : i \in V)$ with distribution function

$$G_\Lambda(x) = \exp\{-\Lambda(\mathcal{E}_+ \setminus [0,x])\}, \quad x \in \mathcal{E}_+. \tag{28}$$

Up to a marginal standardization, these distributions G_Λ are the only possible limits of triangular arrays of maxima (Balkema and Resnick, 1977; de Haan and Resnick, 1977). In the following we assume that Λ has the same marginal measures $\Lambda(u) := \Lambda(y_i > u) > 0$ for all $u \in (0,\infty)$ and $i \in V$, so that each univariate survival function $\mathbb{P}(Z_i > u) = 1 - \exp(-\Lambda(u))$ is asymptotically equivalent to $\Lambda(u)$ in the sense that

$$\lim_{u\to\infty} \frac{\mathbb{P}(Z_i > u)}{\Lambda(u)} = 1. \tag{29}$$

The strength of dependence between the largest observations of the components i-th and j-th component of Z can be summarized by the *extremal correlation coefficient*

$$\chi_{ij} := \lim_{u\to\infty} \mathbb{P}(Z_i > u \mid Z_j > u) = \lim_{u\to\infty} \frac{\Lambda(y_i > u, y_j > u)}{\Lambda(u)} \in [0,1], \tag{30}$$

whenever the limit exists and where the second equation follows from a simple Taylor expansion. Two dependence regimes are usually distinguished: if $\chi_{ij} > 0$ we speak of *asymptotic dependence* between $i, j \in V$, and if $\chi_{ij} = 0$, we say the components are *asymptotically independent*. Accordingly, this section is structured as follows.

In Section 8.1 we focus on the case of asymptotic dependence, where a homogeneous measure Λ can be used to fully describe the extremal dependence properties. Asymptotically independent models are far more complex to describe mathematically. Section 8.2 provides a concrete example of such a max-infinitely divisible distribution that shows how our theory on density factorizations of general Λ measures opens the door to a new theory of asymptotically independent graphical models. Finally, Section 8.3 connects extremal graphical models to other sparsity notions and the field of concomitant extremes.

8.1. *Extremal graphical models.* Let $X \geq 0$ be a random vector and assume for simplicity that it has heavy-tailed marginal distributions with common tail-index $\alpha > 0$. In the case of asymptotic dependence, there are two different, but closely related classical approaches for describing the extremes of the multivariate distribution of X.

The first approach is concerned with scale-normalized componentwise maxima

$$a(n)^{-1} \max_{i=1,2,\dots,n} X^{(i)}$$

of independent copies $X^{(i)}$ of X, where $a(n) > 0$. The only possible limit laws of such maxima as $n \uparrow \infty$ are max-stable with distribution function (28), where the exponent measure Λ satisfies (B) and is $-\alpha$-homogeneous as in (3). In particular, the distribution G_Λ has necessarily α-Fréchet marginals.

The second approach studies the distribution of the scale-normalized exceedances

$$u^{-1} X \,\Big|\, \max_{j=1,\dots,d} X_j > u$$

of the random vector X, conditioning on the event that at least one component X_i exceeds a large threshold u. The only possible limits of these peaks-over-threshold as $u \uparrow \infty$ are multivariate Pareto distributions (Rootzén and Tajvidi, 2006), whose probability laws are induced by a homogeneous measure Λ on the (non-rectangular) set $\mathcal{L} = \mathcal{E}_+ \setminus [0,1]^d$ and take the form $\mathbb{P}_\mathcal{L}(\mathrm{d}y) = \Lambda(\mathrm{d}y)/\Lambda(\mathcal{L})$.

An apparent connection between these two approaches is the exponent measure Λ, which characterizes distribution functions of both, multivariate max-stable distributions and multivariate Pareto distributions. In fact, the connection is due to a fundamental limiting result, which links the two approaches via regular variation. This has been neatly summarized and extended in Dombry and Ribatet (2015, Thm. 1). We only recall here that

$$a(n)^{-1} \max_{i=1,2,\dots,n} X^{(i)} \to G_\Lambda \quad \Longleftrightarrow \quad u^{-1} X \,\Big|\, \max_{j=1,\dots,d} X_j > u \to \mathbb{P}_\mathcal{L}, \tag{31}$$

where $a(n)$ is the $(1-1/n)$-quantile of $\max_{j=1,\dots,d} X_j$. In addition each of these limiting statements is equivalent to the *regular variation* of the random vector X with limiting measure Λ in the sense that the measure $u\mathbb{P}(a(u)^{-1}X \in \cdot)$ converges vaguely to Λ on $\overline{\mathcal{E}}_+ = [0,\infty]^d \setminus \{0\}$, denoted by $X \sim \mathrm{RV}_\alpha^+(\Lambda)$. In other words, for a given $X \sim \mathrm{RV}_\alpha^+(\Lambda)$

the corresponding max-stable distribution G_Λ and multivariate Pareto distribution $\mathbb{P}_{\mathcal{L}}$ are associated via the same exponent measure Λ.

It seems therefore natural to approach conditional independence for multivariate extremes in terms of the exponent measure Λ. We start by linking known versions of extremal (conditional) independence to the Λ-based conditional independence introduced here in Definition 3.1, where Λ assumes now the role of the associated exponent measure. Recall that such exponent measures naturally satisfy our key explosiveness assumption (E1) due to their homogeneity; see Remark 5.4. When writing $(X_A, X_B) \in \mathrm{RV}_\alpha^+(\Lambda)$ below, we tacitly assume that (A, B) constitutes a partition of V, and analogously $(X_A, X_B, X_C) \in \mathrm{RV}_\alpha^+(\Lambda)$ implies (A, B, C) being a partition of V.

DEFINITION 8.1 (Traditional extremal independence). Let $(X_A, X_B) \in \mathrm{RV}_\alpha^+(\Lambda)$, then X_A and X_B are said to exhibit *extremal independence*, denoted by

$$X_A \perp_{\mathrm{tr}} X_B,$$

if we have classical independence $Z_A \perp\!\!\!\perp Z_B$ for $Z \sim G_\Lambda$.

A first observation is that the notion of traditional extremal independence coincides with our new notion of independence with respect to the exponent measure Λ, since both statements are equivalent to $\Lambda(y_A \neq 0_A, y_B \neq 0_B) = 0$; see Proposition 5.1 and, for instance, Strokorb (2020).

COROLLARY 8.2. *Let $(X_A, X_B) \in \mathrm{RV}_\alpha^+(\Lambda)$, then $X_A \perp_{\mathrm{tr}} X_B$ if and only if $A \perp B\ [\Lambda]$.*

REMARK 8.3. If we consider more generally $Z \sim G_\Lambda$, where Λ is not necessarily homogeneous, but still satisfies the basic assumptions (B) and (E0), then also $Z_A \perp\!\!\!\perp Z_B$ if and only if $A \perp B\ [\Lambda]$.

One might be tempted now to extend Definition 8.1 to extremal conditional independence in a similar manner. However, it was shown in Papastathopoulos and Strokorb (2016) that this leads to a flawed approach if one considers models G_Λ with a positive continuous density. Instead two other routes have been successfully pursued. Gissibl and Klüppelberg (2018) study spectrally discrete structural models in the maxima setting where G_Λ does not admit a density; see also Section 7.2. In contrast, Engelke and Hitz (2020) study extremal conditional independence for multivariate Pareto distributions and propose in essence the following definition.

DEFINITION 8.4 (Extremal conditional independence based on Engelke and Hitz (2020)). Let $(X_A, X_B, X_C) \in \mathrm{RV}_\alpha^+(\Lambda)$, then X_A and X_B are said to exhibit *extremal conditional independence* given X_C, denoted by

$$X_A \perp_e X_B \mid X_C,$$

if we have classical conditional independence $Y_A \perp\!\!\!\perp Y_B \mid Y_C$ for $Y \sim \mathbb{P}_{R_{v,1}} = \Lambda(\mathrm{d}y)/\Lambda(R_{v,1})$ for all $v \in V$.

REMARK 8.5. Engelke and Hitz (2020) only consider the case where Λ has a Lebesgue density on $\mathcal{E}_+$ and therefore no mass on any of the sub-faces $\mathcal{E}_D$, $D \subset V$ for $D \neq V$. They show in this case that it suffices to verify classical conditional independence in the above definition for one $v \in C$ in the conditioning set. Our Lemma A.7 in the Appendix shows that indeed any $v \in V$ is sufficient, even if $v \in V \setminus C$.

Since the exponent measure Λ is homogeneous, it is an immediate consequence of Theorem 4.1 (see also Remark 4.2) that the extremal conditional independence from Definition 8.4 and our Λ-based conditional independence coincide. To be precise, we may formulate this finding as follows. In particular, we would like to stress that it is valid for any homogeneous exponent measure Λ without further assumptions; Λ need not have a Lebesgue density and $C = \emptyset$ is allowed.

COROLLARY 8.6. *Let* $(X_A, X_B, X_C) \in \mathrm{RV}_\alpha^+(\Lambda)$, *then*

$$X_A \perp_e X_B \mid X_C \quad \iff \quad A \perp B \mid C \, [\Lambda].$$

In particular, this includes the case $C = \emptyset$.

Originally, Engelke and Hitz (2020) excluded the independence case $C = \emptyset$, as their approach is based on working with a Lebesgue density of Λ and does not allow for positive mass on the lower-dimensional subsets $\mathcal{H}_A \cup \mathcal{H}_B$; in view of Proposition 5.1 we know that mass on such subsets is crucial for the independence case. In our new setup, Definition 8.4 and Corollary 8.6 naturally encapsulate the case $C = \emptyset$ and it is furthermore in line with the traditional notion of extremal independence. We have seen now that for $(X_A, X_B) \in \mathrm{RV}_\alpha^+(\Lambda)$

$$X_A \perp_{\mathrm{tr}} X_B \quad \iff \quad X_A \perp_e X_B \quad \iff \quad A \perp B \, [\Lambda].$$

The first equivalence has already been shown in Strokorb (2020).

In Corollary 8.6 it is also possible to consider discrete spectral measures, which builds a bridge to the structural max-linear models (24) from Gissibl and Klüppelberg (2018); see Section 7.2. Theorem 7.5 shows that we recover all the conditional independencies that are encoded in the structural model equations also within our Λ-based notion of conditional independence.

REMARK 8.7. Instead of restricting oneself to the cone $\mathcal{E}_+ = [0, \infty)^d \setminus \{0\}$, we can also consider a random vector X on $\mathcal{E}$ that is regularly varying in the sense that there exists a scaling function $a(u)$, such that $u\mathbb{P}(a(u)^{-1}X \in \cdot)$ converges vaguely to a measure Λ. Definition 8.4 and Corollary 8.6 also carry over naturally to the situation when $(X_A, X_B, X_C) \in \mathrm{RV}_\alpha(\Lambda)$.

Extremal graphical models have been applied to assess flood risk (Asenova, Mazo and Segers, 2021; Röttger, Engelke and Zwiernik, 2023), financial risks (Engelke, Lalancette and Volgushev, 2021) and large delays in flight networks (Hentschel, Engelke and Segers, 2022). The parametric family of Hüsler–Reiss distributions (Hüsler and Reiss, 1989) can be seen as the analogue of the multivariate Gaussian distribution in extremes, and it is so far the only one used in these kind of statistical applications. One advantage of our new theory here in the context of extremal graphical models is that we lay the foundation to construct much more general models than in Engelke and Hitz (2020), overcoming key limitations that were pointed out in its discussion part. In particular, we do not require Lebesgue densities and the exponent measure Λ can have mass on lower-dimensional sub-faces of $\mathcal{E}_+$. Importantly, the graph can therefore have unconnected components, which corresponds to asymptotically independent groups of variables. The following example provides a generalization of the widely used Hüsler–Reiss distributions that makes use of all of the above features.

EXAMPLE 8.8. Let $V = \{1, \ldots, d\}$ and suppose that $\mathcal{G} = (V, E)$ is a forest, that is, a graph where each connected component is a tree. The (maximal) cliques of such a graph are its edges together with its isolated points, cf. also Figure 4, i.e. $\mathcal{C} = E \cup \mathcal{I}$, where $\mathcal{I}$ consists

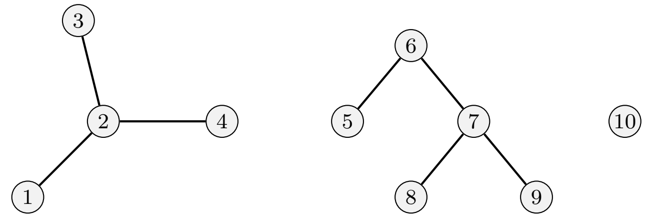

FIG 4. *A forest with cliques* $\mathcal{C} = \{\{1,2\},\{2,3\},\{2,4\},\{5,6\},\{6,7\},\{7,8\},\{7,9\},\{10\}\}$ *(edges and isolated singleton {10}) separated by separators* $\mathcal{S} = \{\{2\},\{2\},\emptyset,\{6\},\{7\},\{7\},\emptyset\}$*. The empty set appears twice in* $\mathcal{S}$*, as the forest consists of three connected components. For each node* $i \in V = \{1,\dots,10\}$*, the singleton* $\{i\}$ *appears* $d_i - 1$ *times in* $\mathcal{S}$*, where* d_i *is the degree of* i *in the forest.*

of isolated singletons $\{i\}$. The separator multiset $\mathcal{S}$ is also comprised of two kinds of sets; on the one hand empty sets separating connected components and on the other hand singletons $\{i\}$ that are separating edges. The multiplicity of each singleton $\{i\}$ in the multiset $\mathcal{S}$ is $d_i - 1$ if d_i denotes the degree of i in the forest. Similarly, the empty set appears once less than the number of trees.

We work with the same setup as in the example in Section 5.4 and assume a dominating product measure of the form (19). As building blocks for any edge $\{i,j\} \in E$ we use the bivariate Hüsler–Reiss densities (e.g., Engelke et al., 2015) given by

$$\kappa_{ij}(y_{ij}) = \frac{1}{\sqrt{2\pi\Gamma_{ij}}} y_i^{-2} y_j^{-1} \exp\{-[\log(y_j/y_i) + \Gamma_{ij}/2]^2/(2\Gamma_{ij})\}, \;\; y_{ij} \in (0,\infty)^2, \tag{32}$$

where $\Gamma_{ij} \in (0,\infty)$ are the corresponding dependence parameters on the edges as in Example 2.2; note that these densities are -3-homogeneous and have univariate marginal densities $m(y) = y^{-2}$. As in Section 5.4, we let λ_{ij} be the corresponding densities on $\mathcal{E}_+^{ij}$ with masses on the axes as in (20) with mixture parameters $p_{ij} \in (0,1]$, $\{i,j\} \in E$. Recall that the marginal densities of λ_{ij} are by construction $m(y_i)$ and $m(y_j)$, respectively.

The exponent measure density of a Hüsler–Reiss forest is defined as

$$\lambda(y) = \frac{\prod_{\{i,j\}\in E} \overline{\lambda}_{ij}(y_i,y_j)}{\prod_{\{i\}\in \mathcal{S}} \overline{\lambda}_i(y_i)} \prod_{\{i\}\in\mathcal{I}} \overline{\lambda}_i(y_i), \tag{33}$$

for μ-almost all $y \in \mathcal{E}_+ \setminus \mathcal{Z}(\mathcal{G})$, and it is implied that $\lambda(y) = 0$ otherwise. By construction, the corresponding measure Λ on $\mathcal{E}_+ = [0,\infty)^V \setminus \{0_V\}$ is -1-homogeneous with marginal densities $\lambda_i(y_i) = m(y_i) = y_i^{-2}$. By Theorem 6.4 it is an undirected graphical model on the forest $\mathcal{G}$.

We write $i \leftrightarrow j$ if i and j are connected, and $i \not\leftrightarrow j$ otherwise. In the former case, we denote by path_{ij} the set of all edges on the unique shortest path between i and j in $\mathcal{G}$. We can explicitly express the extremal correlation between arbitrary nodes $i,j \in V$ as

$$\chi_{ij} = \begin{cases} \{2 - 2\Phi(\sqrt{\Gamma_{ij}}/2)\} \prod_{(s,t)\in\mathrm{path}_{ij}} p_{st} & \text{for } i \leftrightarrow j, \\ 0 & \text{for } i \not\leftrightarrow j, \end{cases}$$

where Φ is the standard normal distribution function, and the $\Gamma_{ij} = \sum_{(s,t)\in\mathrm{path}_{ij}} \Gamma_{st}$ are the tree-completed Hüsler–Reiss coefficients for all $i \leftrightarrow j$ (e.g., Asenova and Segers, 2023; Engelke and Volgushev, 2022); see Appendix D.1 for the proof. This means that nodes in different connected components are asymptotically independent.

8.2. *Asymptotic independence.* In the previous section we have discussed the implications of our results for asymptotically dependent models, where a homogeneous measure Λ characterizes the extremal limits. The theory of this paper is much more general and allows, for the first time, to define non-trivial Λ-based graphical models in the regime of asymptotic independence. In general, there is no unified way of describing the dependence structures of all asymptotically independent distributions, and different approaches exist, including hidden regular variation (Resnick, 2002), conditional extreme value models (Heffernan and Tawn, 2004) or scale mixtures (Engelke, Opitz and Wadsworth, 2019; Wadsworth et al., 2017).

