Received: 21 June 2023 | Revised: 29 July 2023 | Accepted: 9 August 2023
DOI: 10.1002/sta4.613

ORIGINAL ARTICLE

WILEY

On pairwise interaction multivariate Pareto models

Michaël Lalancette

Department of Mathematics, TUM School of Computation, Information and Technology, Technical University of Munich, Munich, Bavaria, 80333, Germany

Correspondence
Michaël Lalancette, Department of Mathematics, TUM School of Computation, Information and Technology, Technical University of Munich, Munich, Bavaria 80333, Germany.
Email: michael.lalancette@tum.de

Present address
Michaël Lalancette, Parkring 13, 85748 Garching b., München, Germany.

Funding information
Natural Sciences and Engineering Research Council of Canada

The rich class of multivariate Pareto distributions forms the basis of recently introduced extremal graphical models. However, most existing literature on the topic is focused on the popular parametric family of Hüsler–Reiss distributions. It is shown that the Hüsler–Reiss family is in fact the only continuous multivariate Pareto model that exhibits the structure of a pairwise interaction model, justifying its use in many high-dimensional problems. Along the way, useful insight is obtained concerning a certain class of distributions that generalize the Hüsler–Reiss family, a result of independent interest.

KEYWORDS
Hüsler–Reiss distribution, multivariate extreme value theory, multivariate Pareto distribution, pairwise interaction model

1 | INTRODUCTION AND MAIN RESULT

Multivariate Pareto distributions play a central role in tail dependence modeling and inference as the only limit laws that can arise from multivariate threshold exceedances. They are defined as the class of possible nondegenerate weak limits of the conditional laws of $u^{-1}\mathbf{X} \mid \|\mathbf{X}\|_\infty > u$, as $u \to \infty$, for random vectors $\mathbf{X}$ with unit Pareto margins. As such, they are usually considered to perfectly describe the possible tail dependence structures of multivariate data. Originally introduced in Rootzén and Tajvidi (2006), they form the basis of multivariate peaks-over-threshold inference (Kiriliouk et al., 2019; Rootzén et al., 2018). Apart from a constraint on their support and a certain marginal standardization arising from their definition, multivariate Pareto distributions consist of all multivariate distributions $\mathbf{Y}$ satisfying the homogeneity property $\mathbb{P}(t^{-1}\mathbf{Y} \in A) = t^{-1}\mathbb{P}(\mathbf{Y} \in A)$, for $t \geq 1$ and A contained in the support of $\mathbf{Y}$. In the absolutely continuous case, which is the focus of this note, they can be exactly defined as follows (cf. Engelke & Volgushev, 2022). Given the existence of a density, the property (MP2) below is equivalent to the aforementioned homogeneity.

Definition 1. An absolutely continuous random vector $\mathbf{Y} := (Y_1, \ldots, Y_d)$ with density f has a multivariate Pareto distribution if:

(MP1) it is supported on $\mathcal{L} := [0,\infty)^d \setminus [0,1]^d$, that is, $f(\mathbf{y}) = 0$ for $\mathbf{y} \notin \mathcal{L}$,
(MP2) its density f is $(d+1)$-homogeneous, that is, $f(t\mathbf{y}) = t^{-(d+1)} f(\mathbf{y})$ for $\mathbf{y} \in \mathcal{L}$ and $t \geq 1$, and
(MP3) the marginal probabilities $P(Y_k > 1)$, $k \in 1, \ldots, d$, are equal to each other.

Remark 1. A multivariate Pareto distributed $\mathbf{Y}$ is said to have unit Pareto conditional margins since for $y > 1$, $\mathbb{P}(Y_k > y \mid Y_k > 1) = y^{-1}$. If threshold exceedances of a random vector $\mathbf{X}$ with equal, but non-Pareto marginal distributions are of interest, the limiting distribution $\mathbf{Y}$ will of course have different margins. So obtained margin-free multivariate Pareto distributions were the focus of Rootzén and Tajvidi (2006), while some authors have considered exponential or short tailed conditional margins (see, e.g., Falk & Guillou, 2008; Röttger, Engelke, & Zwiernik 2023). As models for tail dependence, those formulations are equivalent and the choice of margins is mostly one of notational convenience. In this note, we choose to stick to the historically standard choice of Pareto conditional margins.

As a relaxation, Ho and Dombry (2019) propose to drop the condition that the margins of the original data **X** be all equal, which leads to a generalized multivariate Pareto model that does not satisfy the property (MP3). The main result of this note can be extended to such a model.

The class of multivariate Pareto distributions is of course very rich. It is equivalent to the class of all extreme value copulas, or to all multivariate max-stable distributions (with fixed margins). As such, parametric subfamilies of multivariate Pareto distributions are derived from the corresponding multivariate max-stable models. One of the oldest such models is the extremal logistic distribution (Gumbel, 1960), while the most popular is that associated to the family of Hüsler–Reiss distributions (Hüsler & Reiss, 1989), hereafter termed Hüsler–Reiss distributions themselves for convenience (rather than Hüsler–Reiss multivariate Pareto).

In the examples below and in the rest of this note, let $\mathbf{1}$ denote a vector each element of which is 1, and the dimension of which will be clear from the context. For a vector $\mathbf{y} := (y_1, \ldots, y_d)$, we define $\log \mathbf{y}$ as the elementwise (natural base) logarithm $(\log y_1, \ldots, \log y_d)$. The space of symmetric $d \times d$ matrices is denoted by $\mathcal{S}^{d\times d}$, and $\mathcal{S}^{d\times d}_{\mathbf{1}} \subset \mathcal{S}^{d\times d}$ represents those matrices which have $\mathbf{1}$ in their kernel (i.e., that have zero row and column sums). Finally, $\mathcal{S}^{d\times d}_{\mathbf{1},+} \subset \mathcal{S}^{d\times d}_{\mathbf{1}}$ represents the matrices, which, in addition, are positive semi-definite with rank $d-1$.