Another possibility to construct distributions that exhibit asymptotic independence is given by a max-infinitely divisible distribution Z with distribution function (28) with suitable non-homogeneous exponent measure Λ; here we also assume equal marginal measures with $\Lambda(u) > 0$ for all $u \in (0, \infty)$. In this case, the extremal correlations in (30) are not informative since they satisfy $\chi_{ij} = 0$ for all $i, j \in V$. Following the approach of Ledford and Tawn (1996, 1997), a refined *residual tail dependence coefficient* $\eta_{ij} \in (0, 1]$ can be defined through

$$\mathbb{P}(Z_i > u, Z_j > u) = \ell_{ij}(\mathbb{P}(Z_i > u)^{-1})\mathbb{P}(Z_i > u)^{1/\eta_{ij}}, \tag{34}$$

where ℓ_{ij} is a slowly varying function; see Appendix D.3 for a review of some fundamental properties of slowly varying functions. Due to (29), the relation (34) is equivalent to

$$\mathbb{P}(Z_i > u, Z_j > u) = \widehat{\ell}_{ij}(\Lambda(u)^{-1})\Lambda(u)^{1/\eta_{ij}} \tag{35}$$

for a slowly varying function $\widehat{\ell}_{ij}$, cf. also Lemma D.6. The coefficient η_{ij} characterizes the decay rate of the joint exceedance probability relative to the univariate survival function. We speak about *positive* and *negative extremal association* between Z_i and Z_j if $\eta_{ij} \in (1/2, 1]$ and $\eta_{ij} \in (0, 1/2)$, respectively, and about *near independence* if $\eta_{ij} = 1/2$. For max-infinitely divisible distributions we cannot obtain negative extremal association and the joint survival function $\mathbb{P}(Z_i > u, Z_j > u)$ is asymptotically equivalent to the joint survival function of the exponent measure $\Lambda(y_i > u, y_j > u)$ for positive extremal association. A proof for this result is given in Appendix D.2.

LEMMA 8.9. *Let Z be a max-infinitely divisible distribution with residual tail dependence coefficient η_{ij} as in* (34) *or* (35)*. Then $\eta_{ij} \geq 1/2$ for all $i, j \in V$. If $\eta_{ij} > 1/2$ we have that*

$$\lim_{u\to\infty} \frac{\Lambda(y_i > u, y_j > u)}{\mathbb{P}(Z_i > u, Z_j > u)} = 1 \tag{36}$$

and

$$\Lambda(y_i > u, y_j > u) = \widetilde{\ell}_{ij}(\Lambda(u)^{-1})\Lambda(u)^{1/\widetilde{\eta}_{ij}}, \tag{37}$$

for a slowly varying function $\widetilde{\ell}_{ij}$ and $\widetilde{\eta}_{ij} = \eta_{ij}$.

Conversely, if (37) *holds for some $\widetilde{\eta}_{ij} > 1/2$, this entails* (36) *as well as* (34) *and* (35) *with $\eta_{ij} = \widetilde{\eta}_{ij}$.*

REMARK 8.10. It is known that max-infinitely divisible distributions are positively associated in the usual sense (Resnick, 2008, Prop. 5.29). The above result confirms that this translates into extremal positive association. In many practical examples we have indeed $\eta_{ij} > 1/2$, however for $\eta_{ij} = 1/2$ (the near independence case) one has to be careful. Consider for instance the situation where Z_i and Z_j are independent. Then the limit in (36) will be zero (instead of 1) and the joint survival functions are no longer tail-equivalent.

An example for an exponent measure that induces an asymptotically independent max-infinitely divisible distribution Z exhibiting positive association is the bivariate Gaussian type measure Λ_ρ in Example 2.1 with correlation $\rho > 0$. By construction, in this case (Z_1, Z_2) is a bivariate log-Gaussian distribution, which is asymptotically independent with residual tail dependence coefficient

$$\eta = \frac{1+\rho}{2}.$$

In order to construct examples in higher dimensions $d > 2$ with graphical structure, we can draw on the density factorization in Theorem 6.4.

Here we consider the most basic construction on the chain graph $\mathcal{G} = (V, E)$ with nodes $V = \{1, 2, 3\}$ and edges $E = \{(1, 2), (2, 3)\}$. Following the example in Section 5.4 with no mass on the axes, we consider bivariate measures λ_{12} and λ_{23} that are equal to the densities of the log-Gaussian exponent measures Λ_ρ with correlations $\rho_{12}, \rho_{23} > 0$, respectively. We note that they share the same univariate marginal density $\lambda_1 = \lambda_2 = \lambda_3$. Setting

$$\lambda(y) = \lambda_{12}(y_{12})\lambda_{23}(y_{23})/\lambda_2(y_2), \quad y \in \mathcal{E}_+, \tag{38}$$

we obtain a trivariate density that induces a valid exponent measure Λ and a corresponding max-infinitely divisible distribution $Z = (Z_1, Z_2, Z_3)$. By construction, the exponent measure Λ with density (38) is an undirected graphical model on the chain graph $\mathcal{G}$.

PROPOSITION 8.11. *The max-infinitely divisible distribution Z with exponent measure Λ with density* (38) *is asymptotically independent with residual tail dependence coefficients*

$$\eta_{12} = \frac{1+\rho_{12}}{2}, \qquad \eta_{23} = \frac{1+\rho_{23}}{2}, \qquad \eta_{13} = \frac{1+\rho_{12}\rho_{23}}{2}.$$

We provide a proof of Proposition 8.11 in Appendix D.3. This result resembles the corresponding residual joint asymptotic tail behavior that we would also have detected in a trivariate (log-)Gaussian distribution with the same graphical structure. However, what is quite subtle in this context, is that the distribution of Z is in fact not(!) log-Gaussian, but only arises by combining two bivariate log-Gaussian exponent measure densities.

This construction principle can be extended to more general graphs such as trees. Moreover, any valid exponent measure density of a bivariate max-infinitely divisible distribution can be used; for more examples we refer to Huser, Opitz and Thibaud (2021). This provides a powerful route for statistical modeling of high-dimensional data which exhibit asymptotic independence to be addressed in future research.

8.3. *Sparsity.* Detecting meaningful and interpretable sparse structures in multivariate extremal dependence is an active field of research. Most approaches are formulated in terms of the exponent measure Λ of the respective limit models in (31). The review article Engelke and Ivanovs (2021) distinguishes three groups of methods and formalizes the underlying notions of sparsity. In this section we show that our theory reveals an intimate connection between two of them. First, let us recall the relevant sparsity notions.

The research on *concomitant extremes* is concerned with the identification of elementary sub-faces $\mathcal{E}_D$, $\emptyset \neq D \subset V$, that the exponent measure charges with mass, that is, $\Lambda(\mathcal{E}_D) > 0$. In this context, a sub-face is defined as

$$\mathcal{E}_D = \{y \in \mathcal{E}_+ : y_D > 0, y_{V\setminus D} = 0_{V\setminus D}\}. \tag{39}$$

This is of interest since in the case $\Lambda(\mathcal{E}_D) > 0$, underlying variables indexed by the set D can be jointly extreme. The set $\mathcal{E}_+$ can be partitioned in the $2^d - 1$ disjoint sub-faces $\mathcal{E}_D$ for all non-empty subsets $D \subset V$. The sparsity notion 2(a) in Engelke and Ivanovs (2021) requires that there is only a small number of groups of variables that can be concomitantly extreme, that is,

$$|\{D \subset V : \Lambda(\mathcal{E}_D) > 0\}| \ll 2^d - 1. \tag{40}$$

The sparsity notion 2(b) then states that the maximal cardinality of such groups is much smaller than the dimension $d = |V|$. Many statistical approaches have been proposed to tackle the difficult problem of determining those sets (e.g., Chiapino and Sabourin, 2017; Chiapino, Sabourin and Segers, 2019; Meyer and Wintenberger, 2021; Simpson, Wadsworth and Tawn, 2020).

A more classical notion of sparsity is in terms of conditional independence relations. A sparse graph $\mathcal{G} = (V, E)$, directed or undirected, will induce a distribution that can be explained by lower-dimensional objects; see Sections 6 and 7 here. Accordingly, sparsity notion 3 in Engelke and Ivanovs (2021) is the requirement that a graphical model on $\mathcal{G} = (V, E)$ is sparse if the number of edges $|E|$ is much smaller than the number of all possible edges $d(d-1)/2$, that is,

$$|E| \ll d^2. \tag{41}$$

Methods for sparsity in terms of concomitance as in (40) on the one hand, and in terms of sparse graphs as in (41) on the other hand, have so far not been considered together. In the present conditional independence framework for an exponent measure Λ, we can now establish a link between them. More precisely, we draw on Corollary 6.3, which gives insight how the topology of the graph $\mathcal{G}$ underlying an undirected extremal graphical model limits the number of potentially charged sub-faces $\mathcal{E}_D$. It implies that $\Lambda(\mathcal{E}_D) = 0$ for any index set D such that $\mathcal{G}$ restricted to D is disconnected. This gives an *a priori* upper bound on the number of charged sub-faces.

COROLLARY 8.12. *For an exponent measure Λ corresponding to an extremal graphical model on the undirected graph $\mathcal{G} = (V, E)$, an upper bound on the number of sub-faces charged with mass by Λ is given by*

$$|\{D \subset V : \Lambda(\mathcal{E}_D) > 0\}| \le |\{D \subset V : \mathcal{G}_D \text{ is connected}\}|, \tag{42}$$

where the right-hand side is the cardinality of all connected sub-graphs of $\mathcal{G}$. Moreover, if $\Lambda(\mathcal{E}_D) > 0$, then the cardinality $|D|$ is bounded from above by the number of nodes in the largest connected component of $\mathcal{G}$.

Hence, sparsity in the sense of few edges in the graph (requirement (41)) may also enforce sparsity in terms of fewer potential concomitant extremes (requirement (40)). However, the precise topology matters as well, as the following example illustrates.

EXAMPLE 8.13. Consider the three graphs with $d = |V| = 10$ nodes in Figure 5. Table 1 displays the counts of their edges and connected sub-graphs. The graph on the left corresponds to asymptotic independence between all d nodes. There are no edges, i.e. the graph is sparse in the sense of (41). Trivially, we can only have d sub-faces with mass in this example, which is certainly also a sparse model in the sense of concomitant extremes, i.e., (40), since $d \ll 2^d - 1$. The ring graph in the center is a connected, non-decomposable graph with d nodes and d edges, hence also (41). Using simple combinatorics, one can see that the number of connected sub-graphs of a ring is $(d-1)d + 1 \ll 2^d - 1$ and therefore by Corollary 8.12 it also induces a sparse model for concomitance, cf. (40).

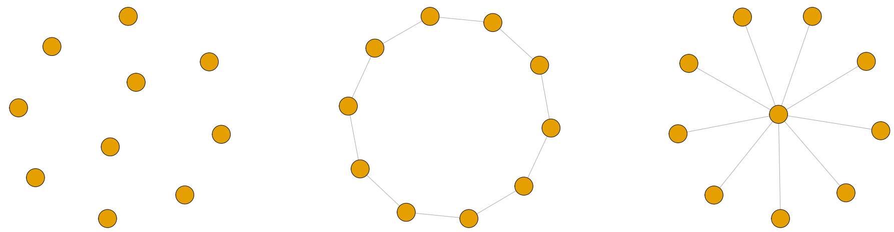

FIG 5. *Three graphs with $d = 10$ nodes: empty (left), ring (center) and star (right) graph.*

However, the star graph on the right-hand side of Figure 5 shows that the number of edges, here $d-1$, is not a sufficient indicator of whether only few sub-faces are charged with mass. Indeed, the star graph has $2^{d-1}+d-1$ connected sub-graphs, which is the largest number among any tree structure (Székely and Wang, 2005, Theorem 3.1). While this tree model is simple in terms of the graph sparsity notion (41), the corresponding exponent measure Λ may have mass on more than half of its sub-faces.

TABLE 1
Counts of edges and connected sub-graphs for graphs with $d = 10$ nodes displayed in Figure 5. The fully connected graph (not displayed in Figure 5) corresponds to the maximal possible counts.

type of graph	empty	ring	star	fully connected
# edges	0	$d = 10$	$d-1=9$	$d(d-1)/2 = 45$
# connected sub-graphs	$d = 10$	$(d-1)d+1 = 91$	$2^{d-1}+d-1 = 521$	$2^d - 1 = 1023$

From these examples it can be seen that connectivity properties of the graph $\mathcal{G} = (V, E)$ are more important for the number of sub-faces with Λ-mass than the cardinality of the edge set $|E|$. Since our extended definition of extremal graphical models for Λ allows for unconnected components in a graph, this enables particularly sparse models in the sense of concomitant extremes.

9. Outlook on applications to Lévy processes. The theory of conditional independence and graphical models developed in this paper applies to general infinite measures Λ. We concentrated on discussing the implications of our results in the field of extreme value theory, since this field has been very active recently; see Sections 7.2 and 8.

Another important instance of a measure Λ studied in this paper is the Lévy measure of an infinitely divisible random vector X as introduced in Section 2.2. In this case, Λ appears in the characteristic function of X and it is not obvious which probabilistic interpretation a conditional independence statement $A \perp B \mid C\ [\Lambda]$ has. In the particular case where the Lévy measure Λ concentrates on a finite number of rays this statement implies the classical conditional independence $X_A \perp\!\!\!\perp X_B \mid X_C$; see Section 7.3. For general Lévy measures this turns out to not be true.

Ongoing research shows that in this case a different and surprisingly intuitive probabilistic interpretation is available in terms of Lévy processes. A Lévy process $\{X(t) : t \geq 0\}$ with values in $\mathbb{R}^d$ is a stochastic process with independent and stationary increments (e.g., Bertoin, 1996; Sato, 2013). Moreover, there is a one-to-one mapping that takes an infinitely divisible random vector X with characteristic triplet $(0,0,\Lambda)$ to the associated Lévy process satisfying $X(1) \stackrel{d}{=} X$; here we assume the drift and Gaussian component in Section 2.2 to be zero since they are well-understood. It turns out that our Λ conditional independence characterizes conditional independence on the level of the sample paths of the Lévy process. Indeed, under the explosiveness assumption (E1) it can be shown that

$$\{X_A(t): t \geq 0\} \perp\!\!\!\perp \{X_B(t): t \geq 0\} \mid \{X_C(t): t \geq 0\} \quad \Longleftrightarrow \quad A \perp B \mid C \, [\Lambda]$$

This result enables the definition of graphical models for Lévy processes with large potential for new theory, statistical methods and applications.

REFERENCES

AMÉNDOLA, C., KLÜPPELBERG, C., LAURITZEN, S. and TRAN, N. M. (2022). Conditional independence in max-linear Bayesian networks. *Ann. Appl. Probab.* **32** 1–45. MR4386833

ASENOVA, S., MAZO, G. and SEGERS, J. (2021). Inference on extremal dependence in the domain of attraction of a structured Hüsler-Reiss distribution motivated by a Markov tree with latent variables. *Extremes* **24** 461–500. MR4301085

ASENOVA, S. and SEGERS, J. (2023). Extremes of Markov random fields on block graphs: max-stable limits and structured Hüsler-Reiss distributions. *Extremes* **26** 433–468. MR4627351

BALKEMA, A. A. and RESNICK, S. I. (1977). Max-infinite divisibility. *J. Appl. Probability* **14** 309–319. MR438425

BERG, C., CHRISTENSEN, J. P. R. and RESSEL, P. (1984). *Harmonic analysis on semigroups*. *Graduate Texts in Mathematics* **100**. Springer-Verlag, New York Theory of positive definite and related functions. MR747302

BERTOIN, J. (1996). *Lévy processes*. Cambridge University Press: Cambridge, UK.

CHIAPINO, M. and SABOURIN, A. (2017). Feature Clustering for Extreme Events Analysis, with Application to Extreme Stream-Flow Data. In *New Frontiers in Mining Complex Patterns* (A. APPICE, M. CECI, C. LOGLISCI, E. MASCIARI and Z. W. RAŚ, eds.) 132–147. Springer International Publishing, Cham.