Example 1 Extremal logistic distribution. The multivariate Pareto distributed random vector **Y** has an extremal logistic distribution with parameter $\theta \in (0,1)$ if its density f is given by

$$f(\mathbf{y}) \propto \left(\sum_{i=1}^{d} y_i^{-1/\theta} \right)^{\theta - d} \prod_{i=1}^{d} y_i^{-1/\theta - 1}, \ \mathbf{y} \in \mathcal{L}.$$

Example 2 Hüsler–Reiss distribution. The multivariate Pareto distributed random vector **Y** has a *Hüsler–Reiss* distribution if for a matrix $\Theta \in \mathcal{S}^{d\times d}_{\mathbf{1},+}$, its density f is given by

$$f(\mathbf{y}) \propto \exp\left\{ -\mu_{\mathrm{HR}}(\Theta)^\top (\log \mathbf{y}) - (\log \mathbf{y})^\top \Theta (\log \mathbf{y}) \right\}, \ \mathbf{y} \in \mathcal{L},$$

where $\mu_{\mathrm{HR}}(\Theta) := (1 + \frac{2}{d})\mathbf{1} - \frac{1}{d}\Theta\Gamma\mathbf{1}$, $\Gamma := \mathbf{1}\mathrm{diag}(\Theta^+)^\top + \mathrm{diag}(\Theta^+)\mathbf{1}^\top - 2\Theta^+$, and Θ^+ denotes the Moore–Penrose pseudoinverse of Θ. While traditionally parametrized by the variogram matrix Γ, it has recently been suggested that the Hüsler–Reiss family can be elegantly parametrized by the *Hüsler–Reiss precision matrix* Θ (Hentschel et al., 2022). By Proposition 3.4 of that paper, these two matrices are uniquely determined by each other through a homeomorphic mapping between $\mathcal{S}^{d\times d}_{\mathbf{1},+}$ and the space of symmetric, strictly conditionally negative definite matrices, to which Γ belongs. We follow Hentschel et al. (2022) and shall refer to the Hüsler–Reiss distribution with precision matrix Θ.

Hüsler–Reiss distributions enjoy a nice connection to recently introduced extremal graphical models (Engelke & Hitz, 2020). The authors of that paper declare that two components Y_i and Y_j of a multivariate Pareto random vector are conditionally independent given the other variables $(Y_k)_{k \notin \{i,j\}}$ in the extremal sense if, roughly speaking, the density of **Y** admits the factorization

$$f(\mathbf{y}) = f_i(\mathbf{y}) f_j(\mathbf{y}), \tag{1}$$

where f_i (respectively f_j) does not depend on its ith (respectively jth) argument; see their Proposition 1. While this would be equivalent to the usual notion of conditional independence were **Y** supported on a product space, this is not the case for multivariate Pareto distributions, which are supported on $\mathcal{L}$. The authors then define an extremal graphical model as a multivariate Pareto distribution satisfying a Markov property (with respect to this weaker notion of conditional independence) on a given undirected graph. For the relevant notions of graphical modeling, the reader is referred to Engelke and Hitz (2020) or to Maathuis et al. (2019).

If **Y** is Hüsler–Reiss distributed with precision matrix Θ, then Y_i and Y_j are conditionally independent given the other variables in the extremal sense of Engelke and Hitz (2020) if and only if $\Theta_{ij} = 0$. This forms an example of a *pairwise interaction model*: an exponential family of multivariate distributions where the (i,j)th element of a parameter matrix fully governs the dependence between the ith and jth variables.

Definition 2. Let $q \in \mathbb{N}$. A (curved) exponential family of probability distributions supported on a common set $\mathcal{Y} \subseteq \mathbb{R}^d$ and indexed by a parameter space $\Omega \subseteq (\mathbb{R}^q)^d \times \mathcal{S}^{d\times d}$ is called a pairwise interaction model if:

(PI1) it corresponds to a family of Lebesgue densities

$$\mathcal{F} := \left\{ \mathbf{y} \mapsto f_{\mu,\Theta}(\mathbf{y}) := \frac{1}{Z(\mu,\Theta)} \exp\left\{ -\sum_{i=1}^{d} \mu_i^\top S_i(y_i) - \sum_{i=1}^{d}\sum_{j=1}^{d} \Theta_{ij} T_i(y_i) T_j(y_j) \right\}, (\mu,\Theta) \in \Omega \right\} \tag{2}$$

for some measurable functions $S_i : \mathbb{R} \to \mathbb{R}^q$ and measurable, nonconstant functions $T_i : \mathbb{R} \to \mathbb{R}$, and

(PI2) for every pair (i,j), $i \neq j$, there exists at least one parameter $(\mu,\Theta) \in \Omega$ such that $\Theta_{ij} \neq 0$.

Remark 2. 1. Only Lebesgue densities are considered for the purpose of the present note, but Definition 2 can be readily adapted to pairwise interaction models dominated by any base (product) measure.

2. Following standard definitions of exponential families, a so-called carrying density $\prod_{i=1}^{d} h_i(y_i)$ could be included as a factor in the functions $f_{\mu,\Theta}$. However, since we do not require the family to be regular (in particular, the Hüsler–Reiss distributions form a curved exponential family), the carrying density can be absorbed in the marginal terms $S_i(y_i)$ by possibly adding a dimension to each μ_i and reducing the domain $\mathcal{Y}$. The reader is referred to standard texts such as Brown (1986) for notions of exponential families.
3. The slightly nonstandard property (PI2) is a technical requirement for the characterization of all pairwise interaction multivariate Pareto models in Lemma 1 below. It is very minor: it holds if at least one possible value of the parameter Θ has no zeros, that is, if the model allows for the simultaneous presence of all pairwise interactions. Upon inspection of the proof of Lemma 1, it could even be further relaxed. For instance, consider a graph G with an edge between i and j if and only if $\Theta_{ij} \neq 0$ for some $(\mu,\Theta) \in \Omega$. The property (PI2) states that G is fully connected, but Lemma 1 holds under the mere requirement that every edge in G is part of a cycle of odd length (e.g., a triangle).

Pairwise interaction models are ubiquitous in dependence modeling. When the common support $\mathcal{Y}$ is a product space, they are elegant examples of undirected graphical models, where the conditional independence graph contains the edge (i,j) if and only if $\Theta_{ij} \neq 0$. Each variable then typically satisfies a generalized linear model conditionally on the other variables, with the regression coefficients being extracted from the parameter matrix. Structure learning and parameter inference on a given graph structure can be efficiently carried out via (possibly penalized) likelihood, but also regression (Yang et al., 2015) or score matching (Lin et al., 2016) based methods. Gaussian and Gaussian copula models (Liu et al., 2009) as well as the continuous square root graphical model (Inouye et al., 2016) are examples of pairwise interaction graphical models used for high-dimensional dependence modeling in Euclidean settings. Klein et al. (2020) introduce pairwise interaction graphical models for multivariate angular data. Discrete analogs include the Ising model and more general log-linear interaction models (Darroch et al., 1980) as well as the discrete square root graphical model (Inouye et al., 2016).

Even in a pairwise interaction model without a product space support (such as multivariate Pareto distributions), the (i,j)th element of the parameter matrix Θ is zero if and only if the density admits a factorization as in Equation (1). Yu et al. (2022) show that score matching can be adapted to perform model selection and inference for the interaction parameter Θ in such a setting.

The Hüsler–Reiss family has been the focus of many recent papers on high-dimensional modeling and inference for tail dependence, especially in relation to extremal graphical models (see, e.g., Asenova et al., 2021; Engelke et al., 2022; Hentschel et al., 2022; Lederer & Oesting, 2023; Röttger, Coons, & Grosdos, 2023; Röttger, Engelke, & Zwiernik, 2023). In fact, it is the only pairwise interaction family of multivariate Pareto distributions that can be found in the literature, despite the fact that such families are naturally related to the density factorization property underlying extremal graphical models. A natural question is whether there exists another such family.

Theorem 1. Let $\mathcal{P}$ be a family of absolutely continuous multivariate Pareto distributions in dimension $d \geq 3$ that forms a pairwise interaction model. Then $\mathcal{P}$ is a subset of the Hüsler–Reiss family.

It has already been seen that the family of multivariate Pareto distributions is nonparametric. Moreover, the class of all pairwise interaction models is also extremely rich. If $\mathcal{Y} = \mathcal{L}$ in Definition 2, any choice of functions S_i and T_i, which, in absolute value, increase fast enough at ∞ and slowly enough at 0, gives rise to a nontrivial pairwise interaction model, with constraints on the parameter space being imposed by the choice of those functions. As will be seen in the proofs, the interplay between the additive structure of the log densities in pairwise interaction models and the homogeneity property (MP2) required of multivariate Pareto distributions means that the intersection between those two nonparametric classes is the Hüsler–Reiss family, the complexity of which is comparable to Gaussian distributions. Theorem 1 thus justifies the focus on this (relatively small) parametric model in the recent literature on graphical extremes. It would be tempting to extend some of the models in the aforementioned papers into more complicated structures while retaining the practicality of estimating pairwise interaction models, but this is in fact not possible.

In particular, in order to develop an efficient score matching algorithm, Lederer and Oesting (2023) introduce the class of functions

$$\mathcal{F}^{\text{gen HR}} := \left\{ \mathbf{y} \mapsto f_{\mu,\Theta}^{\text{gen HR}}(\mathbf{y}) := \frac{1}{Z(\mu,\Theta)} \exp\{-\mu^\top (\log \mathbf{y}) - (\log \mathbf{y})^\top \Theta (\log \mathbf{y})\} : \mu \in \mathbb{R}^d, \Theta \in \mathcal{S}_{\mathbf{1}}^{d\times d} \right\} \tag{3}$$

as surrogates of the densities of Hüsler–Reiss distributions. As the authors rightfully point out, $\mathcal{F}^{\text{gen HR}}$ strictly generalizes the class of Hüsler–Reiss densities; some of the functions therein are not even integrable on $\mathcal{L}$. While there are functions in $\mathcal{F}^{\text{gen HR}}$, which are in between, that is, integrable but not Hüsler–Reiss densities, Theorem 1 guarantees that none of them corresponds to a multivariate Pareto distribution. In fact, a corollary of the proof of Theorem 1 is a full characterization of the functions $f_{\mu,\Theta}^{\text{gen HR}} \in \mathcal{F}^{\text{gen HR}}$ based on the values of μ and Θ; see Lemma 2 below.

Theorem 1 is a direct consequence of the following two results of independent interest, which are proved in Section 2.