CHIAPINO, M., SABOURIN, A. and SEGERS, J. (2019). Identifying groups of variables with the potential of being large simultaneously. *Extremes* **22** 193–222. MR3949044

DAVYDOV, Y., MOLCHANOV, I. and ZUYEV, S. (2008). Strictly stable distributions on convex cones. *Electron. J. Probab.* **13** no. 11, 259–321. MR2386734

DAWID, A. P. (1979). Conditional independence in statistical theory. *J. Roy. Statist. Soc. Ser. B* **41** 1–31. MR535541

DE HAAN, L. (1984). A spectral representation for max-stable processes. *Ann. Probab.* **12** 1194–1204. MR757776

DE HAAN, L. and FERREIRA, A. (2006). *Extreme value theory*. *Springer Series in Operations Research and Financial Engineering*. Springer, New York An introduction. MR2234156

DE HAAN, L. and RESNICK, S. I. (1977). Limit theory for multivariate sample extremes. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **40** 317–337. MR478290

DOMBRY, C. and RIBATET, M. (2015). Functional regular variations, Pareto processes and peaks over threshold. *Stat. Interface* **8** 9–17. MR3320385

ENGELKE, S. and HITZ, A. S. (2020). Graphical models for extremes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **82** 871–932. With discussions. MR4136498

ENGELKE, S. and IVANOVS, J. (2021). Sparse structures for multivariate extremes. *Annu. Rev. Stat. Appl.* **8** 241–270. MR4243547

ENGELKE, S., LALANCETTE, M. and VOLGUSHEV, S. (2021). Learning extremal graphical structures in high dimensions. https://doi.org/10.48550/ARXIV.2111.00840

ENGELKE, S., OPITZ, T. and WADSWORTH, J. (2019). Extremal dependence of random scale constructions. *Extremes* **22** 623–666. MR4031852

ENGELKE, S. and VOLGUSHEV, S. (2022). Structure learning for extremal tree models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **84** 2055–2087. MR4515566

ENGELKE, S., MALINOWSKI, A., KABLUCHKO, Z. and SCHLATHER, M. (2015). Estimation of Hüsler-Reiss distributions and Brown-Resnick processes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **77** 239–265. MR3299407

FELLER, W. (1971). *An introduction to probability theory and its applications. Vol. II*, Second ed. John Wiley & Sons, Inc., New York-London-Sydney. MR0270403

GISSIBL, N. and KLÜPPELBERG, C. (2018). Max-linear models on directed acyclic graphs. *Bernoulli* **24** 2693–2720. MR3779699

GISSIBL, N., KLÜPPELBERG, C. and OTTO, M. (2018). Tail dependence of recursive max-linear models with regularly varying noise variables. *Econom. Stat.* **6** 149–167. MR3797980

GISSIBL, N., KLÜPPELBERG, C. and LAURITZEN, S. (2021). Identifiability and estimation of recursive max-linear models. *Scand. J. Stat.* **48** 188–211. MR4233170

HEFFERNAN, J. E. and TAWN, J. A. (2004). A conditional approach for multivariate extreme values. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **66** 497–546. With discussions and reply by the authors. MR2088289

HENTSCHEL, M., ENGELKE, S. and SEGERS, J. (2022). Statistical inference for Hüsler-Reiss graphical models through matrix completions. https://doi.org/10.48550/ARXIV.2210.14292

HULT, H. and LINDSKOG, F. (2007). Extremal behavior of stochastic integrals driven by regularly varying Lévy processes. *Ann. Probab.* **35** 309–339. MR2303951

HUSER, R., OPITZ, T. and THIBAUD, E. (2021). Max-infinitely divisible models and inference for spatial extremes. *Scand. J. Stat.* **48** 321–348. MR4233175

HÜSLER, J. and REISS, R.-D. (1989). Maxima of normal random vectors: between independence and complete dependence. *Statist. Probab. Lett.* **7** 283–286. MR980699

IVANOVS, J. (2020). Discussion of 'Graphical models for extremes' by Engelke and Hitz.

KABLUCHKO, Z. (2009). Spectral representations of sum- and max-stable processes. *Extremes* **12** 401–424. MR2562988

KABLUCHKO, Z., SCHLATHER, M. and DE HAAN, L. (2009). Stationary max-stable fields associated to negative definite functions. *Ann. Probab.* **37** 2042–2065. MR2561440

KALLENBERG, O. (2002). *Foundations of modern probability*, second ed. *Probability and its Applications (New York)*. Springer-Verlag, New York. MR1876169

KLÜPPELBERG, C. and LAURITZEN, S. (2019). Bayesian networks for max-linear models. In *Network science* 79–97. Springer, Cham. MR3971300

LAURITZEN, S. L. (1996). *Graphical models*. *Oxford Statistical Science Series* **17**. The Clarendon Press, Oxford University Press, New York Oxford Science Publications. MR1419991

LAURITZEN, S. L., DAWID, A. P., LARSEN, B. N. and LEIMER, H. G. (1990). Independence properties of directed Markov fields. In *Networks*, **20** 491–505. Special issue on influence diagrams. MR1064735

LEDFORD, A. W. and TAWN, J. A. (1996). Statistics for near independence in multivariate extreme values. *Biometrika* **83** 169–187. MR1399163

LEDFORD, A. W. and TAWN, J. A. (1997). Modelling dependence within joint tail regions. *J. Roy. Statist. Soc. Ser. B* **59** 475–499. MR1440592

MAATHUIS, M., DRTON, M., LAURITZEN, S. and WAINWRIGHT, M., eds. (2019). *Handbook of graphical models*. *Chapman & Hall/CRC Handbooks of Modern Statistical Methods*. CRC Press, Boca Raton, FL. MR3889064

MEYER, N. and WINTENBERGER, O. (2021). Sparse regular variation. *Adv. in Appl. Probab.* **53** 1115–1148. MR4342579

MISRA, N. and KURUOGLU, E. E. (2016). Stable graphical models. *J. Mach. Learn. Res.* **17** Paper No. 168, 36. MR3567436

MOLCHANOV, I. (2009). Convex and star-shaped sets associated with multivariate stable distributions. I. Moments and densities. *J. Multivariate Anal.* **100** 2195–2213. MR2560363

PAPASTATHOPOULOS, I. and STROKORB, K. (2016). Conditional independence among max-stable laws. *Statist. Probab. Lett.* **108** 9–15. MR3420420

PEARL, J. (1988). *Probabilistic reasoning in intelligent systems: networks of plausible inference*. *The Morgan Kaufmann Series in Representation and Reasoning*. Morgan Kaufmann, San Mateo, CA. MR965765

PEARL, J. (2009). *Causality*, Second ed. Cambridge University Press, Cambridge Models, reasoning, and inference. MR2548166

PEARL, J. and PAZ, A. (2022). GRAPHOIDS: graph-based logic for reasoning about relevance relations – or – when would x tell you more about y if you already know z? In *Probabilistic and causal inference. The works of Judea Pearl* 189–200. New York, NY: Association for Computing Machinery (ACM). https://doi.org/10.1145/3501714.3501729

DLMF *NIST Digital Library of Mathematical Functions*. `http://dlmf.nist.gov/`, Release 1.1.7 of 2022-10-15. F. W. J. Olver, A. B. Olde Daalhuis, D. W. Lozier, B. I. Schneider, R. F. Boisvert, C. W. Clark, B. R. Miller, B. V. Saunders, H. S. Cohl, and M. A. McClain, eds.

RESNICK, S. (2002). Hidden regular variation, second order regular variation and asymptotic independence. *Extremes* **5** 303–336 (2003). MR2002121

RESNICK, S. I. (2008). *Extreme values, regular variation and point processes*. *Springer Series in Operations Research and Financial Engineering*. Springer, New York Reprint of the 1987 original. MR2364939

ROOTZÉN, H. and TAJVIDI, N. (2006). Multivariate generalized Pareto distributions. *Bernoulli* **12** 917–930. MR2265668

RÖTTGER, F., ENGELKE, S. and ZWIERNIK, P. (2023). Total positivity in multivariate extremes. *Ann. Statist.* **51** 962–1004. MR4630937

SATO, K.-I. (2013). *Lévy processes and infinitely divisible distributions*. *Cambridge Studies in Advanced Mathematics* **68**. Cambridge University Press, Cambridge Translated from the 1990 Japanese original, Revised edition of the 1999 English translation. MR3185174

SEGERS, J. (2020). One- versus multi-component regular variation and extremes of Markov trees. *Adv. in Appl. Probab.* **52** 855–878. MR4153591

SIMPSON, E. S., WADSWORTH, J. L. and TAWN, J. A. (2020). Determining the dependence structure of multivariate extremes. *Biometrika* **107** 513–532. MR4138974

SKOROHOD, A. V. (1957). Limit theorems for stochastic processes with independent increments. *Teor. Veroyatnost. i Primenen.* **2** 145–177. MR0094842

STROKORB, K. (2020). Extremal independence old and new. Discussion of 'Graphical models for extremes' by Engelke and Hitz. https://doi.org/10.48550/ARXIV.2002.07808

SZÉKELY, L. A. and WANG, H. (2005). On subtrees of trees. *Adv. in Appl. Math.* **34** 138–155. MR2102279

WADSWORTH, J. L., TAWN, J. A., DAVISON, A. C. and ELTON, D. M. (2017). Modelling across extremal dependence classes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 149–175. MR3597968

WANG, Y. and STOEV, S. A. (2010). On the association of sum- and max-stable processes. *Statist. Probab. Lett.* **80** 480–488. MR2593589

YUEN, R. and STOEV, S. (2014). CRPS M-estimation for max-stable models. *Extremes* **17** 387–410. MR3252818

ZHANG, Y. and WANG, L. (2020). Conditional Independence Beyond Domain Separability: Discussion of Engelke and Hitz (2020). https://doi.org/10.48550/ARXIV.2006.01036

APPENDIX A: ADDITIONAL PROOFS FOR SECTIONS 3 AND 4: CONDITIONAL INDEPENDENCE AND ADDITIONAL CHARACTERIZATION RESULTS

A.1. Consistency properties of the basic operations with measures.

LEMMA A.1. *Let (A, B, C) be a partition of V with $C \neq \emptyset$. Then*

$$(\Lambda_{A\cup C})_C = \Lambda_C, \qquad (\Lambda^0_{A\cup C})^0_C = \Lambda^0_C, \qquad (\Lambda^0_{A\cup C})_C = (\Lambda_{B\cup C})^0_C.$$

PROOF. The domain of these measures is $\mathcal{E}^C = \mathbb{R}^C \setminus \{0_C\}$, and expressed in terms of Λ they coincide with $\Lambda(y_C \in \cdot)$, $\Lambda(y_C \in \cdot, y_B = 0_B, y_A = 0_A)$ and $\Lambda(y_C \in \cdot, y_B = 0_B)$, respectively. ■

LEMMA A.2. *For a non-empty $D \subset V$ consider $Y \sim \mathbb{P}_{R_D \times \mathbb{R}^{V\setminus D}}$ with $R_D \in \mathcal{R}(\Lambda_D)$. Then*

(a) $Y_D \sim \mathbb{P}_{R_D}$.
(b) *For any $R = R_D \times R_{V\setminus D}$ it holds that $\mathbb{P}_R$ is the conditional law of Y given $Y_{V\setminus D} \in R_{V\setminus D}$.*

PROOF. (a) follows immediately from (7) and the definition of the marginal measure in (9).
(b) is seen from

$$\frac{\Lambda(\mathrm{d}y)}{\Lambda(R)} = \frac{\Lambda(\mathrm{d}y)}{\Lambda(R_D \times \mathbb{R}^{V\setminus D})} \frac{\Lambda(R_D \times \mathbb{R}^{V\setminus D})}{\Lambda(R)} = \frac{\mathbb{P}(Y \in \mathrm{d}y)}{\mathbb{P}(Y \in R)}$$

for $y \in R$ and $Y_D \in R_D$ a.s. ■

A.2. Restriction and extension. We start with a few fundamental results which follow directly from basic probability theory. We provide short arguments for completeness. Consider a partition (A,B,C) of V. For two product-form sets $R,R' \in \mathcal{R}(\Lambda)$ the corresponding random vectors as in (7) are denoted by $Y \sim \mathbb{P}_R$ and $Y' \sim \mathbb{P}_{R'}$ in the following.

LEMMA A.3. *Let $C \neq \emptyset$. Assume that $R_{A\cup B} \supset R'_{A\cup B} \supset E_{A\cup B}$, where the latter is some Borel set. Then*

$$\mathbb{P}(Y'_{A\cup B} \in E_{A\cup B} | Y'_C = y_C) = \frac{\mathbb{P}(Y_{A\cup B} \in E_{A\cup B} | Y_C = y_C)}{\mathbb{P}(Y_{A\cup B} \in R'_{A\cup B} | Y_C = y_C)}$$

for $\mathbb{P}_{Y'_C}$-almost all $y_C \in R_C \cap R'_C$.

PROOF. Choose a Borel set $E_C \subset R_C \cap R'_C$ and observe that

$$\mathbb{P}(Y'_{A\cup B} \in E_{A\cup B}, Y'_C \in E_C) = \mathbb{P}(Y_{A\cup B} \in E_{A\cup B}, Y_C \in E_C)\frac{\Lambda(R)}{\Lambda(R')}.$$

The right hand side can be rewritten as

$$\int_{E_C} \frac{\mathbb{P}(Y_{A\cup B} \in E_{A\cup B} | Y_C = y_C)}{\mathbb{P}(Y_{A\cup B} \in R'_{A\cup B} | Y_C = y_C)} \mathbb{P}(Y_{A\cup B} \in R'_{A\cup B}, Y_C \in \mathrm{d}y_C)\frac{\Lambda(R)}{\Lambda(R')},$$

where the denominator is non-zero for $\mathbb{P}_{Y'_C}$-almost all $y_C \in E_C$. Note that the scaled measure corresponds to $\mathbb{P}(Y'_C \in \mathrm{d}y)$ and thus we obtain

$$\mathbb{P}(Y'_{A\cup B} \in E_{A\cup B}, Y'_C \in E_C) = \int_{E_C} \frac{\mathbb{P}(Y_{A\cup B} \in E_{A\cup B} | Y_C = y_C)}{\mathbb{P}(Y_{A\cup B} \in R'_{A\cup B} | Y_C = y_C)} \mathbb{P}(Y'_C \in \mathrm{d}y_C),$$

and the result follows. ■

REMARK A.4. (a) If in addition $Y_A \perp\!\!\!\perp Y_B \mid Y_C$ in Lemma A.3, by setting $E_{A\cup B} = E_A \times R'_B$, we can further conclude that for $\mathbb{P}_{Y'_C}$-almost all $y_C \in R_C \cap R'_C$

$$\mathbb{P}(Y'_A \in E_A | Y'_C = y_C) = \frac{\mathbb{P}(Y_A \in E_A | Y_C = y_C)}{\mathbb{P}(Y_A \in R'_A | Y_C = y_C)},$$

(and the analogous identity with reversed roles for A and B).

(b) We also note that, when $R'_{A\cup B} = R_{A\cup B}$ in Lemma A.3, the two conditional probabilities $\mathbb{P}(Y'_{A\cup B} \in E_{A\cup B} | Y'_C = y_C)$ and $\mathbb{P}(Y_{A\cup B} \in E_{A\cup B} | Y_C = y_C)$ coincide for almost all $y_C \in R_C \cap R'_C$ with respect to both probability measures $\mathbb{P}_{Y_C}$ and $\mathbb{P}_{Y'_C}$.

(c) For $C = \emptyset$ the analogous identity

$$\mathbb{P}(Y'_{A\cup B} \in E_{A\cup B}) = \frac{\mathbb{P}(Y_{A\cup B} \in E_{A\cup B})}{\mathbb{P}(Y_{A\cup B} \in R'_{A\cup B})}$$

holds trivially.

LEMMA A.5 (Restriction). *Assume that $R' \subset R$. Then $Y_A \perp\!\!\!\perp Y_B \mid Y_C$ implies $Y'_A \perp\!\!\!\perp Y'_B \mid Y'_C$. This result is also true for $C = \emptyset$ and the respective independence statements.*

PROOF. Consider Borel subsets $E_A \subset R'_A$ and $E_B \subset R'_B$. According to Lemma A.3 and conditional independence $Y_A \perp\!\!\!\perp Y_B \mid Y_C$ we have

$$\mathbb{P}(Y'_A \in E_A, Y'_B \in E_B | Y'_C = y_C) = \frac{\mathbb{P}(Y_A \in E_A | Y_C = y_C)}{\mathbb{P}(Y_A \in R'_A | Y_C = y_C)} \frac{\mathbb{P}(Y_B \in E_B | Y_C = y_C)}{\mathbb{P}(Y_B \in R'_B | Y_C = y_C)}$$

for $\mathbb{P}_{Y'_C}$-almost all $y_C \in R'_C$. By considering the same identity with $E_B = R'_B$ and then $E_A = R'_A$ and applying Lemma A.3 again, we find that the two ratios on the right-hand side are $\mathbb{P}(Y'_A \in E_A | Y'_C = y_C)$ and $\mathbb{P}(Y'_B \in E_B | Y'_C = y_C)$, respectively, and the first claim is proven. The case $C = \emptyset$ follows analogously from the trivial identity

$$\mathbb{P}(Y'_A \in E_A, Y'_B \in E_B) = \frac{\mathbb{P}(Y_A \in E_A, Y_B \in E_B)}{\mathbb{P}(Y_A \in R'_A, Y_B \in R'_B)}. \qquad \blacksquare$$

LEMMA A.6 (Extension via C). *Let $C \neq \emptyset$ and assume that $R'_{A\cup B} = R_{A\cup B}$. Then $Y_A \perp\!\!\!\perp Y_B \mid Y_C$ and $Y'_A \perp\!\!\!\perp Y'_B \mid Y'_C$ imply $Y''_A \perp\!\!\!\perp Y''_B \mid Y''_C$ with $Y'' \sim \mathbb{P}_{R\cup R'}$.*

PROOF. Lemma A.3 shows that for any Borel subsets $E_A \subset R_A, E_B \subset R_B$ we have

$$\begin{aligned}
&\mathbb{P}(Y''_A \in E_A, Y''_B \in E_B | Y''_C = y_C) = \mathbb{P}(Y_A \in E_A, Y_B \in E_B | Y_C = y_C) \\
&= \mathbb{P}(Y_A \in E_A | Y_C = y_C)\mathbb{P}(Y_B \in E_B | Y_C = y_C) \\
&= \mathbb{P}(Y''_A \in E_A | Y''_C = y_C)\mathbb{P}(Y''_B \in E_B | Y''_C = y_C)
\end{aligned}$$

for $\mathbb{P}_{Y''_C}$-almost all $y_C \in R_C$. Analogously, we get a factorization for $\mathbb{P}_{Y''_C}$-almost all $y_C \in R'_C$ when considering Y' instead of Y. This shows the stated result. $\blacksquare$

In Lemma A.6 we verify an extension result for conditional independence from two sets R and R' with identical $A \cup B$ component to their union $R \cup R'$. Naturally, the analogous extension result holds also for extending conditional independence from countably many product-form sets $R^{(n)}$ with $R^{(n)}_{A\cup B} = R_{A\cup B}$ for all n, resulting in the conditional independence statement on the union $\bigcup_n R^{(n)}$.

A.3. Main proofs for characterization results. We are now ready to prove our first characterization result.