Lemma 1. Let $\mathcal{P}$ be a family of absolutely continuous multivariate Pareto distributions in dimension $d \geq 3$ that forms a pairwise interaction model and let $\mathcal{F}$ be the corresponding class of densities. Then $\mathcal{F} \subseteq \mathcal{F}^{\text{gen HR}}$.

Lemma 2. The functions $f_{\mu,\Theta}^{\text{gen HR}}$ in $\mathcal{F}^{\text{gen HR}}$ can be categorized as follows.

(i) If $\Theta \in \mathcal{S}_{\mathbf{1},+}^{d\times d}$ and $\mu = \mu_{\text{HR}}(\Theta)$, then $f_{\mu,\Theta}^{\text{gen HR}}$ is the density of a Hüsler–Reiss distribution.
(ii) If $\Theta \in \mathcal{S}_{\mathbf{1},+}^{d\times d}$, $\mu \neq \mu_{\text{HR}}(\Theta)$ and $\mu^\top \mathbf{1} > d$, then $f_{\mu,\Theta}^{\text{gen HR}}$ is integrable on $\mathcal{L}$ but is not the density of a multivariate Pareto distribution.
(iii) If either $\Theta \notin \mathcal{S}_{\mathbf{1},+}^{d\times d}$ or $\mu^\top \mathbf{1} \leq d$, then $f_{\mu,\Theta}^{\text{gen HR}}$ is not integrable on $\mathcal{L}$.

Remark 3. Part (i) in Lemma 2 was already stated and proved by Lederer and Oesting (2023) and is in fact the main justification for working with $\mathcal{F}^{\text{gen HR}}$. The contribution here is in giving a complete picture of the functions in this class. In particular, there are no multivariate Pareto densities in $\mathcal{F}^{\text{gen HR}}$ other than Hüsler–Reiss densities. The distributions arising in Part (ii) are of the generalized Hüsler–Reiss type of Ho and Dombry (2019).

2 | PROOFS

Throughout the proofs, for a vector $\mathbf{y} \in \mathbb{R}^d$, we generically define y_i as the ith entry of $\mathbf{y}$ and $\mathbf{y}_{\setminus i} \in \mathbb{R}^{d-1}$ as the subvector obtained by removing its ith entry. The same conventions are used for indexing the rows and columns of a matrix.

2.1 | Proof of Lemma 1

By assumption, the density class $\mathcal{F}$ is defined as in Equation (2). Using the assumed properties (MP2) and (MP3) of multivariate Pareto distributions as well as the property (PI2) of pairwise interaction models, we shall establish all the required properties that will ensure that each density in $\mathcal{F}$ is an element of $\mathcal{F}^{\text{gen HR}}$. Specifically, it will be shown that the dimension q of the marginal parameters μ_i can be taken as 1, that the functions S_i and T_i must be logarithmic, and that moreover the parameter matrix Θ must (or rather, can be assumed to) have zero row and column sums.

2.1.1 | The functions T_i must be logarithmic and Θ can be assumed to have zero row sums

The required homogeneity property (MP2) implies that for any $\mathbf{y} \in \mathcal{L}$ and $t > 1$,

$$\sum_{i=1}^{d} \mu_i^\top (S_i(ty_i) - S_i(y_i)) + \sum_{i=1}^{d}\sum_{j=1}^{d} \Theta_{ij}(T_i(ty_i)T_j(ty_j) - T_i(y_i)T_j(y_j)) = (d+1)\log t. \tag{4}$$

Since for every pair (i,j), $i \neq j$, Θ_{ij} can be nonzero, the above implies that for every such pair, any $\mathbf{y} \in \mathcal{L}$ and any $t > 1$,

$$T_i(ty_i)T_j(ty_j) - T_i(y_i)T_j(y_j) = a_{ij}(\mathbf{y}_{\setminus j}, t) + b_{ij}(\mathbf{y}_{\setminus i}, t), \tag{5}$$

for some functions a_{ij} and b_{ij} not depending on y_j and on y_i, respectively. Consider now the following auxiliary result.

Lemma 3. Let $\xi, \psi : (0,\infty) \to \mathbb{R}$ be nonconstant functions such that for any $x, y \in (0,\infty)$ and $t > 1$,

$$\xi(tx)\psi(ty) - \xi(x)\psi(y) = \alpha(x,t) + \beta(y,t),$$

for some functions α and β. Then there exist functions δ_ξ and δ_ψ such that for all positive x_1, x_2, y_1 and y_2 and all $t > 0$,

$$\xi(tx_1) - \xi(tx_2) = \delta_\xi(t)(\xi(x_1) - \xi(x_2)), \quad \psi(ty_1) - \psi(ty_2) = \delta_\psi(t)(\psi(y_1) - \psi(y_2))$$

and such that $\delta_\xi(t)$ and $\delta_\psi(t)$ are nonzero, for any $t > 0$.