PROOF OF THEOREM 4.1. By definition $A \perp B \mid C\,[\Lambda]$ implies $Y_A \perp\!\!\!\perp Y_B \mid Y_C$ with $Y \sim \mathbb{P}_{R_{v,\varepsilon}}$ for any $v \in V$ and $\varepsilon > 0$ such that $\Lambda(R_{v,\varepsilon}) > 0$. Additionally, it implies $A \perp B\,[\Lambda^0_{A\cup B}]$, see Lemma 3.3. Thus, $A \perp B \mid C\,[\Lambda]$ implies both (i) and (ii), also when $C = \emptyset$.

(i) In the opposite direction, we take any $R = \times R_v \in \mathcal{R}(\Lambda)$ and note that there exists $v \in V$ such that the closure of R_v does not contain 0. Hence we may choose $\varepsilon > 0$ such that $R \subset R_{v,\varepsilon}$. According to Lemma A.5 we have $Y_A \perp\!\!\!\perp Y_B \mid Y_C$ with $Y \sim \mathbb{P}_R$. Hence $A \perp B \mid C\,[\Lambda]$. The argument also applies when $C = \emptyset$.

(ii) It is sufficient to establish (i), and so we take any $v \in A \cup B$, $\varepsilon > 0$ and consider the set $R_{v,\varepsilon}$. We partition this set according to $y_C \neq 0_C$ and $y_C = 0_C$, letting Y and Y' be the respective random vectors. Here we assume that both sets are charged by Λ, since otherwise the proof is only simpler. According to Lemma A.6 it is sufficient to show both $Y_A \perp\!\!\!\perp Y_B \mid Y_C$ and $Y'_A \perp\!\!\!\perp Y'_B \mid Y'_C$.

The latter follows directly from $A \perp B\,[\Lambda^0_{A\cup B}]$, see also the proof of Lemma 3.3. The former follows from the countably-infinite version of Lemma A.6 by considering the sets $R^{(n,c)} = \{|y_v| \geq \varepsilon, |y_c| \geq 1/n\}$ for all $c \in C$ and natural n. The respective conditional independence statements are implied by the assumption in (ii) together with the restriction Lemma A.5. $\blacksquare$

As a side result the same argument that we used in proof of Theorem 4.1 (ii) above reveals that a single $v \in V$ can be taken when checking for conditional independence under the condition that $\Lambda(y \neq 0_V, y_v = 0) = 0$. To be precise we formulate the following lemma.

LEMMA A.7 (Single v test). *Assume that a fixed $v \in V$ is such that $\Lambda(y \neq 0_V, y_v = 0) = 0$. Then $Y_A \perp\!\!\!\perp Y_B \mid Y_C$ for all $Y \sim \mathbb{P}_{R_{v,\varepsilon}}, \varepsilon > 0$ implies $A \perp B \mid C\ [\Lambda]$. This result is also true for $C = \emptyset$ and the respective independence statements.*

Next we prove uniqueness and representation of the kernel.

PROOF OF LEMMA 4.3. Let R^1_C be some product form set bounded away from 0_C and define $R^1 = \mathbb{R}^{A\cup B} \times R^1_C$. Suppose for a moment that $R_C \subset R^1_C$. Letting $Y \sim \mathbb{P}_{R^1}$ (the case of zero mass being trivial) we observe that

$$\Lambda(R) = \mathbb{P}(Y_{A\cup B} \in R_{A\cup B}, Y_C \in R_C)\Lambda(R^1) = \int_{y_C \in R_C} \nu^1_C(y_C, R_{A\cup B})\mathbb{P}(Y_C \in \mathrm{d}y_C)\Lambda(R^1),$$

where $\nu^1_C(y_C, \cdot)$ is a probability kernel corresponding to Y, see Kallenberg (2002, Thm. 6.3). Note that $\Lambda_C(\mathrm{d}y_C) = \mathbb{P}(Y_C \in \mathrm{d}y_C)\Lambda(R^1), y_C \in R_C$ giving the stated representation of $\Lambda(R)$. In general, we partition $\mathcal{E}^C$ into countably many product form sets R^i_C and define $\nu_C(y_C, \cdot) = \nu^i_C(y_C, \cdot)$ for y_C in the respective partition. The stated form of ν_C and Λ_C-uniqueness follow from Lemma A.3. ∎

APPENDIX B: FURTHER DETAILS FOR THE GENERIC CONSTRUCTION OF SECTION 5.4

Here we provide further details on the constructive example of a measure Λ in $d = 3$ dimensions with index set $V = \{1, 2, 3\}$ such that the conditional independence $\{1\} \perp \{3\} \mid \{2\}\ [\Lambda]$ holds. Table 2 contains the reformulation of the factorization of the modified density (21) in terms of original building blocks, that is, the densities κ_{ij}, their joint marginal density m and the mixture probabilities p_{ij} with $q_{ij} = 1 - p_{ij}$. Figure 6 illustrates them for faces up to dimension 2.

We note that if the starting density κ is $-\alpha$-homogeneous for some $\alpha > 2$, it is easily checked that its marginal density m will be $-(\alpha-1)$-homogeneous and the resulting measure Λ will be $-(\alpha - 2)$-homogeneous; see Example 8.8.

TABLE 2

Modified densities of the trivariate example from Section 5.4; see also Figure 6. The resulting measure Λ satisfies $\{1\} \perp \{3\} \mid \{2\}\ [\Lambda]$ and has identical one-dimensional marginals. It is globally Markov for the chain graph $1 - 2 - 3$ with zero mass on $\mathcal{Z} = \{y_1 \neq 0, y_2 = 0, y_3 \neq 0\}$; see Theorem 6.4.

Dimension	**Faces** y_1	y_2	y_3	**Modified densities** $\overline{\lambda}_{12}(y_{12})$	$\overline{\lambda}_{23}(y_{23})$	$\overline{\lambda}_2(y_2)$	$\overline{\lambda}(y) = \lambda(y)$
1	0	0	$\neq 0$	1	$q_{23}m(y_3)$	1	$q_{23}m(y_3)$
	$\neq 0$	0	0	$q_{12}m(y_1)$	1	1	$q_{12}m(y_1)$
	0	$\neq 0$	0	$q_{12}m(y_2)$	$q_{23}m(y_2)$	$m(y_2)$	$q_{12}q_{23}m(y_2)$
2	$\neq 0$	$\neq 0$	0	$p_{12}\kappa_{12}(y_{12})$	$q_{23}m(y_2)$	$m(y_2)$	$p_{12}q_{23}\kappa_{12}(y_{12})$
	0	$\neq 0$	$\neq 0$	$q_{12}m(y_2)$	$p_{23}\kappa_{23}(y_{23})$	$m(y_2)$	$q_{12}p_{23}\kappa_{23}(y_{23})$
	$\neq 0$	0	$\neq 0$	$q_{12}m(y_1)$	$q_{23}m(y_3)$	1	0
3	$\neq 0$	$\neq 0$	$\neq 0$	$p_{12}\kappa_{12}(y_{12})$	$p_{23}\kappa_{23}(y_{23})$	$m(y_2)$	$\frac{p_{12}p_{23}\kappa_{12}(y_{12})\kappa_{23}(y_{23})}{m(y_2)}$

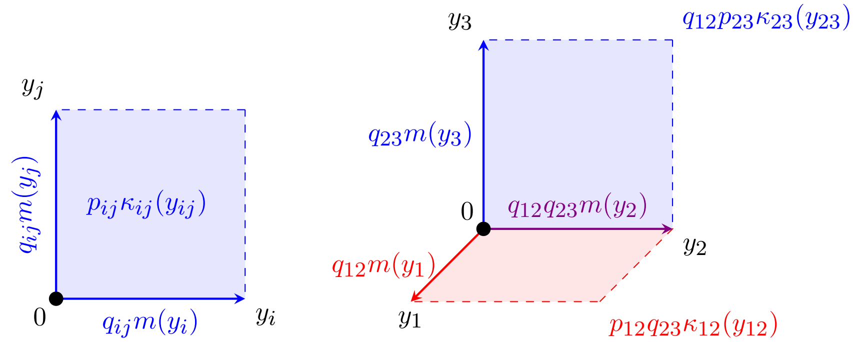

FIG 6. *Illustration of the generic construction of a measure Λ satisfying $\{1\} \perp \{3\} \mid \{2\}\ [\Lambda]$ from Section 5.4. Left: measure λ_{ij}, which serves as building block. Right: resulting densities λ on the faces up to dimension 2; see Table 2.*

APPENDIX C: PROOFS AND EXAMPLES FOR SECTION 6: UNDIRECTED GRAPHICAL MODELS

C.1. Auxiliary result for the Hammersley-Clifford-type theorem.

PROPOSITION C.1. *Assume that Λ satisfies (E1) and that $V = D \cup C$ is the vertex set of a graph $\mathcal{G}$, where C is a clique and $\widetilde{S} = D \cap C$ separates $D \setminus \widetilde{S}$ from $C \setminus \widetilde{S}$. Then Λ satisfies the global Markov property with respect to $\mathcal{G}$ if and only if*

$$D \setminus \widetilde{S} \perp C \setminus \widetilde{S} \mid \widetilde{S}\ [\Lambda] \tag{43}$$

and Λ is globally Markov when restricted to D.

PROOF OF PROPOSITION C.1. The global Markov property for Λ implies (43) and, since $\widetilde{S} \subset C$ is fully connected, the global Markov property for Λ restricted to D, cf. Lauritzen (1996, Prop. 3.17 (start of proof)). So we only need to establish the other direction, i.e. show the global Markov property for Λ from these two properties.

To this end, consider disjoint sets $A, B, S \subset V$ with S separating A from B. As outlined in Remark 6.2, it suffices to consider the partition situation, where $A \cup B \cup S = V$. Since C is a clique, the sets A and B cannot both intersect C, otherwise they could not be separated by any S, so we may assume that $B \cap C = \emptyset$.

Let us first consider the case that $S \subset D$. Since Λ satisfies (E1), we can employ the semi-graphoid property (L4). We do so with the four sets $(B, A \cap D, S, A \setminus D)$ as the sets (A, B, C, D) from the original formulation of (L4). The first condition there reads as $B \perp (A \cap D) \mid S$, which is implied by the global Markov property for D. The second condition reads as $B \perp (A \setminus D) \mid S \cup (A \cap D)$, which follows from the other assumption, $(D \setminus \widetilde{S}) \perp (A \setminus D) \mid \widetilde{S}$, by means of (L3). As a result we get $A \perp B \mid S$ as required.

If instead $E = S \cap (C \setminus \widetilde{S}) \neq \emptyset$, set $A' = A \cup E$ and $S' = S \setminus E$. Then we can finish the proof by showing that S' still separates A' from B, but with $S' \subset D$, so that we can employ the above argument again to obtain $A' \perp B \mid S'$, which subsequently implies $A \perp B \mid S$ by (L3). It remains to be seen that indeed S' separates A' from B. Let $b \in B$ and $a \in A' = A \cup E$. We need to show that any path π from b to a (if such a path exists) necessarily visits $S' = S \setminus E$.

Case 1: $a \in E$. Since $\widetilde{S}$ separates $D \setminus \widetilde{S}$ from $C \setminus \widetilde{S}$, $\widetilde{S}$ also separates their subsets B and E. In particular, π must pass through an element $v \in \widetilde{S}$ before reaching $C \setminus \widetilde{S}$ for the first time. Since $\widetilde{S} \subset A \cup S'$, either $v \in S'$ (and we are done) or $v \in A$. However, if $v \in A$, recall that

S separates B from A, so π must first pass through an element $s \in S$ before reaching $v \in A$. Here, s cannot belong to E, since π has not yet arrived in $C \setminus \widetilde{S}$ for the first time. Hence $s \in S' = S \setminus E$.

Case 2: $a \in A$. Since S separates B from A, π needs to visit some $s \in S$. If $s \in S' = S \setminus E$, we are done. Else $s \in E$. But then the argument from Case 1 shows that on the way between b and $s \in E$, the path π needs to visit S'. ■

C.2. Proof of the Hammersley-Clifford-type theorem.

PROOF OF THEOREM 6.4. Let us prove the 'if' and 'only if' directions separately.

"Global Markov property $\Rightarrow$ Factorization:" First we prove that the global Markov property of Λ implies the factorization (23) by induction on the number of cliques. The statement holds trivially for a single clique, as there are no separators, so the factorization reads $\overline{\lambda}(y) = \overline{\lambda}(y)$ for all $y \in \mathbb{R}^{C_1}$, which is always true. We let $\mathcal{G}_j$ be the decomposable graph that arises from restricting $\mathcal{G}$ to D_j and abbreviate $\mathcal{Z}_j = \mathcal{Z}(\mathcal{G}_j)$ henceforth. Let us assume that the implication holds for $n-1$ cliques, that is, $\Lambda_{D_{n-1}}$ being globally Markov with respect to $\mathcal{G}_{n-1}$ implies

$$\overline{\lambda}(y_{D_{n-1}}) \prod_{j<n-1} \overline{\lambda}_{S_j}(y_{S_j}) = \prod_{j<n} \overline{\lambda}_{C_j}(y_{C_j}). \tag{44}$$

for $\mu_{D_{n-1}}$-almost all $y_{D_{n-1}} \notin \mathcal{Z}_{n-1}$. For the induction step, suppose now that Λ is globally Markov with respect to $\mathcal{G}_n$. We need to establish (23) for μ-almost all $y \notin \mathcal{Z}_n$.

The global Markov property of Λ for $\mathcal{G}_n$ implies also the global Markov property of $\Lambda_{D_{n-1}}$ for $\mathcal{G}_{n-1}$ and we obtain (44) as well by the induction assumption. It is easily seen that

$$\mathbb{R}^V \setminus \mathcal{Z}_n \subset (\mathbb{R}^{D_{n-1}} \setminus \mathcal{Z}_{n-1}) \times \mathbb{R}^{V \setminus D_{n-1}}$$

and so (44) holds for μ-almost all $y \notin \mathcal{Z}_n$. According to the running intersection property we have $C_n \setminus S_{n-1} \perp D_{n-1} \setminus S_{n-1} \mid S_{n-1}\ [\Lambda]$, which in view of Proposition 5.5 is equivalent to

$$\overline{\lambda}(y)\overline{\lambda}_{S_{n-1}}(y_{S_{n-1}}) = \overline{\lambda}_{C_n}(y_{C_n})\overline{\lambda}_{D_{n-1}}(y_{D_{n-1}})$$

for μ-almost all $y \in \mathbb{R}^V$ excluding those with $y_{S_{n-1}} = 0_{S_{n-1}}$ and $y_{C_n \setminus S_{n-1}} \neq 0_{C_n \setminus S_{n-1}}$ and $y_{D_{n-1} \setminus S_{n-1}} \neq 0_{D_{n-1} \setminus S_{n-1}}$. The factorization therefore holds for μ-almost all $y \notin \mathcal{Z}_n$, since for any partition (A, B, S) of V, where $S \in \mathcal{S}$ separates A from B in the graph $\mathcal{G}$ we have

$$\{y : y_A \neq 0_A, y_B \neq 0_B, y_S = 0_S\} = \bigcup_{a \in A, b \in B} \{y : y_a \neq 0, y_b \neq 0, y_S = 0_S\} \subset \mathcal{Z}(\mathcal{G}).$$

Here S_{n-1} is also allowed to be an empty set and we recall that $\overline{\lambda}_\emptyset = 1$. Multiplying both sides by $\prod_{j<n-1} \overline{\lambda}_{S_j}(y_{S_j})$ and using (44) yields (23) for μ-almost all $y \notin \mathcal{Z}_n$, and taken together this completes the induction step.

"Factorization $\Rightarrow$ Global Markov property:" The proof in the other direction is more subtle. We need to show that, if the factorization (23) holds for μ-almost all $y \notin \mathcal{Z}(\mathcal{G})$, then Λ is globally Markov with respect to $\mathcal{G}$. Again, we use induction on the number of cliques, where the case of a single clique is trivial, as there are no separators, so Λ_{C_1} is always globally Markov with respect to $\mathcal{G}_1$, a true conclusion. Our induction assumption is now that (44) for $\mu_{D_{n-1}}$-almost all $y_{D_{n-1}} \notin \mathcal{Z}_{n-1}$ implies that $\Lambda_{D_{n-1}}$ is globally Markov with respect to $\mathcal{G}_{n-1}$. We need to show that Λ is globally Markov with respect to $\mathcal{G}_n$, whilst assuming (23) for μ-almost all $y \notin \mathcal{Z}_n$.