Proof. Let $x_1, x_2, y \in (0,\infty)$ and $t > 1$. Applying our assumption to the points (x_1, y) and (x_2, y), find that

$$\psi(ty)(\xi(tx_1) - \xi(tx_2)) - \psi(y)(\xi(x_1) - \xi(x_2)) = \alpha(x_1,t) - \alpha(x_2,t). \tag{6}$$

At this point, note that $\xi(x_1) = \xi(x_2)$ if and only if $\xi(tx_1) = \xi(tx_2)$. Indeed, if one of these equalities holds but not the other, Equation (6) contradicts the assumption that ψ is not constant. Now supposing that $\xi(x_1) \neq \xi(x_2)$, which is possible since ξ is not constant, Equation (6) is equivalent to

$$\psi(ty) - \psi(y)\frac{\xi(x_1) - \xi(x_2)}{\xi(tx_1) - \xi(tx_2)} = \frac{\alpha(x_1,t) - \alpha(x_2,t)}{\xi(tx_1) - \xi(tx_2)}.$$

Applying the same reasoning with any other pair of points x_3, x_4 such that $\xi(x_3) \neq \xi(x_4)$ yields

$$\psi(ty) - \psi(y)\frac{\xi(x_3) - \xi(x_4)}{\xi(tx_3) - \xi(tx_4)} = \frac{\alpha(x_3,t) - \alpha(x_4,t)}{\xi(tx_3) - \xi(tx_4)}.$$

Subtracting the latter equation from the former,

$$\psi(y)\left(\frac{\xi(x_1) - \xi(x_2)}{\xi(tx_1) - \xi(tx_2)} - \frac{\xi(x_3) - \xi(x_4)}{\xi(tx_3) - \xi(tx_4)}\right) = \frac{\alpha(x_1,t) - \alpha(x_2,t)}{\xi(tx_1) - \xi(tx_2)} - \frac{\alpha(x_3,t) - \alpha(x_4,t)}{\xi(tx_3) - \xi(tx_4)},$$

which is constant in y. However, ψ was assumed nonconstant. Deduce that the difference between parentheses has to be zero, hence for any (fixed) $t > 1$, among all pairs x_1, x_2 such that $\xi(x_1) \neq \xi(x_2)$, the ratio $(\xi(tx_1) - \xi(tx_2))/(\xi(x_1) - \xi(x_2))$ is constant, say equal to $\delta_\xi(t)$.

To summarize, we have shown that

$$\xi(tx_1) - \xi(tx_2) = \delta_\xi(t)(\xi(x_1) - \xi(x_2))$$

holds for every $t > 1$ and x_1, x_2 such that $\xi(x_1) \neq \xi(x_2)$. By extension, it holds for every positive x_1 and x_2, since $\xi(x_1) = \xi(x_2)$ makes both sides vanish. Finally, the same clearly holds for $t \in (0,1]$ if $\delta_\xi(t)$ is defined as 1 for $t = 1$, and as $\delta_\xi(t^{-1})^{-1}$ for $t < 1$.

This is the desired result for the function ξ. By symmetry, the same holds for ψ. □

By Lemma 3, there exist functions δ_i, $i \in \{1,\ldots,d\}$, such that for every positive $y_i^{(1)}$, $y_i^{(2)}$ and t,

$$T_i\left(ty_i^{(1)}\right) - T_i\left(ty_i^{(2)}\right) = \delta_i(t)\left(T_i\left(y_i^{(1)}\right) - T_i\left(y_i^{(2)}\right)\right). \tag{7}$$

Now, let $y_i^{(1)}$, $y_i^{(2)}$, $y_j^{(1)}$, and $y_j^{(2)}$ be such that $T_i\left(y_i^{(1)}\right) \neq T_i\left(y_i^{(2)}\right)$ and $T_j\left(y_j^{(1)}\right) \neq T_j\left(y_j^{(2)}\right)$. Define four vectors $\mathbf{y}^{(1)},\ldots,\mathbf{y}^{(4)}$, the ith and jth entries of which are $\left(y_i^{(1)}, y_j^{(1)}\right)$, $\left(y_i^{(1)}, y_j^{(2)}\right)$, $\left(y_i^{(2)}, y_j^{(1)}\right)$, and $\left(y_i^{(2)}, y_j^{(2)}\right)$, respectively, and which agree with each other in the other $d-2$ entries.

Applying Equation (5) to those four vectors, followed by Equation (7), we have

$$
\begin{aligned}
0 &= a_{ij}\left(\mathbf{y}^{(1)}_{\setminus j},t\right)+b_{ij}\left(\mathbf{y}^{(1)}_{\setminus i},t\right)-a_{ij}\left(\mathbf{y}^{(2)}_{\setminus j},t\right)-b_{ij}\left(\mathbf{y}^{(2)}_{\setminus i},t\right)-a_{ij}\left(\mathbf{y}^{(3)}_{\setminus j},t\right)-b_{ij}\left(\mathbf{y}^{(3)}_{\setminus i},t\right)+a_{ij}\left(\mathbf{y}^{(4)}_{\setminus j},t\right)+b_{ij}\left(\mathbf{y}^{(4)}_{\setminus i},t\right) \\
&= \left(T_i\left(ty_i^{(1)}\right)-T_i\left(ty_i^{(2)}\right)\right)\times\left(T_j\left(ty_j^{(1)}\right)-T_j\left(ty_j^{(2)}\right)\right)-\left(T_i\left(y_i^{(1)}\right)-T_i\left(y_i^{(2)}\right)\right)\times\left(T_j\left(y_j^{(1)}\right)-T_j\left(y_j^{(2)}\right)\right) \\
&= \left(\delta_i(t)\delta_j(t)-1\right)\times\left(T_i\left(y_i^{(1)}\right)-T_i\left(y_i^{(2)}\right)\right)\times\left(T_j\left(y_j^{(1)}\right)-T_j\left(y_j^{(2)}\right)\right)
\end{aligned}
$$