Note that $\{y_{D_{n-1}} = 0_{D_{n-1}}\} \subset \mathbb{R}^V \setminus \mathcal{Z}_n$. So we obtain from (23) that for μ-almost all $y \in \mathcal{E}$ with $y_{D_{n-1}} = 0_{D_{n-1}}$

$$\lambda(0_{D_{n-1}}, y_{C_n \setminus S_{n-1}})\mu_{D_{n-1}}(\{0_{D_{n-1}}\}) = \lambda_{C_n}(0_{S_{n-1}}, y_{C_n \setminus S_{n-1}})\mu_{S_{n-1}}(\{0_{S_{n-1}}\}),$$

where the left-hand side and right-hand side are the densities of the reduced measures

$$\Lambda^0_{C_n\setminus S_{n-1}} = (\Lambda^0_{V\setminus S_{n-1}})^0_{C_n\setminus S_{n-1}} \quad \text{and} \quad (\Lambda_{C_n})^0_{C_n\setminus S_{n-1}} = (\Lambda^0_{V\setminus S_{n-1}})_{C_n\setminus S_{n-1}},$$

respectively, where the last two identities follow from Lemma A.1 with $A = D_{n-1}\setminus S_{n-1}$, $B = S_{n-1}$ and $C = C_n \setminus S_{n-1}$ therein. By Corollary 5.2 we obtain the independence statement for the restricted measure $\Lambda^0_{V\setminus S_{n-1}}$, that is,

$$C_n \setminus S_{n-1} \perp D_{n-1}\setminus S_{n-1}\,[\Lambda^0_{V\setminus S_{n-1}}], \tag{45}$$

which further implies by Corollary 5.2 and Lemma A.1 that

$$\Lambda^0_{D_{n-1}\setminus S_{n-1}} = (\Lambda^0_{V\setminus S_{n-1}})^0_{D_{n-1}\setminus S_{n-1}} = (\Lambda^0_{V\setminus S_{n-1}})_{D_{n-1}\setminus S_{n-1}} = (\Lambda_{D_{n-1}})^0_{D_{n-1}\setminus S_{n-1}}.$$

Therefore, we have for μ-almost all $y\in\mathcal{E}$ with $y_{C_n} = 0_{C_n}$ that

$$\lambda(y_{D_n\setminus S_{n-1}}, 0_{C_n})\mu_{C_n}(\{0_{C_n}\}) = \lambda_{D_{n-1}}(y_{D_{n-1}\setminus S_{n-1}}, 0_{S_{n-1}})\mu(\{0_{S_{n-1}}\}). \tag{46}$$

Further we obtain from (23) that for μ-almost all $y \in \mathcal{E}\setminus\mathcal{Z}_n$ with $y_{C_n} = 0_{C_n}$

$$\Big[\lambda(y_{D_n\setminus S_{n-1}}, 0_{C_n})\mu_{C_n\setminus S_{n-1}}(\{0_{C_n\setminus S_{n-1}}\})\Big]\prod_{j<n-1}\overline{\lambda}_{S_j}(y_{S_j}) = \prod_{j<n}\overline{\lambda}_{C_j}(y_{C_j}),$$

where the term in brackets can be replaced by $\lambda_{D_{n-1}}(y_{D_{n-1}\setminus S_{n-1}}, 0_{S_{n-1}})$ due to (46). Note that $y\in\mathcal{E}\setminus\mathcal{Z}_n$ with $y_{C_n} = 0_{C_n}$ implies $y_{D_{n-1}} \in \mathcal{E}^{D_{n-1}}\setminus\mathcal{Z}_{n-1}$ with $y_{S_{n-1}} = 0_{S_{n-1}}$, and conversely, if $y_{D_{n-1}} \in \mathcal{E}^{D_{n-1}}\setminus\mathcal{Z}_{n-1}$ with $y_{S_{n-1}} = 0_{S_{n-1}}$, then $(y_{D_{n-1}}, 0_{C_n\setminus S_{n-1}}) \in \mathcal{E}\setminus\mathcal{Z}_n$ and $y_{C_n} = 0_{C_n}$. Hence, we recover (44) for $\mu_{D_{n-1}}$-almost all $y_{D_{n-1}}\notin\mathcal{Z}_{n-1}$ with $y_{S_{n-1}} = 0_{S_{n-1}}$ (note that it is trivially satisfied for $y_{D_{n-1}} = 0_{D_{n-1}}$). If $S_{n-1} = \emptyset$, the same line of reasoning (with $\mu_{S_{n-1}}(\{0_{S_{n-1}}\})$ replaced by 1 above) even recovers (44) for $\mu_{D_{n-1}}$-almost all $y_{D_{n-1}}\notin\mathcal{Z}_{n-1}$.

Else $S_{n-1}\neq\emptyset$ and we consider $y\notin\mathcal{Z}_n$ with $y_{S_{n-1}}\neq 0_{S_{n-1}}$, so that $y_{D_{n-1}}\notin\mathcal{Z}_{n-1}$ with $y_{S_{n-1}}\neq 0_{S_{n-1}}$. Conversely, such a $y_{D_{n-1}}$ satisfies $y\notin\mathcal{Z}_n$ with $y_{S_{n-1}}\neq 0_{S_{n-1}}$ for any(!) $y_{C_n\setminus S_{n-1}}\in\mathbb{R}^{C_n\setminus S_{n-1}}$. So we can integrate out the $C_n\setminus S_{n-1}$ components in (23) for such y to obtain

$$\lambda_{D_{n-1}}(y_{D_{n-1}})\lambda_{S_{n-1}}(y_{S_{n-1}})\prod_{j<n-1}\overline{\lambda}_{S_j}(y_{S_j}) = \lambda_{S_{n-1}}(y_{S_{n-1}})\prod_{j<n}\overline{\lambda}_{C_j}(y_{C_j})$$

for $\mu_{D_{n-1}}$-almost all $y_{D_{n-1}}\notin\mathcal{Z}_{n-1}$ with $y_{S_{n-1}}\neq 0_{S_{n-1}}$. This formula is still true when the term $\lambda_{S_{n-1}}(y_{S_{n-1}})$ is dropped on both sides, since for μ-almost all y with $y_{S_{n-1}}\neq 0_{S_{n-1}}$, for which the marginal density $\lambda_{S_{n-1}}(y_{S_{n-1}})$ is zero, we also must have zero densities $\lambda_{D_{n-1}}(y_{D_{n-1}})$ and $\lambda_{C_{n-1}}(y_{C_{n-1}}) = \overline{\lambda}_{C_{n-1}}(y_{C_{n-1}})$, as $\lambda_{S_{n-1}}(y_{S_{n-1}})$ arises as a marginal density from them. So, collectively, we recover (44) for $\mu_{D_{n-1}}$-almost all $y_{D_{n-1}}\notin\mathcal{Z}_{n-1}$.

The induction assumption now gives us that $\Lambda_{D_{n-1}}$ is globally Markov with respect to $\mathcal{G}_{n-1}$. Hence, by Proposition C.1 it suffices to establish

$$C_n\setminus S_{n-1}\perp D_{n-1}\setminus S_{n-1}\mid S_{n-1}\,[\Lambda],$$

which is our goal in what follows. If $S_{n-1}=\emptyset$, it follows readily from (45) and the proof is complete.

Else $S_{n-1}\neq\emptyset$ and (45) only yields the independence of the reduced measure and by Theorem 4.6 all that remains to be seen is that

$$\lambda(y)\lambda_{S_{n-1}}(y_{S_{n-1}}) = \lambda_{C_n}(y_{C_n})\lambda_{D_{n-1}}(y_{D_{n-1}}) \tag{47}$$

for μ-almost all y with $y_{S_{n-1}}\neq 0_{S_{n-1}}$.

Here, we note that (23) for μ-almost all $y \notin \mathcal{Z}_n$ together with (44) for $\mu_{D_{n-1}}$-almost all $y_{D_{n-1}} \notin \mathcal{Z}_{n-1}$ gives

$$\lambda(y)\lambda_{S_{n-1}}(y_{S_{n-1}}) \prod_{j<n-1} \overline{\lambda}_{S_j}(y_{S_j}) = \lambda_{C_n}(y_{C_n})\lambda_{D_{n-1}}(y_{D_{n-1}}) \prod_{j<n-1} \overline{\lambda}_{S_j}(y_{S_j})$$

for μ-almost all $y \notin \mathcal{Z}_n$, $y_{S_{n-1}} \neq 0_{S_{n-1}}$. The product terms $\overline{\lambda}_{S_j}(y_{S_j})$ can either be canceled by the same marginal cancellation argument as above if $y_{S_j} \neq 0_{S_j}$ or else due to $\overline{\lambda}_{S_j}(0_{S_j}) = 1/\mu_{S_j}(\{0_{S_j}\})$. Hence we have (47) for μ-almost all $y \notin \mathcal{Z}_n$, $y_{S_{n-1}} \neq 0_{S_{n-1}}$.

Finally, we note that, if $y_{S_{n-1}} \neq 0_{S_{n-1}}$, then $y \notin \mathcal{Z}_n$ is equivalent to $y_{D_{n-1}} \notin \mathcal{Z}_{n-1}$. By the global Markov property on D_{n-1} we know $\Lambda_{D_{n-1}}(\mathcal{Z}_{n-1}) = 0$. Hence, both sides of (47) are zero for μ-almost all y with $y_{S_{n-1}} \neq 0_{S_{n-1}}$ and $y_{D_{n-1}} \in \mathcal{Z}_{n-1}$. Taken together, this establishes (47) in the general case for μ-almost all y with $y_{S_{n-1}} \neq 0_{S_{n-1}}$ and the proof is complete. ∎

C.3. Example for Remark 6.6. Consider the chain graph $1 - 2 - 3 - 4 - 5$ on the vertices $V = \{1, \dots, 5\}$ where the neighboring indices are connected. We construct a measure Λ on $\mathcal{E}$ which is not(!) globally Markov with respect to this graph $\mathcal{G}$, even though

(i) (E1) holds;
(ii) the density λ satisfies the factorization

$$\lambda(y)\lambda_2(y_2)\lambda_3(y_3)\lambda_4(y_4) = \lambda_{12}(y_{12})\lambda_{23}(y_{23})\lambda_{34}(y_{34})\lambda_{45}(y_{45})$$

for μ-almost all y such that $y_2, y_3, y_4 \neq 0$;
(iii) $\Lambda(\mathcal{Z}(\mathcal{G})) = 0$, that is, $\Lambda(y_k \neq 0, y_l = 0, y_m \neq 0) = 0$ whenever $k < l < m$.

As in Section 5.4 we consider the same dominating measure μ as in (19) (here for $d = 5$) and use some -3-homogeneous bivariate densities κ_{ij} on $(0, \infty)^2$ with marginal densities y^{-2} as building blocks, now for $(i, j) \in \{(1, 2), (2, 3), (3, 4), (4, 5)\}$. In addition, we ensure that κ_{12} and κ_{23} are compatible with a -4-homogeneous trivariate density η_{123} on $(0, \infty)^3$ as follows:

- $\eta_{12} = \kappa_{12}, \eta_{23} = \kappa_{23}$ and, thus, respective univariate marginals have densities y^{-2},
- $\eta_{123}(y_{123})\eta_2(y_2) \neq \eta_{12}(y_{12})\eta_{23}(y_{23})$ on a subset of $(0, \infty)^3$ of positive Lebesgue measure.

The second condition is equivalent to saying that the measure corresponding to η_{123} does not induce the conditional independence of 1 and 3 given 2, see Theorem 4.6. Such a construction is always possible, e.g. one can start from η_{123} with the respective univariate marginals and then work with κ_{12} and κ_{23} that arise from η_{123}. From there, we define

$$\lambda(y) = \begin{cases} \kappa_{12}(y_{12})\kappa_{23}(y_{23})\kappa_{34}(y_{34})\kappa_{45}(y_{45})y_2^2 y_3^2 y_4^2, & y_1, y_2, y_3, y_4, y_5 > 0, \\ \eta_{123}(y_{123}), & y_1, y_2, y_3 > 0, y_4 = y_5 = 0, \end{cases}$$

and let it be 0 otherwise. By construction, the measure Λ satisfies the conditions (i) and (iii). Furthermore, the bivariate marginals on $(0, \infty)^2$ are given by

$$\lambda_{12} = \kappa_{12} + \eta_{12} = 2\kappa_{12}, \quad \lambda_{23} = \kappa_{23} + \eta_{23} = 2\kappa_{23}, \quad \lambda_{34} = \kappa_{34}, \quad \lambda_{45} = \kappa_{45},$$

and the univariate ones by $\lambda_2(y_2) = 2y_2^{-2}, \lambda_3(y_3) = 2y_3^{-2}, \lambda_4(y_4) = y_4^{-2}$. This readily yields the factorization in (ii). Finally, we note that Λ is not globally Markov for the chain graph since $\{1\} \perp \{3, 4, 5\} \mid \{2\}\ [\Lambda]$ is not true, because

$$\lambda(y)\lambda_2(y_2) = \eta_{123}(y_{123})2y_2^{-2} \neq 2\eta_{12}(y_{12})\eta_{23}(y_{23}) = \lambda_{12}(y_{12})\lambda_{2345}(y_{2345})$$

on the subset $\{y : y_1, y_2, y_3 > 0, y_4 = y_5 = 0\}$, which has positive μ-mass. Hence, according to Theorem 4.6 the condition $\{1\} \perp \{3, 4, 5\} \mid \{2\}\ [\Lambda]$ is violated.

Had we checked instead the factorization condition from Theorem 6.4 in terms of $\overline{\lambda}$ on this set, we would have picked up the violation of the global Markov property. The left-hand side of (23) reads $\eta_{123}(y_{123})2y_2^{-2}2y_3^{-2}$ in this case, and the right-hand side is $2\eta_{12}(y_{12})2\eta_{23}(y_{23})y_3^{-2}$. They cannot be equal due to our choice of η_{123}.

APPENDIX D: PROOFS FOR SECTION 8: RELATION TO CONDITIONAL INDEPENDENCE IN EXTREMES

D.1. Derivation of extremal correlation χ_{ij} in Example 8.8.

PROOF. For $i \nleftrightarrow j$, since Λ is a graphical model on the forest $\mathcal{G}$, the global Markov property implies that $i \perp j\ [\Lambda]$. By Proposition 5.1 we therefore have that $\Lambda(y_i \neq 0, y_j \neq 0) = 0$, which yields $\chi_{ij} = 0$ according to (30).

For $i \leftrightarrow j$ we note that by homogeneity of Λ we have $\chi_{ij} = \Lambda(y_i > 1, y_j > 1)/\Lambda(y_i > 1)$ and due to all marginal densities being $m(y) = y^{-2}$ we have $\Lambda(y_i > 1) = 1$. Because of the density factorization (33) we can first integrate out all components $k \notin \mathrm{path}_{ij}$ to get

$$\Lambda(y_i > 1, y_j > 1) = \int_1^\infty \int_1^\infty \int_{(0,\infty)^{|V(ij)|}} \frac{\prod_{(s,t)\in\mathrm{path}_{ij}} \overline{\lambda}_{st}(y_s, y_t)}{\prod_{s\in V(ij)} \overline{\lambda}_s(y_s)} \mathrm{d}y_i \mathrm{d}y_j \mathrm{d}y_{V(ij)},$$

where $V(ij)$ denotes all nodes on the path from i to j, excluding i and j. Thus

$$\chi_{ij} = \int_1^\infty \int_1^\infty \int_{(0,\infty)^{|V(ij)|}} \frac{\prod_{(s,t)\in\mathrm{path}_{ij}} p_{st}\kappa_{st}(y_s, y_t)}{\prod_{s\in V(ij)} y_s^{-2}} \mathrm{d}y_i \mathrm{d}y_j \mathrm{d}y_{V(ij)} = \chi_{ij}^{\mathrm{HR}} \prod_{(s,t)\in\mathrm{path}_{ij}} p_{st}$$

where $\chi_{ij}^{\mathrm{HR}} = 2 - 2\Phi(\sqrt{\Gamma_{ij}}/2)$ with $\Gamma_{ij} = \sum_{(s,t)\in\mathrm{path}_{ij}} \Gamma_{st}$ is the extremal correlation of the corresponding tree model on the connected component that contains i and j (Engelke and Volgushev, 2022, Proposition 4). This result is true since the domain of the above integral is contained in the set $\{y_i > 0, y_j > 0, y_{V(ij)} > 0\}$, on which the density coincides with the usual Hüsler–Reiss tree model.

■

D.2. Max-infinitely divisibility and asymptotic equivalence of the joint survival functions.

PROOF OF LEMMA 8.9. Denoting $\Lambda_{ij}(u,u) = \Lambda_{ij}(y_i > u \text{ or } y_j > u)$, we can rewrite

$$\mathbb{P}(Z_i > u, Z_j > u) = 1 - 2e^{-\Lambda(u)} + e^{-\Lambda_{ij}(u,u)}. \tag{48}$$

Using $\Lambda_{ij}(u,u) \leq 2\Lambda(u)$ for all $u > 0$ gives

$$\mathbb{P}(Z_i > u, Z_j > u) \geq 1 - 2e^{-\Lambda(u)} + e^{-2\Lambda(u)} = (1 - e^{-\Lambda(u)})^2 = \mathbb{P}(Z_i > u)^2,$$

and hence,

$$\mathbb{P}(Z_i > u, Z_j > u)\mathbb{P}(Z_i > u)^{-2} \geq 1$$

for all $u > 0$. On the other hand (34) tells us that the left-hand side is a function of the form $r(\mathbb{P}(Z_i > u)^{-1})$ with r regularly varying with index $-1/\eta_{ij} + 2$. If we had $\eta_{ij} < 1/2$, the left side would converge to 0 as $u \to \infty$ (e.g., de Haan and Ferreira, 2006, Appendix B), contradicting the last inequality. So we obtain $\eta_{ij} \geq 1/2$.

Further, recall the bounds $1-x \le e^{-x} \le 1-x+x^2/2$ for all $x>0$. Applied to (48) we obtain upper and lower bounds for $\mathbb{P}(Z_i>u, Z_j>u)$ as follows:

$$\begin{aligned}2\Lambda(u)-\Lambda_{ij}(u,u)-\Lambda(u)^2 \le \mathbb{P}(Z_i>u,Z_j>u) &\le 2\Lambda(u)-\Lambda_{ij}(u,u)+\Lambda_{ij}(u,u)^2/2\\ &\le 2\Lambda(u)-\Lambda_{ij}(u,u)+2\Lambda(u)^2.\end{aligned}$$

Since $\Lambda(y_i>u,y_j>u)=2\Lambda(u)-\Lambda_{ij}(u,u)$, we may rewrite these bounds as

$$\mathbb{P}(Z_i>u,Z_j>u)-2\Lambda(u)^2 \le \Lambda(y_i>u,y_j>u) \le \mathbb{P}(Z_i>u,Z_j>u)+\Lambda(u)^2. \tag{49}$$

Hence, if we assume $\eta_{ij}>1/2$, we obtain the assertion (36) from

$$\lim_{u\to\infty}\frac{\Lambda(u)^2}{\mathbb{P}(Z_i>u,Z_j>u)}=0.$$

Note that (36) together with (35) entails (37) with $\widetilde{\eta}_{ij}=\eta_{ij}$.

Conversely, observe that (49) can be rearranged to

$$\Lambda(y_i>u,y_j>u)-\Lambda(u)^2 \le \mathbb{P}(Z_i>u,Z_j>u) \le \Lambda(y_i>u,y_j>u)+2\Lambda(u)^2$$

so that (36) also follows from (37), as (37) entails

$$\lim_{u\to\infty}\frac{\Lambda(u)^2}{\Lambda(y_i>u,y_j>u)}=0.$$

Finally, note that (36) together with (37) gives (35) with $\eta_{ij}=\widetilde{\eta}_{ij}$. ■

D.3. Proof of Proposition 8.11: An asymptotically independent example based on Gaussian distributions. We provide an explicit construction of an exponent measure Λ in dimension 3, which satisfies the conditional independence $\{1\}\perp\{3\}\mid\{2\}\,[\Lambda]$, but exhibits otherwise asymptotic independence. It is based on two bivariate Gaussian distributions and we will need some auxiliary results in order to derive and quantify the asymptotic independence in the sense of Ledford and Tawn (1996).

Let Φ and φ be the cdf and pdf of the univariate standard Gaussian distribution, and let $\overline{\Phi}(x)=1-\Phi(x)=\Phi(-x)$ be the corresponding survival function. It is well-known (DLMF, Eq. 7.8.2) that

$$\frac{2}{x+\sqrt{x^2+4}} < \frac{\overline{\Phi}(x)}{\varphi(x)} < \frac{2}{x+\sqrt{x^2+8/\pi}} < \frac{1}{x}, \qquad x>0. \tag{50}$$

Using (50), one may easily derive the following result by showing positivity of the respective derivatives (e.g., Resnick, 2008, Lemma 5.4).

LEMMA D.1. *The functions Φ/φ and $\varphi/\overline{\Phi}$ are strictly monotonously increasing on its entire domain $\mathbb{R}=(-\infty,\infty)$ from 0 to ∞.*

As any univariate distribution, Φ is max-infinitely divisible. Its exponent measure Λ_1 is supported on $\mathbb{R}=(-\infty,\infty)$ with survival function

$$\Lambda(u)=\Lambda_1((u,\infty))=-\log\Phi(u), \qquad u\in\mathbb{R}$$

and density

$$\lambda_1(u)=\frac{\partial}{\partial u}\log\Phi(u)=\frac{\varphi(u)}{\Phi(u)}, \qquad u\in\mathbb{R}.$$

It is immediate that the survival functions of Λ_1 and Φ are asymptotically equivalent in the sense that

$$\lim_{u\to\infty}\frac{\Lambda(u)}{\overline{\Phi}(u)}=1. \tag{51}$$

Let Φ_ρ and φ_ρ be the cdf and pdf of the bivariate normal distribution with standard Gaussian margins and correlation ρ. Some useful identities, which we will use in several arguments below, are

$$\frac{\partial}{\partial x_1}\Phi_\rho(x_1,x_2)=\Phi\left(\frac{x_2-\rho x_1}{\sqrt{1-\rho^2}}\right)\varphi(x_1),\qquad \frac{\partial}{\partial x_2}\Phi_\rho(x_1,x_2)=\Phi\left(\frac{x_1-\rho x_2}{\sqrt{1-\rho^2}}\right)\varphi(x_2),$$

$$\sqrt{1-\rho^2}\,\varphi_\rho(x_1,x_2)=\varphi\left(\frac{x_2-\rho x_1}{\sqrt{1-\rho^2}}\right)\varphi(x_1)=\varphi\left(\frac{x_1-\rho x_2}{\sqrt{1-\rho^2}}\right)\varphi(x_2).$$

LEMMA D.2. *For $x_1,x_2\in\mathbb{R}$ and $\rho\in(0,1)$*

$$\max\left[\Phi(x_1)\Phi\left(\frac{x_2-\rho x_1}{\sqrt{1-\rho^2}}\right),\Phi(x_2)\Phi\left(\frac{x_1-\rho x_2}{\sqrt{1-\rho^2}}\right)\right]\le\Phi_\rho(x_1,x_2)$$

and

$$\Phi_\rho(x_1,x_2)\le\min\left[\Phi(x_1)\Phi\left(\frac{x_2-\rho x_1}{\sqrt{1-\rho^2}}\right)+\Phi(x_2),\Phi(x_2)\Phi\left(\frac{x_1-\rho x_2}{\sqrt{1-\rho^2}}\right)+\Phi(x_1)\right].$$

PROOF. Consider

$$H_\rho(x_1,x_2)=\Phi_\rho(x_1,x_2)-\Phi(x_1)\Phi\left(\frac{x_2-\rho x_1}{\sqrt{1-\rho^2}}\right).$$

It is easily seen that $H_\rho(x_1,x_2)$ converges to 0 for $x_1\to-\infty$ and to $\Phi(x_2)$ for $x_1\to\infty$. Moreover,

$$\frac{\partial}{\partial x_1}H_\rho(x_1,x_2)=\frac{\rho}{\sqrt{1-\rho^2}}\varphi\left(\frac{x_2-\rho x_1}{\sqrt{1-\rho^2}}\right)\Phi(x_1)$$

is positive. Hence $H_\rho(x_1,x_2)$ is strictly increasing in x_1 and ranges between 0 and $\Phi(x_2)$. Exchanging the roles of x_1 and x_2 establishes the assertion. ■

We shall abbreviate the corresponding joint survival function of Φ_ρ by

$$S_\rho(x_1,x_2)=\mathbb{P}(X_1>x_1,X_2>x_2),$$

where (X_1,X_2) is distributed according to Φ_ρ. By an argument of Balkema (Resnick, 2008, Section 5.2), the bivariate Gaussian distribution Φ_ρ is max-infinitely divisible if and only if $\rho\ge0$. If $\rho\in(0,1)$, the associated exponent measure $\Lambda^{(\rho)}$ has a positive continuous density on $\mathbb{R}^2=(-\infty,\infty)^2$ as given by

$$\lambda^{(\rho)}(x_1,x_2)=\frac{\partial}{\partial x_1}\frac{\partial}{\partial x_2}\log\Phi_\rho(x_1,x_2)=\frac{\varphi_\rho(x_1,x_2)}{\Phi_\rho(x_1,x_2)}\kappa^{(\rho)}(x_1,x_2),$$

where

$$\kappa^{(\rho)}(x_1,x_2)=1-\frac{\varphi(x_1)\varphi(x_2)\Phi\left(\frac{x_1-\rho x_2}{\sqrt{1-\rho^2}}\right)\Phi\left(\frac{x_2-\rho x_1}{\sqrt{1-\rho^2}}\right)}{\varphi_\rho(x_1,x_2)\Phi_\rho(x_1,x_2)}.$$

Evidently, $\kappa^{(\rho)}\le1$. By the max-infinitely divisibility property of Φ_ρ, it is clear that $\kappa^{(\rho)}\ge0$. It can be further uniformly bounded as follows.

LEMMA D.3. *Let* $\rho \in (0,1)$. *For* $\max(x_1, x_2) \geq 0$ *we have*

$$0 < 1 - \sqrt{1-\rho^2} \leq \kappa^{(\rho)}(x_1, x_2). \tag{52}$$

Moreover, there exists a constant $C_\rho > 0$ *such that for* $u > 0$

$$1 - C_\rho u^{-1} \leq \inf_{x_1, x_2 \geq u} \kappa^{(\rho)}(x_1, x_2). \tag{53}$$

PROOF. First note that for all $x_1, x_2 \in \mathbb{R}$

$$\frac{\varphi(x_1)\varphi(x_2)}{\varphi_\rho(x_1, x_2)} = \frac{\sqrt{1-\rho^2}\varphi(x_1)}{\varphi\left(\frac{x_1 - \rho x_2}{\sqrt{1-\rho^2}}\right)}$$

and by Lemma D.2

$$\frac{\Phi\left(\frac{x_1-\rho x_2}{\sqrt{1-\rho^2}}\right)\Phi\left(\frac{x_2-\rho x_1}{\sqrt{1-\rho^2}}\right)}{\Phi_\rho(x_1, x_2)} \leq \frac{1}{\Phi(x_1)}\Phi\left(\frac{x_1 - \rho x_2}{\sqrt{1-\rho^2}}\right),$$

which gives (by repeating this argument with reversed roles for x_1 and x_2)

$$\begin{aligned} 1 - \kappa^{(\rho)}(x_1, x_2) &\leq \sqrt{1-\rho^2} \min\left[\frac{\varphi(x_1)}{\Phi(x_1)} \frac{\Phi\left(\frac{x_1-\rho x_2}{\sqrt{1-\rho^2}}\right)}{\varphi\left(\frac{x_1-\rho x_2}{\sqrt{1-\rho^2}}\right)}, \frac{\varphi(x_2)}{\Phi(x_2)} \frac{\Phi\left(\frac{x_2-\rho x_1}{\sqrt{1-\rho^2}}\right)}{\varphi\left(\frac{x_2-\rho x_1}{\sqrt{1-\rho^2}}\right)}\right] \\ &= \sqrt{1-\rho^2} \min\left[H\left(\frac{x_1 - \rho x_2}{\sqrt{1-\rho^2}}\right)/H(x_1), H\left(\frac{x_2 - \rho x_1}{\sqrt{1-\rho^2}}\right)/H(x_2)\right], \end{aligned} \tag{54}$$

where $H = \Phi/\varphi$.

In order to establish (52), it suffices to show that the last minimum is not larger than 1 for $x_1 \geq 0$. We distinguish three cases.

Case 1: $0 \leq x_1 \leq x_2$. First note that $\rho \in (0,1)$ implies $q = (1 - \sqrt{1-\rho^2})/\rho \in (0,1)$ and hence $qx_1 \leq x_2$, which implies

$$\frac{x_1 - \rho x_2}{\sqrt{1-\rho^2}} \leq x_1.$$

By the monotonicity of H (see Lemma D.1) the assertion follows.

Case 2a: $0 \leq x_2 \leq x_1$. We can repeat the argument of Case 1 with reversed roles for x_1 and x_2.

Case 2b: $x_2 \leq x_1$, $x_1 \geq 0$, $x_2 = -y < 0$. In this situation, we can rewrite

$$H\left(\frac{x_2 - \rho x_1}{\sqrt{1-\rho^2}}\right)/H(x_2) = \widetilde{H}(y)/\widetilde{H}\left(\frac{y + \rho x_1}{\sqrt{1-\rho^2}}\right),$$

where $\widetilde{H}(x) = 1/H(-x) = \varphi(x)/\overline{\Phi}(x)$ is also monotonously increasing (see Lemma D.1). Since $y \leq (y + \rho x_1)/\sqrt{1-\rho^2}$, we can bound the right-hand side by 1 and the assertion follows. Taken together, this shows (52).

In order to establish the second assertion (53), without loss of generality, assume $x_1 \geq x_2 \geq u > 0$ and recall from (54) that

$$\frac{\Phi(0)}{\sqrt{1-\rho^2}}\left(1 - \kappa^{(\rho)}(x_1, x_2)\right) \leq \frac{\Phi(0)}{\Phi(x_2)}\varphi(x_2)\frac{\Phi\left(\frac{x_2-\rho x_1}{\sqrt{1-\rho^2}}\right)}{\varphi\left(\frac{x_2-\rho x_1}{\sqrt{1-\rho^2}}\right)} \leq \varphi(x_2)\frac{\Phi\left(\frac{x_2-\rho x_1}{\sqrt{1-\rho^2}}\right)}{\varphi\left(\frac{x_2-\rho x_1}{\sqrt{1-\rho^2}}\right)}. \tag{55}$$

We distinguish three cases according to the value of $q = x_2/x_1 \in (0,1]$:
Case 1: $q \in (0, \rho/2]$. Then $-(x_2 - \rho x_1) = (\rho - q)x_1 > \rho/2 \cdot u$ and the right-hand side of (55) is bounded by $2\varphi(0)\sqrt{1-\rho^2}/\rho \cdot u^{-1}$ according to (50).
Case 2: $q \in (\rho/2, \rho)$. Then $x_2 - \rho x_1 = -(\rho - q)x_1 < 0$ and $x_2 = qx_1 > \rho/2 \cdot u$. With Lemma D.1, we obtain that the right-hand side of (55) is bounded by $\Phi(0)/\varphi(0)\varphi(\rho/2 \cdot u)$.
Case 3: $q \in [\rho, 1]$. In this situation $x_2 - \rho x_1 = (q - \rho)x_1$ is positive. We also recall $\rho/(1 - \sqrt{1-\rho^2}) > 1 \geq q$, which implies

$$q > \frac{q - \rho}{\sqrt{1-\rho^2}} > 0.$$

Hence, we can bound the right-hand side of (55) by $\Phi(0)$ multiplied with

$$\sup_{q\in[\rho,1]} \frac{\varphi(qx_1)}{\varphi\Big(\frac{(q-\rho)x_1}{\sqrt{1-\rho^2}}\Big)} \leq \exp\left\{ -\frac{1}{2} \inf_{q\in[\rho,1]} \left[q^2 - \left(\frac{q-\rho}{\sqrt{1-\rho^2}}\right)^2\right] x_1^2 \right\} \leq \exp(-\rho^2 u^2/2).$$

Taken together, all three cases establish also the assertion (53). ■

LEMMA D.4. *Let $\rho \in (0,1)$. The joint survival functions $\Lambda^{(\rho)}$ and S_ρ are asymptotically equivalent in the sense that*

$$\lim_{u\to\infty} \frac{\Lambda^{(\rho)}((u,\infty)^2)}{S_\rho(u,u)} = 1.$$

PROOF. By definition, we have

$$\Lambda^{(\rho)}((u,\infty)^2) = 2(-\log \Phi(u)) - (-\log \Phi_\rho(u,u))$$

and

$$S_\rho(u,u) = 2(1 - \Phi(u)) - (1 - \Phi_\rho(u,u)).$$

Applying L'Hopital's rule yields

$$\begin{aligned}
\lim_{u\to\infty} \frac{\Lambda^{(\rho)}((u,\infty)^2)}{S_\rho(u,u)} &= \lim_{u\to\infty} \frac{2\varphi(u)/\Phi(u) - 2\varphi(u)\Phi(su)/\Phi_\rho(u,u)}{2\varphi(u) - 2\varphi(u)\Phi(su)} \\
&= \lim_{u\to\infty} \frac{1}{\Phi_\rho(u,u)\Phi(u)} \cdot \lim_{u\to\infty} \frac{\Phi_\rho(u,u) - \Phi(su)\Phi(u)}{1 - \Phi(su)},
\end{aligned}$$

where $s = (1-\rho)/\sqrt{1-\rho^2} = \sqrt{(1-\rho)/(1+\rho)} \in (0,1)$. As the involved cdfs converge to 1, it remains to be seen that the second limit equals 1. Indeed,

$$\begin{aligned}
\lim_{u\to\infty} \frac{\Phi_\rho(u,u) - \Phi(su)\Phi(u)}{1 - \Phi(su)} &= \lim_{u\to\infty} \frac{2\varphi(u)\Phi(su) - s\varphi(su)\Phi(u) - \varphi(u)\Phi(su)}{-s\varphi(su)} \\
&= \lim_{u\to\infty} \Phi(u) - \lim_{u\to\infty} \frac{\varphi(u)}{\varphi(su)} \cdot \lim_{u\to\infty} \frac{\Phi(su)}{s} = 1 - 0/s = 1,
\end{aligned}$$

hence the assertion. ■

In what follows we will need a few simple facts about regularly varying functions. An eventually positive measurable function that is defined in a neighborhood of $\infty \in \mathbb{R}$, is *regularly varying* with index $\alpha \in \mathbb{R}$ (we write $f \in \mathrm{RV}_\alpha$) if $f(st)/f(t) \to s^{-\alpha}$ for $t \to \infty$. The case $\alpha = 0$ is a special case, f is then called *slowly varying*. Trivially, $\log(t)$, $1/\log(t)$ or any measurable function that converges to a positive constant is slowly varying. Moreover, $f \in \mathrm{RV}_\alpha$ if and only if $f(t)t^{-\alpha}$ is slowly varying. Some well-known closure properties are as follows (e.g., de Haan and Ferreira, 2006, Appendix B).

LEMMA D.5. *If $f_1 \in \mathrm{RV}_{\alpha_1}$, $f_2 \in \mathrm{RV}_{\alpha_2}$, then $f_1 + f_2 \in \mathrm{RV}_{\max(\alpha_1,\alpha_2)}$ and $f_1 f_2 \in \mathrm{RV}_{\alpha_1+\alpha_2}$. If in addition $f_2(t) \to \infty$ for $t \to \infty$, then the composition satisfies $f_1 \circ f_2 \in \mathrm{RV}_{\alpha_1\alpha_2}$.*

We will need to be aware of the following consequence for $F(u) = \Phi(u)$ in this section.