for any $t>1$. By assumption, the last two terms in the product are nonzero. Deduce that $\delta_i(t)\delta_j(t)=1$. However, for a third index $k\notin\{i,j\}$, we may apply the same logic to find that similarly, $\delta_i(t)\delta_k(t)=\delta_j(t)\delta_k(t)=1$. This is only possible if $\delta_i(t)=\delta_j(t)=\delta_k(t)=1$ for every $t>1$, and by extension for every $t>0$, recalling that $\delta_i(t)$ is defined as 1 for $t=1$ and as $\delta_i(t^{-1})^{-1}$ for $t<1$. The same argument applies to every triple (i,j,k), so that Equation (7) can be rewritten as

$$
T_i\left(ty_i^{(1)}\right)-T_i\left(ty_i^{(2)}\right)=T_i\left(y_i^{(1)}\right)-T_i\left(y_i^{(2)}\right),
$$

which now holds for every index $i\in\{1,\ldots,d\}$ and positive $y_i^{(1)}$, $y_i^{(2)}$ and t. Equivalently, for every $t>0$, the function $y\mapsto T_i(ty)-T_i(y)$ is constant. We now apply the following.

Lemma 4. Let $\xi:(0,\infty)\to\mathbb{R}$ be a measurable function such that for any $t>0$, the function $x\mapsto\xi(tx)-\xi(x)$ is constant over $x>0$. Then $\xi(x)=c\log x+\xi(1)$, for some $c\in\mathbb{R}$.

Proof. For every positive x and t, ξ satisfies $\xi(tx)-\xi(x)=\xi(t)-\xi(1)$, that is $\xi(tx)-\xi(1)=\xi(x)+\xi(t)-2\xi(1)$. Equivalently, in terms of the function $\varphi:=\xi\circ\exp$,

$$
\varphi(u+v)-\varphi(0)=\varphi(u)+\varphi(v)-2\varphi(0),\ u,v\in\mathbb{R}.
$$

That is, $\varphi-\varphi(0)$ satisfies Cauchy's functional equation, the only measurable solutions to which are additive functions of the form $u\mapsto cu$, for some $c\in\mathbb{R}$ (see Theorem 1.1.8 of Bingham et al., 1987; Kestelman, 1947, and the references therein). Thus, $\xi(x)=\varphi(\log x)=c\log x+\xi(1)$. □

Applying Lemma 4, deduce that T_i is a logarithmic function of the form $T_i(y)=c_i\log y+T_i(1)$. Note however that the values of c_i and $T_i(1)$ can be absorbed into the parameters μ_i and Θ_{ij}, $j\in\{1,\ldots,d\}$, and into the normalizing constant, and are thus not important in characterizing the possible distributions in $\mathcal{P}$. We may therefore assume that all the functions $T_1,\ldots,T_d$ are equal to the logarithm function.

With our current formulation of the distributions in $\mathcal{P}$, the diagonal elements of Θ are not identifiable. Indeed, their value can be changed arbitrarily by adding a $(q+1)$th dimension to each μ_i and letting $S_i(y_i)_{q+1}=(\log y_i)^2$. Therefore, we shall assume without loss of generality that $\Theta_{ii}=-\sum_{j\neq i}\Theta_{ij}$, so that the row sums of Θ (and by symmetry, its column sums) are all equal to zero.

2.1.2 | The parameters μ_i can be assumed scalar and the functions S_i must be logarithmic

Replacing the functions T_i by logarithms in Equation (4) and using the assumption that the row and columns sums of Θ are zero, we find

$$
(d+1)\log t \ =\sum_{i=1}^{d}\mu_i^{\top}\left(S_i(ty_i)-S_i(y_i)\right)+\sum_{i=1}^{d}\sum_{j=1}^{d}\Theta_{ij}(\log y_i+\log y_j+\log t)\log t=\sum_{i=1}^{d}\mu_i^{\top}\left(S_i(ty_i)-S_i(y_i)\right).
$$

Similarly to what was argued about the functions T_i, deduce that for all y_i and $t>0$,

$$
\mu_i^{\top}\left(S_i(ty_i)-S_i(y_i)\right)=\mu_i^{\top}\left(S_i(t)-S_i(1)\right),
$$

which by Lemma 4 means that $\mu_i^{\top}S_i$ can be chosen to be simply a logarithm, up to scaling. Thus, we may assume without loss of generality that $q=1$ and that the real-valued functions S_i are logarithms.

To summarize, we have established that all the densities in $\mathcal{F}$ must be of the form

$$f_{\mu,\Theta}(\mathbf{y}) = \frac{1}{Z(\mu,\Theta)} \exp\left\{ -\sum_{i=1}^{d} \mu_i (\log y_i) - \sum_{i=1}^{d}\sum_{j=1}^{d} \Theta_{ij} (\log y_i)(\log y_j) \right\}$$

with Θ symmetric with zero row (and column) sums. This concludes the proof. □

2.2 | Proof of Lemma 2

As mentioned after the statement of the result, (*i*) is already obtained by Lederer and Oesting (2023), so only (*ii*) and (*iii*) shall be proved here.

Let $f_{\mu,\Theta}^{\text{gen HR}} \in \mathcal{F}^{\text{gen HR}}$. It will first be shown that for $f_{\mu,\Theta}^{\text{gen HR}}$ to be integrable on $\mathcal{L}$, it is necessary for Θ to be a Hüsler–Reiss precision matrix, that is, an element of $\mathcal{S}_{\mathbf{1},+}^{d\times d}$, and for μ to satisfy $\mu^\top \mathbf{1} > d$, establishing (*iii*). Finally, it will be shown that for such a given matrix Θ, for $f_{\mu,\Theta}^{\text{gen HR}}$ to be a multivariate Pareto density, it is necessary for μ to have the specific form $\mu_{\text{HR}}(\Theta)$ in which case $f_{\mu,\Theta}^{\text{gen HR}}$ is a Hüsler–Reiss density, thus establishing (*ii*).