LEMMA D.6. *Let F be a (univariate) montonously increasing continuous cumulative distribution function with upper endpoint ∞ and survival function $\overline{F} = 1 - F$. Let $\Lambda(u) = -\log F(u)$. Let $a(u) = \overline{F}(u)^{-1}$ and $b(u) = \Lambda(u)^{-1}$. Then $f \circ a^{-1} \in \mathrm{RV}_\alpha$ if and only if $f \circ b^{-1} \in \mathrm{RV}_\alpha$. In particular, $f \circ a^{-1}$ is slowly varying if and only if $f \circ b^{-1}$ is slowly varying.*

PROOF. Since $b(u) \to \infty$ and $a(u)/b(u) \to 1$ for $u \to \infty$, we have $a \circ b^{-1}(t)/t \to 1$ for $t \to \infty$. This implies $a \circ b^{-1} \in \mathrm{RV}_1$ and $a \circ b^{-1}(t) \to \infty$ for $t \to \infty$. So, if $f \circ a^{-1} \in \mathrm{RV}_\alpha$, so is $f \circ b^{-1} = (f \circ a^{-1}) \circ (a \circ b^{-1})$ by Lemma D.5. The converse follows analogously. ∎

Coming back to the Gaussian joint survival functions, it is well-known that for $\eta_\rho = (1 + \rho)/2$

$$S_\rho(u,u) = \ell_\rho(\overline{\Phi}(u)^{-1})\overline{\Phi}(u)^{1/\eta_\rho} \tag{56}$$

for a slowly varying function ℓ_ρ as $u \to \infty$ (Ledford and Tawn, 1996). In other words, the distribution Φ_ρ exhibits *asymptotic independence*; see Section 8.2.

Due to the asymptotic equivalences from (51) and Lemma D.4 and, in addition, Lemma D.6, this relation carries over to the asymptotically equivalent counterparts, that is,

$$\Lambda^{(\rho)}((u,\infty)^2) = \widetilde{\ell}_\rho(\Lambda(u)^{-1})\Lambda(u)^{1/\eta_\rho} \tag{57}$$

for a slowly varying function $\widetilde{\ell}_\rho$. In particular,

$$\chi_\rho = \lim_{u\to\infty} \frac{\Lambda^{(\rho)}((u,\infty)^2)}{\Lambda(u)} = \lim_{u\to\infty} \frac{S_\rho(u,u)}{\overline{\Phi}(u)} = 0. \tag{58}$$

Moreover, moving from a multivariate Gaussian distribution F on $\mathbb{R}^d = (-\infty,\infty)^d$ to a log-Gaussian distribution $F \circ \log$ on $(0,\infty)^d$ only requires a (componentwise) log-transformation, and the same applies to the corresponding exponent measures; that is, $\Lambda_1 \circ \log$ is the exponent measure of $\Phi \circ \log$ and $\Lambda^{(\rho)} \circ \log$ is the exponent measure of $\Phi_\rho \circ \log$. Hence, the relations (51), (57) and (58) all carry over to the log-Gaussian variant on $(0,\infty)$ and $(0,\infty)^2$ by realizing that one can replace u by $\log(u)$ in all situations. For notational convenience, we shall remain on the Gaussian scale in what follows, even though the construction could also be carried out on the log-Gaussian scale. In essence, this corresponds to replacing 0 by $-\infty$ as the absorbing state, whilst the framework of conditional independence carries over with only minor modifications; see also Remark 3.2.

Now let $a, b \in (0,1)$ and consider a trivariate exponent measure Λ on $\mathbb{R}^3$ with density

$$\lambda(x) = \frac{\lambda^{(a)}(x_1,x_2)\lambda^{(b)}(x_2,x_3)}{\lambda_1(x_2)}, \qquad x \in \mathbb{R}^3,$$

similarly to the example of Section 5.4 with $p_{12} = p_{23} = 1$. This ensures that the marginal measures of Λ are given by

$$\Lambda_{12} = \Lambda^{(a)} \qquad \text{and} \qquad \Lambda_{23} = \Lambda^{(b)},$$

whilst the transformation of Λ to log-Gaussian scale satisfies the conditional independence relation

$$\{1\} \perp \{3\} \mid \{2\} \; [\Lambda \circ \log].$$

From the above considerations, we know that

$$\chi_{12} = \chi_a = 0, \quad \eta_{12} = \frac{1+a}{2}, \quad \chi_{23} = \chi_b = 0 \quad \text{and} \quad \eta_{23} = \eta_b = \frac{1+b}{2}. \tag{59}$$

In order to derive the corresponding values for χ_{13} and η_{13}, we need to investigate the asymptotic behavior of

$$\Lambda_{13}((u,\infty)^2) = \int_u^\infty \int_{-\infty}^\infty \int_u^\infty \lambda(x)\mathrm{d}x_1\mathrm{d}x_2\mathrm{d}x_3 \tag{60}$$

in relation to $\Lambda(u)$ as $u \uparrow \infty$.

Note that Λ_{13} is different from the bivariate exponent measure $\Lambda^{(ab)}$, which would have arisen as marginal exponent measure of (X_1, X_3) of a trivariate Gaussian distribution (X_1, X_2, X_3) with $(X_1, X_2) \sim \Phi_a$, $(X_2, X_3) \sim \Phi_b$ and $X_1 \perp\!\!\!\perp X_3 \mid X_2$ (it is easily seen that the inverse Q of the correlation matrix Σ with entries $\rho_{12} = a$, $\rho_{23} = b$, $\rho_{13} = ab$ satisfies $Q_{13} = 0$; an easy way to verify that $\Lambda_{13} \neq \Lambda^{(ab)}$ is by considering the trivariate zero mean Gaussian distribution with covariance matrix Σ and by observing that the value at $0 \in \mathbb{R}^3$ of a hypothetical density of its exponent measure, if it existed, deviates from $\lambda(0)$).

However, we bear in mind that

$$\varphi_{ab}(x_1, x_3) = \int_{-\infty}^\infty \widetilde{\varphi}(x)\mathrm{d}x_2 \qquad \text{with} \qquad \widetilde{\varphi}(x) = \frac{\varphi_a(x_1,x_2)\varphi_b(x_2,x_3)}{\varphi(x_2)},$$

and that

$$\lambda^{(ab)}(x_1,x_3) = \frac{\varphi_{ab}(x_1,x_3)}{\Phi_{ab}(x_1,x_3)}\kappa^{(ab)}(x_1,x_3) = \int_{-\infty}^\infty \widetilde{\varphi}(x)\,\mathrm{d}x_2\,\frac{\kappa^{(ab)}(x_1,x_3)}{\Phi_{ab}(x_1,x_3)}.$$

The density of Λ_{13} is instead given by

$$\begin{aligned}
\lambda_{13}(x_1,x_3) &= \int_{-\infty}^\infty \lambda(x)\mathrm{d}x_2 \\
&= \int_{-\infty}^\infty \frac{\lambda^{(a)}(x_1,x_2)\lambda^{(b)}(x_2,x_3)}{\lambda_1(x_2)}\mathrm{d}x_2 \\
&= \int_{-\infty}^\infty \widetilde{\varphi}(x)\,\frac{\Phi(x_2)}{\Phi_a(x_1,x_2)\Phi_b(x_2,x_3)}\,\kappa^{(a)}(x_1,x_2)\kappa^{(b)}(x_2,x_3)\mathrm{d}x_2.
\end{aligned}$$

Although, $\Lambda_{13} \neq \Lambda^{(ab)}$, these measures do not deviate too significantly from each other in their joint upper tails. In order to see this, all we need to do is to control the behavior of the additional multiplicative terms in the integrand above that appear in addition to the term $\widetilde{\varphi}(x)$.

LEMMA D.7. *There exist $C > 0$ such that for $x_1, x_3 \geq 0$*

$$\frac{\Phi(x_2)}{\Phi_a(x_1,x_2)\Phi_b(x_2,x_3)} \leq C\frac{1}{\Phi(x_2)}.$$

PROOF. For $x_1, x_3 \geq 0$, we have

$$\frac{\Phi(x_2)}{\Phi_a(x_1,x_2)\Phi_b(x_2,x_3)} \leq \frac{\Phi(x_2)}{\Phi_a(0,x_2)\Phi_b(x_2,0)}.$$

If $x_2 \geq 0$ the latter is bounded by $(\Phi_a(0,0)\Phi_b(0,0))^{-1}$. If $x_2 < 0$ we can apply Lemma D.2 to see that we may instead use the bound

$$\frac{\Phi(x_2)}{\Phi(-ax_2/\sqrt{1-a^2})\Phi(x_2)\Phi(-bx_2/\sqrt{1-b^2})\Phi(x_2)} \leq \frac{1}{\Phi(0)^2}\frac{1}{\Phi(x_2)}.$$

Hence, the assertion is valid for $C = 1/\min(\Phi(0)^2, \Phi_a(0,0)\Phi_b(0,0))$. ■

LEMMA D.8. *For $\beta \in (0,1)$ and $\delta > 0$*

$$\lim_{u\to\infty} \frac{1}{\overline{\Phi}(u)^{2-\delta}} \int_u^\infty \int_{-\infty}^{u^\beta} \int_u^\infty \frac{\widetilde{\varphi}(x)}{\Phi(x_2)} \,\mathrm{d}x_1\mathrm{d}x_2\mathrm{d}x_3 = 0.$$

PROOF. We can rewrite the integrand as

$$\frac{\widetilde{\varphi}(x)}{\Phi(x_2)} = C_{ab}\frac{\varphi(x_2)}{\Phi(x_2)}\varphi\left(\frac{x_1 - ax_2}{\sqrt{1-a^2}}\right)\varphi\left(\frac{x_3 - bx_2}{\sqrt{1-b^2}}\right)$$

for a constant $C_{ab} > 0$. For sufficiently large u, and $x_1, x_3 \geq u$, $x_2 \leq u^\beta$ it is clear that $x_1 - ax_2 \geq u - au^\beta > 0$ and $x_3 - bx_2 > u - bu^\beta > 0$. Therefore, and since $\sqrt{1-\rho^2} < 1$ for $\rho \in \{a, b\}$, we obtain together with the left-hand side of (50)

$$\frac{\widetilde{\varphi}(x)}{\Phi(x_2)} \leq C_{ab}\max(-2x_2, c)\varphi(x_1 - ax_2)\varphi(x_3 - bx_2)$$

for a constant $c > 0$. Moreover

$$\begin{aligned}\frac{\varphi(x_1 - ax_2)}{\varphi(x_1 - au^\beta)} &= \sqrt{2\pi}\exp(-ax_1(u^\beta - x_2))\exp\left(\frac{1}{2}a^2u^{2\beta}\right)\varphi(ax_2)\\ &\leq \sqrt{2\pi}\exp\left(\frac{1}{2}a^2u^{2\beta}\right)\varphi(ax_2)\end{aligned}$$

and analogously for the terms involving x_3. Thus,

$$\frac{\widetilde{\varphi}(x)}{\Phi(x_2)} \leq C'\exp\left(\frac{1}{2}(a^2+b^2)u^{2\beta}\right)\max(-2x_2, c)\varphi\big(\sqrt{a^2+b^2}x_2\big)\varphi(x_1 - au^\beta)\varphi(x_3 - bu^\beta)$$

for a constant $C' > 0$. Now, the function $\max(-2x_2, c)\varphi\big(\sqrt{a^2+b^2}x_2\big)$ is integrable over $\mathbb{R} = (-\infty, \infty)$ (and not only over $(-\infty, u^\beta)$). So, we can find a constant $C'' > 0$, such that

$$\int_u^\infty \int_{-\infty}^{u^\beta} \int_u^\infty \frac{\widetilde{\varphi}(x)}{\Phi(x_2)} \,\mathrm{d}x_1\mathrm{d}x_2\mathrm{d}x_3 \leq C''\exp\left(\frac{1}{2}(a^2+b^2)u^{2\beta}\right)\overline{\Phi}(u - au^\beta)\overline{\Phi}(u - bu^\beta).$$

Finally, using (50) again, we find that the right-hand side of the last expression divided by $\overline{\Phi}(u)^{2-\delta}$ is (up to a multiplicative constant) asymptotically equivalent to

$$\begin{aligned}&p(u)\exp\left\{\frac{1}{2}\left[(a^2+b^2)u^{2\beta} - [u - au^\beta]^2 - [u - bu^\beta]^2 + (2-\delta)u^2\right]\right\}\\ &= p(u)\exp\left(-\frac{1}{2}\left[\delta u^2 - 2(a+b)u^{1+\beta}\right]\right) \to 0\end{aligned}$$

as $u \to \infty$, where $p(u)$ is absorbing the ratio of polynomial terms arising from the approximation via (50). ■

We can now give the proof for the main result of this section.

THEOREM D.9. *The measure Λ_{13} from (60) satisfies*

$$\Lambda_{13}((u,\infty)^2) = \ell(\overline{\Phi}(u)^{-1})\overline{\Phi}(u)^{1/\eta_{ab}}.$$

for a slowly varying function ℓ and $\eta_{ab} = (1+ab)/2$.

PROOF. Let $\beta \in (0,1)$. We consider the following decompositions of integrals $\Lambda_{13}((u,\infty)^2) = A_\lambda(u) + B_\lambda(u)$ and $\Lambda^{(ab)}((u,\infty)^2) = A_{\widetilde{\varphi}}(u) + B_{\widetilde{\varphi}}(u)$, where

$$A_f(u) = \int_u^\infty \int_{-\infty}^{u^\beta} \int_u^\infty f(x)\,\mathrm{d}x_1\mathrm{d}x_2\mathrm{d}x_3 \quad \text{and} \quad B_f(u) = \int_u^\infty \int_{u^\beta}^\infty \int_u^\infty f(x)\,\mathrm{d}x_1\mathrm{d}x_2\mathrm{d}x_3.$$

Due to Lemma D.3 we have

$$(1 - C_a u^{-\beta})(1 - C_b u^{-\beta}) B_{\widetilde{\varphi}}(u) \le B_\lambda(u) \le \frac{1}{\Phi_a(u, u^\beta)\Phi_b(u^\beta, u)} B_{\widetilde{\varphi}}(u)$$

and therefore $B_\lambda(u) = (1 + e(u))B_{\widetilde{\varphi}}(u)$ with $e(u) \to 0$ for $u \to \infty$. We note that therefore the mapping $u \mapsto 1 + e(u)$ is slowly varying, as it converges to a positive constant.
Denoting $\eta = \eta_{ab} = (1+ab)/2$, we have

$$\frac{\Lambda_{13}((u,\infty)^2)}{\overline{\Phi}(u)^{1/\eta}} = \frac{A_\lambda(u)}{\overline{\Phi}(u)^{1/\eta}} + (1 + e(u))\left[\frac{\Lambda^{(ab)}((u,\infty)^2)}{\overline{\Phi}(u)^{1/\eta}} - \frac{A_{\widetilde{\varphi}}(u)}{\overline{\Phi}(u)^{1/\eta}}\right]$$

and we need to show that

$$\frac{\Lambda_{13}((h^{-1}(v),\infty)^2)}{\overline{\Phi}(h^{-1}(v))^{1/\eta}} \qquad \text{with} \qquad h(u) = \frac{1}{\overline{\Phi}(u)}$$

is slowly varying as a function of v. To start with, we recall from the knowledge about bivariate Gaussian survival functions from above that

$$\frac{\Lambda^{(ab)}((h^{-1}(v),\infty)^2)}{\overline{\Phi}(h^{-1}(v))^{1/\eta}}$$

is slowly varying, which is a reformulation of (56) and (57) for $\rho = ab$. Moreover, the function h^{-1} is slowly varying, because Φ is in the Gumbel domain of attraction and h^{-1} is monotone (see Resnick, 2008, Prop. 0.10 and Exercise 0.4.3.1.), and in addition $h^{-1}(v) \to \infty$ for $v \to \infty$. Hence $(1 + e(h^{-1}(v)))$ is slowly varying.
What remains to be seen to complete the proof, is that adding (or subtracting) the term $A_f(h^{-1}(v))/\overline{\Phi}(h^{-1}(v))^{1/\eta}$ to (or from) a slowly varying function for $f \in \{\lambda, \widetilde{\varphi}\}$, preserves slow variation. Indeed, with $C > 0$ as in Lemma D.7 we have

$$\max(A_\lambda(u), A_{\widetilde{\varphi}}(u)) \le \max(1, C) \int_u^\infty \int_{-\infty}^{u^\beta} \int_u^\infty \frac{\widetilde{\varphi}(x)}{\Phi(x_2)}\,\mathrm{d}x_1\mathrm{d}x_2\mathrm{d}x_3.$$

According to Lemma D.8 and since $\eta > 1/2$, there exists (in both cases, $f = \lambda$ or $f = \widetilde{\varphi}$) some $\alpha > 0$ such that

$$\lim_{v\to\infty} \frac{A_f(h^{-1}(v))}{\overline{\Phi}(h^{-1}(v))^{1/\eta}} \cdot v^\alpha = \lim_{u\to\infty} \frac{A_f(u)}{\overline{\Phi}(u)^{1/\eta+\alpha}} = 0.$$

It is easily checked that adding such a function g (with $g(v) \cdot v^\alpha \to 0$) to a slowly varying function will not alter the slow variation property.
Taken together, we have seen that

$$\Lambda_{13}((u,\infty)^2) = \ell(\overline{\Phi}(u)^{-1})\overline{\Phi}(u)^{1/\eta}$$

for a slowly varying function ℓ. ∎

As above, we may also work in Theorem D.9 with $\Lambda(u) = -\log \Phi(u)$ instead of $\overline{\Phi}(u)$ (see (51) and Lemma D.6), that is,

$$\Lambda_{13}((u,\infty)^2) = \widetilde{\ell}(\Lambda(u)^{-1})\Lambda(u)^{1/\eta_{ab}}$$

for a slowly varying function $\widetilde{\ell}$. Stated in this form, the asymptotic behavior of the joint survival function of Λ_{13} resembles (57) with correlation $\rho = ab$ except that the slowly varying function $\widetilde{\ell}$ does not coincide with $\widetilde{\ell}_{ab}$. Nevertheless, we may conclude

$$\chi_{13} = \chi_{ab} = 0 \quad \text{and} \quad \eta_{13} = \eta_{ab} = \frac{1+ab}{2},$$

complementing our knowledge from (59). Hence, although the marginal measure Λ_{13} does not coincide with the bivariate exponent measure $\Lambda^{(ab)}$, they share the same asymptotic behavior on their joint survival sets $(u,\infty)^2$ as $u \uparrow \infty$. Collectively, this establishes Proposition 8.11.

APPENDIX E: BEYOND SEMI-GRAPHOIDS — WHEN WE HAVE A GRAPHOID

Under the explosiveness condition (E1), the conditional independence with respect to Λ satisfies the four semi-graphoid properties, (L1)–(L4), cf. Section 3.3 and Theorem 5.3. It would even be a *graphoid* if in addition the following property (L5) was satisfied for all disjoint subsets A, B, C, D of V:

(L5) If $A \perp B \,|\, C \cup D\,[\Lambda]$ and $A \perp C \,|\, B \cup D\,[\Lambda]$, then $A \perp B \cup C \,|\, D\,[\Lambda]$ (Intersection)

Importantly, the presence of a graphoid-type conditional independence guarantees the equivalence of global, local and pairwise Markov properties, be it directed or undirected; see Lauritzen (1996, Theorem 3.7 and page 52) or Pearl and Paz (2022). Hence the question if sufficiently rich conditions can be found, under which the Λ-based conditional independence defined here is in fact a graphoid, and not only a semi-graphoid; see also the discussion contribution of Steffen Lauritzen in Engelke and Hitz (2020). It can be easily seen that (L5) is violated in general.

EXAMPLE E.1 (Violation of (L5) for a measure on a ray). Let η be a measure on the positive real line $(0,\infty)$ that explodes at the 0 and is finite on sets bounded away from 0, e.g., $\eta((t,\infty)) = t^{-1}$. Then the measure

$$\Lambda(E) = \eta\big(\{t > 0 : t(1,1,1) \in E\}\big).$$

on $\mathbb{R}^3$, which is concentrated on the ray through $(1,1,1) \in \mathbb{R}^3$, satisfies our basic requirements (B) and (E1), and thus (L1)–(L4) hold. Let $Y \sim \mathbb{P}_R$ for an admissible $R \in \mathcal{R}(\Lambda)$. Then the law of $(Y_1,Y_2)|Y_3$ is a point mass at (Y_3,Y_3), which is a product measure. Hence, $\{1\} \perp \{2\} \,|\, \{3\}[\Lambda]$, and by symmetry of the argument, $\{1\} \perp \{3\} \,|\, \{2\}[\Lambda]$. If (L5) were true, we would be able to conclude $\{1\} \perp \{2,3\}[\Lambda]$ and thereby $\Lambda(y_1 \neq 0, (y_2,y_3) \neq (0,0)) = 0$; see Prop. 5.1. This contradicts $\Lambda(y_1 \neq 0, (y_2,y_3) \neq (0,0)) = \eta((0,\infty)) = \infty$.

In the classical probabilistic case the existence of a positive continuous density with respect to a product measure is a well-known condition to guarantee (L5) (Lauritzen, 1996, discussion of (3.10)). This translates as follows to our situation.

COROLLARY E.2. *Assume (E1) and let $\mu = \bigotimes_{v \in V} \mu_v$ be a product measure. Assume that each marginal measure Λ_K, $\emptyset \neq K \subset V$ (including $\Lambda = \Lambda_V$) has a positive continuous density on $\mathcal{E}^K = \mathbb{R}^K \setminus \{0_K\}$ with respect to $\mu_K = \bigotimes_{v \in K} \mu_v$, then conditional independence as defined in Definition 3.1 satisfies all five graphoid properties (L1)–(L5).*

PROOF. Properties (L1)–(L4) are already established in Theorem 5.3. To prove (L5), it is sufficient to assume that the union of A, B, C, D is V; see Lemma A.2 and our assumption that each Λ_K has a positive continuous density. Then (L5) follows directly from Definition 3.1 and the respective counterpart of (L5) for probability laws. ∎

In Corollary E.2 we require each marginal Λ_K to have a positive continuous density on every $\mathcal{E}^K = \mathbb{R}^K \setminus \{0_K\}$. Instead we may consider a one-sided version (in some or all directions) in this requirement and, for instance, ask for Λ_K to have a positive continuous density on $\mathcal{E}_+^K = [0,\infty)^K \setminus \{0_K\}$ for every $\emptyset \neq K \subset V$, and the same argument shows that (L5) is then valid. If Λ is $-\alpha$-homogenous and Λ has a positive continuous Lebesgue-density on $\mathcal{E}_+ = [0,\infty)^d \setminus \{0\}$, then the marginal measures Λ_K will also have a positive continuous Lebesgue-density in their respective dimensions, so the assumptions in Corollary E.2 are met and (L5) follows in addition to (L1)–(L4). This is the case for most of the measures considered in Engelke and Hitz (2020), in particular the Hüsler–Reiss exponent measures, which have a positive continuous Lebesgue-density. Hence, they define a graphoid-type conditional independence.

However, we need to be cautious, when drawing conclusions from Corollary E.2 in our more general context. Take, for example, a measure Λ that has a positive continuous μ-density on each of the elementary sub-faces $\mathcal{E}_K$ in (14) for $\emptyset \neq K \subset V$, for μ as in (19), where each of the factors μ_v has an atom at 0. In this case, the conditional independence notion induced by Λ defines indeed a graphoid, but likely not for the intended reason. In fact, it is sufficient to assume that the measure Λ, which satisfies our basic explosiveness assumption (E1), has mass on each of the $\mathcal{E}_K$ for $\emptyset \neq K \subset V$, in order to conclude that *no* conditional independence statement $A \perp B \,|\, C\ [\Lambda]$ with non-empty A and B can hold; see Prop. 5.1 and Lemma 3.3. So there is nothing to check in order to conclude that (L5) is valid. On the other extreme end, if Λ is a zero measure, then *all* conditional independence relations are true, and so the zero measure induces trivially a graphoid relation, too.

More interestingly, let us turn our attention again to situations, when some elementary sub-faces are charged and others not. We have already considered one such case above, where we have only mass in the upper most elementary sub-face $\mathcal{E}_V$ and, consequently, it does not matter if we add an atom at 0 in each reference measure μ_v or not. Indeed, it is reassuring to see that a positive continuous Lebesgue-density for Λ and its marginal measures then ensures that Λ induces a graphoid-type conditional independence, and that this leads to the Hüsler–Reiss exponent measures from Engelke and Hitz (2020) to define a graphoid. Still, restricting the mass of Λ to only $\mathcal{E}_V$ comes along with the drawback that associated graphical models cannot be disconnected as no independence statement can hold true; see Section 8.1. More generally, we would therefore like to allow for mass on several $\mathcal{E}_K$ for non-empty $K \subset V$. The following instructive example demonstrates that caution should be exercised in such situations. Even when requiring the existence of a positive continuous density on each of the charged sub-faces $\mathcal{E}_K$, the property (L5) may be violated.

EXAMPLE E.3 (Violation of (L5) for a measure with positive continuous density on each of the charged sub-faces). Let $A = \{1\}$, $B = \{2\}$, $C = \{3\}$, $D = \{4\}$, $V = \{1,2,3,4\}$ and consider an explosive measure Λ on $\mathbb{R}^4 \setminus \{0\}$ with positive continuous μ-density $h_{14}(y_1, y_4) = \lambda(y_1, 0_2, 0_3, y_4)$ on $\mathcal{E}_{14}$ and $h_{234}(y_2, y_3, y_4) = \lambda(0_1, y_2, y_3, y_4)$ on $\mathcal{E}_{234}$ for μ as in (19) (with $d = 4$) and zero mass on any other $\mathcal{E}_K$, $K \subset \{1,2,3,4\}$. Then $\Lambda(y_4 = 0_4) = 0$, and from Table 3, which documents all densities that are relevant for the relations that appear in (L5), we can read off that $\{1\} \perp \{2\} \,|\, \{3,4\}\ [\Lambda]$ and $\{1\} \perp \{3\} \,|\, \{2,4\}\ [\Lambda]$, but not $\{1\} \perp \{2,3\} \,|\, \{4\}\ [\Lambda]$. So (L5) cannot hold.

What is interesting to note in this example is that we have still the proportionalities

$$\frac{\lambda(y)}{\lambda_{234}(y_{234})} = \frac{\lambda_{134}(y_{134})}{\lambda_{34}(y_{34})} = \frac{\lambda_{124}(y_{124})}{\lambda_{24}(y_{24})} = \begin{cases} h_{14}(y_{14})/\int h_{14}(y_{14})dy_1, & y \in \mathcal{E}_{14}, \\ 0 & y \in \mathcal{E}_{1234}. \end{cases}$$

and

$$\frac{\lambda(y)}{\lambda_{234}(y_{234})} = \frac{\lambda_{134}(y_{134})}{\lambda_{34}(y_{34})} = \frac{\lambda_{124}(y_{124})}{\lambda_{24}(y_{24})} = \begin{cases} 0, & y \in \mathcal{E}_{4}, \\ 1, & y \in \mathcal{E}_{234}. \end{cases}$$

When fixing y_{14}, the ratios still differ according to the y_{23}-component – not their precise value, but it matters if these components are (jointly) zero or not. If one wanted to conclude $\{1\} \perp \{2,3\} \,|\, \{4\} \; [\Lambda]$, one would need an argument that connects these ratios, e.g., in the form of a path through the y_{23}-components that can handle the phase transitions between different charged $\mathcal{E}_K$.

We would like to conclude the discussion with a final insight how to easily construct a measure Λ that has mass on several elementary sub-faces $\mathcal{E}_K$ for non-empty $K \subset V$, so that the induced conditional independence notion defines indeed a graphoid. In fact, one possibility to do so is to use an *independent concatenation* of graphoids. Any of the situations above, where we know that we have a graphoid, may serve as an independent component in such a construction. That is, each independent component could be, for instance, a measure as in Corollary E.2, a zero measure, a measure with mass on every elementary sub-face, or an independent concatenation thereof.

PROPOSITION E.4 (Independent concatenation of graphoids). *For $i = 1,2$ let Λ_i be a measure on $\mathbb{R}^{V_i}$ satisfying* (B) *and* (E1)*, so that the conditional independence for Λ_i defines a graphoid. Consider the measure Λ on $\mathbb{R}^{V_1 \cup V_2} = \mathbb{R}^{V_1} \times \mathbb{R}^{V_2}$ such that $\Lambda^0_{V_1} = \Lambda_{V_1} = \Lambda_1$ and $\Lambda^0_{V_2} = \Lambda_{V_2} = \Lambda_2$. Then Λ satisfies* (B) *and* (E1) *and the Λ-induced conditional independence defines a graphoid.*

The proposition follows from the following auxiliary result, which may be of independent interest. Its proof depends only on semi-graphoid properties. We added the proof here, as we did not find the following lemma explicitly in the literature.

LEMMA E.5. *Assume* (E1)*, and suppose $(A \cup B \cup C) \perp (A' \cup B' \cup C') \; [\Lambda]$. Then*

$$(A \cup A') \perp (B \cup B') \,|\, (C \cup C') \; [\Lambda] \quad \iff \quad A \perp B \,|\, C \; [\Lambda] \;\; \textit{and} \;\; A' \perp B' \,|\, C' \; [\Lambda].$$

PROOF. Assumption (E1) ensures the validity of the semi-graphoid properties (L1)–(L4), cf. Theorem 5.3. Let us abbreviate the involved statements as follows:
$(A \cup B \cup C) \perp (A' \cup B' \cup C') \; [\Lambda]$ by (Indep), $(A \cup A') \perp (B \cup B') \,|\, (C \cup C') \; [\Lambda]$ by (CI-joint), $A \perp B \,|\, C \; [\Lambda]$ by (CI-a) and $A' \perp B' \,|\, C' \; [\Lambda]$ by (CI-b).

We show first that (Indep) and (CI-joint) implies (CI-a) (and by symmetry (CI-b)): Indeed, (CI-joint) implies $A \perp B \,|\, (C \cup C') \; [\Lambda]$ by (L2), whereas (Indep) gives $A \cup C \perp C' \; [\Lambda]$ by (L2) and, subsequently, $A \perp C' \,|\, C \; [\Lambda]$ by (L3). Taken together this gives $A \perp B \cup C' \,|\, C \; [\Lambda]$ by (L4), and hence $A \perp B \,|\, C \; [\Lambda]$ by (L2).

Conversely, let us assume (CI-a) and (CI-b) in the presence of (Indep). Now, (Indep) and (L3) imply $A \perp (A' \cup B' \cup C') \,|\, (B \cup C) \; [\Lambda]$. Together with (CI-a) and (L4) this gives $A \perp B \cup (A' \cup B' \cup C') \,|\, C \; [\Lambda]$. Applying (L3) again gives

$$A \perp B \cup (A' \cup B') \,|\, C \cup C' \; [\Lambda]. \tag{61}$$

TABLE 3. *Densities from (Counter-)Example E.3 on all elementary sub-faces $\mathcal{E}_K$, cf.* (14)*, with $y_4 \neq 0$ (i.e., $4 \in K$). Since $\lambda(y)\lambda_{34}(y_{34}) = \lambda_{134}(y_{134})\lambda_{234}(y_{234})$ and $\lambda(y)\lambda_{24}(y_{24}) = \lambda_{124}(y_{124})\lambda_{234}(y_{234})$ for $y_4 \neq 0$, and $\Lambda(y_4 = 0) = 0$, we have $\{1\} \perp \{2\} \,|\, \{3,4\}\ [\Lambda]$ and $\{1\} \perp \{3\} \,|\, \{2,4\}\ [\Lambda]$. However, we may not conclude $\{1\} \perp \{2,3\} \,|\, \{4\}\ [\Lambda]$, since we do not(!) have $\lambda(y)\lambda_4(y_4) = \lambda_{14}(y_{134})\lambda_{234}(y_{234})$ for μ-almost all y with $y_4 \neq 0$; see Theorem 4.6.*

	$\lambda(y)$	$\lambda_{124}(y_{124})$	$\lambda_{134}(y_{134})$	$\lambda_{234}(y_{234})$	$\lambda_{14}(y_{14})$	$\lambda_{24}(y_{24})$	$\lambda_{34}(y_{34})$	$\lambda_4(y_4)$
$\mathcal{E}_4$	0	0	0	$\int h_{14}(y_{14})dy_1$	$\int h_{234}(y_{234})dy_{23}$	$\int h_{14}(y_{14})dy_1$	$\int h_{14}(y_{14})dy_1$	$*$
$\mathcal{E}_{14}$	$h_{14}(y_{14})$	$h_{14}(y_{14})$	$h_{14}(y_{14})$	$\int h_{14}(y_{14})dy_1$	$h_{14}(y_{14})$	$\int h_{14}(y_{14})dy_1$	$\int h_{14}(y_{14})dy_1$	$*$
$\mathcal{E}_{24}$	0	$\int h_{234}(y_{234})dy_3$	0	0	$\int h_{234}(y_{234})dy_{23}$	$\int h_{234}(y_{234})dy_3$	$\int h_{14}(y_{14})dy_1$	$*$
$\mathcal{E}_{34}$	0	0	$\int h_{234}(y_{234})dy_2$	0	$\int h_{234}(y_{234})dy_{23}$	$\int h_{14}(y_{14})dy_1$	$\int h_{234}(y_{234})dy_2$	$*$
$\mathcal{E}_{124}$	0	0	$h_{14}(y_{14})$	0	$h_{14}(y_{14})$	$\int h_{234}(y_{234})dy_3$	$\int h_{14}(y_{14})dy_1$	$*$
$\mathcal{E}_{134}$	0	$h_{14}(y_{14})$	0	0	$h_{14}(y_{14})$	$\int h_{14}(y_{14})dy_1$	$\int h_{234}(y_{234})dy_2$	$*$
$\mathcal{E}_{234}$	$h_{234}(y_{234})$	$\int h_{234}(y_{234})dy_3$	$\int h_{234}(y_{234})dy_2$	$h_{234}(y_{234})$	$\int h_{234}(y_{234})dy_{23}$	$\int h_{234}(y_{234})dy_3$	$\int h_{234}(y_{234})dy_2$	$*$
$\mathcal{E}_{1234}$	0	0	0	$h_{234}(y_{234})$	$h_{14}(y_{14})$	$\int h_{234}(y_{234})dy_3$	$\int h_{234}(y_{234})dy_2$	$*$

$* = \int h_{14}(y_{14})dy_1 + \int h_{234}(y_{234})dy_{23}$

By symmetry of the prerequisites, the analogous argument gives

(62) $$A' \perp (B \cup B') \cup A \,|\, C \cup C' \; [\Lambda].$$

With (61) and (L3) and (L1) we get

(63) $$(B \cup B') \perp A \,|\, A' \cup (C \cup C') \; [\Lambda],$$

whereas (62) and (L2) and (L1) imply

(64) $$(B \cup B') \perp A' \,|\, C \cup C' \; [\Lambda].$$

Finally, (63) and (64) and (L4) give the desired (CI-joint). ■

Acknowledgments. SE thankfully acknowledges funding from an Eccellenza grant of the Swiss National Science Foundation (Grant 186858). JI was supported by a Sapere Aude Starting Grant of the Independent Research Fund Denmark (Grant 8049-00021B). The authors would also like to thank two anonymous referees for their constructive feedback on this manuscript. In particular, this has led us to investigate the graphoid question in more depth.