2.2.1 | If $\Theta \notin \mathcal{S}_{\mathbf{1},+}^{d\times d}$, then $f_{\mu,\Theta}^{\text{gen HR}}$ is not integrable

For any index k, by the change of variable $\mathbf{x} = \log \mathbf{y}$, we find that

$$\begin{aligned} Z(\mu,\Theta) \int_{\{\mathbf{y}\in\mathcal{L}: y_k>1\}} f_{\mu,\Theta}^{\text{gen HR}}(\mathbf{y}) d\mathbf{y} &= \int_{\{\mathbf{y}\in\mathcal{L}: y_k>1\}} \exp\left\{ -\sum_{i=1}^{d} \mu_i \log(y_i) - \sum_{i=1}^{d}\sum_{j=1}^{d} \Theta_{ij} (\log y_i)(\log y_j) \right\} d\mathbf{y} \\ &= \int_{\{\mathbf{x}\in\mathbb{R}^d: x_k>0\}} \exp\{ -(\mu-\mathbf{1})^\top \mathbf{x} - \mathbf{x}^\top \Theta \mathbf{x} \} d\mathbf{x}. \end{aligned} \tag{8}$$

Decomposing $\mathbf{x}$ into $\mathbf{x}_{\setminus k}$ and x_k, we may write $(\mu-\mathbf{1})^\top \mathbf{x}$ as $(\mu-\mathbf{1})_{\setminus k}^\top \mathbf{x}_{\setminus k} + (\mu_k - 1)x_k$, and $\mathbf{x}^\top \Theta \mathbf{x}$ as

$$\mathbf{x}_{\setminus k}^\top \Theta^{(k)} \mathbf{x}_{\setminus k} + 2x_k \Theta_{k,\setminus k} \mathbf{x}_{\setminus k} + \Theta_{kk} x_k^2,$$

where $\Theta^{(k)} := \Theta_{\setminus k,\setminus k}$. We may then rewrite Equation (8) as

$$\int_0^\infty \int_{\mathbb{R}^{d-1}} \exp\{ -\mathbf{x}_{\setminus k}^\top \Theta^{(k)} \mathbf{x}_{\setminus k} - 2x_k \Theta_{k,\setminus k} \mathbf{x}_{\setminus k} - (\mu-\mathbf{1})_{\setminus k}^\top \mathbf{x}_{\setminus k} \} d\mathbf{x}_{\setminus k} \times \exp\{ -(\mu_k - 1)x_k - \Theta_{kk} x_k^2 \} dx_k.$$

The inner integral is a Gaussian type integral. It is straightforward to show that it is finite if and only if $\Theta^{(k)}$ is positive definite, using a spectral decomposition of that matrix. Deduce that for $f_{\mu,\Theta}^{\text{gen HR}}$ to be integrable, all the matrices $\Theta^{(k)}$ must be positive definite (hence, of full rank $d-1$). This implies that Θ must also be of rank $d-1$. Moreover, Θ has to have only nonnegative eigenvalues. Indeed, suppose it doesn't. Then there is an $\mathbf{x} \in \mathbb{R}^d$ such that $\mathbf{x}^\top \Theta \mathbf{x} < 0$. However, since $\Theta \mathbf{1} = 0$, it is also true that $0 > (\mathbf{x} - x_k \mathbf{1})^\top \Theta (\mathbf{x} - x_k \mathbf{1}) = (\mathbf{x} - x_k \mathbf{1})_{\setminus k}^\top \Theta^{(k)} (\mathbf{x} - x_k \mathbf{1})_{\setminus k}$, contradicting the positive definiteness of $\Theta^{(k)}$.

It is therefore necessary for the integrability of $f_{\mu,\Theta}^{\text{gen HR}}$ on $\mathcal{L}$ that $\Theta \in \mathcal{S}_{\mathbf{1},+}^{d\times d}$. For the remainder of the proof, we shall assume that this is the case.

2.2.2 | If $\mu^\top \mathbf{1} \le d$, then $f_{\mu,\Theta}^{\text{gen HR}}$ is not integrable

Now that we assume the matrices $\Theta^{(k)}$ to be invertible, we may obtain a more precise expression for Equation (8). By tedious but elementary computations involving completion of the quadratic form $\mathbf{x}_{\setminus k}^\top \Theta^{(k)} \mathbf{x}_{\setminus k}$, we may rewrite the argument of the exponential in Equation (8) as

$$-\left(\mathbf{x}_{\setminus k} - \beta^{(k)}(x_k)\right)^\top \Theta^{(k)} \left(\mathbf{x}_{\setminus k} - \beta^{(k)}(x_k)\right) + \beta^{(k)}(x_k)^\top \Theta^{(k)} \beta^{(k)}(x_k) + (1-\mu_k)x_k - \Theta_{kk} x_k^2,$$

where $\beta^{(k)}(x_k) := \frac{1}{2}\Sigma^{(k)}(\mathbf{1}-\mu)_{\setminus k} + x_k \mathbf{1}_{\setminus k}$ and $\Sigma^{(k)} := (\Theta^{(k)})^{-1}$. Since $\mathbf{x}_{\setminus k}$ only appears in the first term, we may rewrite Equation (8) as

$$\begin{aligned}&\int_0^\infty \left[\int_{\mathbb{R}^{d-1}} \exp\left\{-\left(\mathbf{x}_{\setminus k}-\beta^{(k)}(x_k)\right)^\top \Theta^{(k)}\left(\mathbf{x}_{\setminus k}-\beta^{(k)}(x_k)\right)\right\} d\mathbf{x}_{\setminus k} \times \exp\left\{\beta^{(k)}(x_k)^\top \Theta^{(k)}\beta^{(k)}(x_k) + (1-\mu_k)x_k - \Theta_{kk}x_k^2\right\}\right] dx_k \\ &= \frac{(2\pi)^{(d-1)/2}}{\det(\Theta^{(k)})^{1/2}} \int_0^\infty \exp\left\{\beta^{(k)}(x_k)^\top \Theta^{(k)}\beta^{(k)}(x_k) + (1-\mu_k)x_k - \Theta_{kk}x_k^2\right\} dx_k.\end{aligned} \tag{9}$$

Expanding the quadratic form $\beta^{(k)}(x_k)^\top \Theta^{(k)}\beta^{(k)}(x_k)$, we find that

$$\begin{aligned}\beta^{(k)}(x_k)^\top \Theta^{(k)}\beta^{(k)}(x_k) + (1-\mu_k)x_k - \Theta_{kk}x_k^2 &= -\left((\mu-\mathbf{1})^\top \mathbf{1}\right)x_k + \frac{1}{4}(\mu-\mathbf{1})_{\setminus k}^\top \Sigma^{(k)}(\mu-\mathbf{1})_{\setminus k} \\ &= -(\mu^\top \mathbf{1} - d)x_k + \frac{1}{4}(\mu-\mathbf{1})_{\setminus k}^\top \Sigma^{(k)}(\mu-\mathbf{1})_{\setminus k}.\end{aligned}$$

Therefore, Equation (9) is equal to

$$\frac{(2\pi)^{(d-1)/2}}{\det(\Theta^{(k)})^{1/2}} \exp\left\{\frac{1}{4}(\mu-\mathbf{1})_{\setminus k}^\top \Sigma^{(k)}(\mu-\mathbf{1})_{\setminus k}\right\} \int_0^\infty e^{-(\mu^\top \mathbf{1}-d)x_k} dx_k, \tag{10}$$

which is finite if and only if $\mu^\top \mathbf{1} > d$. This establishes that $f_{\mu,\Theta}^{\text{gen HR}}$ is integrable on $\mathcal{L}$ if and only if $\Theta \in \mathcal{S}_{\mathbf{1},+}^{d\times d}$ and $\mu^\top \mathbf{1} > d$, which in particular implies (*iii*).

2.2.3 | If $f_{\mu,\Theta}^{\text{gen HR}}$ is a multivariate Pareto density, then $\mu = \mu_{\text{HR}}(\Theta)$

Now suppose that $f_{\mu,\Theta}^{\text{gen HR}}$ is a multivariate Pareto density. By the last two sections, it follows that $\Theta \in \mathcal{S}_{\mathbf{1},+}^{d\times d}$. Moreover, by the homogeneity property (MP2) and the fact that $\Theta\mathbf{1} = 0$,

$$\log f_{\mu,\Theta}^{\text{gen HR}}(\mathbf{y}) - \log f_{\mu,\Theta}^{\text{gen HR}}(t\mathbf{y}) = \sum_{i=1}^d \mu_i \log t + \sum_{i=1}^d \sum_{j=1}^d \Theta_{ij}(\log y_i + \log y_j + \log t)\log t = \sum_{i=1}^d \mu_i \log t$$

must be equal to $(d+1)\log t$. That is, $\mu^\top \mathbf{1} = d+1$.

We shall now enforce the marginal standardization property (MP3). Recalling Equation (10), we now have

$$Z(\mu,\Theta) \int_{\{\mathbf{y}\in\mathcal{L}: y_k>1\}} f_{\mu,\Theta}^{\text{gen HR}}(\mathbf{y})d\mathbf{y} = \frac{(2\pi)^{(d-1)/2}}{\det(\Theta^{(k)})^{1/2}} \exp\left\{\frac{1}{4}(\mu-\mathbf{1})_{\setminus k}^\top \Sigma^{(k)}(\mu-\mathbf{1})_{\setminus k}\right\}.$$

By eq. (23) in Röttger et al. (2023), $\det(\Theta^{(k)})$ equals $1/d$ times the pseudodeterminant of Θ and as such, does not depend on k. Hence, the marginal standardization property holds if and only if the value of $(\mu-\mathbf{1})_{\setminus k}^\top \Sigma^{(k)}(\mu-\mathbf{1})_{\setminus k}$ is the same for each k.

The matrices $\Sigma^{(k)}$, however, enjoy a special structure. Let us augment $\Sigma^{(k)}$ by adding a row and column of zeros in its kth position, forming a matrix $\tilde{\Sigma}^{(k)} \in \mathbb{R}^{d\times d}$. Then the matrices $\tilde{\Sigma}^{(k)}$ satisfy

$$\tilde{\Sigma}_{ij}^{(k)} = \frac{1}{2}(\Gamma_{ik} + \Gamma_{jk} - \Gamma_{ij}),\; i,j \in \{1,\ldots,d\},$$

where Γ is the variogram matrix associated to Θ, as defined in Example 2. These are the same matrices $\tilde{\Sigma}^{(k)}$ as that appearing in Section 4.3 of Engelke and Hitz (2020). Then, for any two indices k and ℓ,

$$\begin{aligned}(\mu-\mathbf{1})_{\setminus k}^{\top}\Sigma^{(k)}(\mu-\mathbf{1})_{\setminus k}-(\mu-\mathbf{1})_{\setminus \ell}^{\top}\Sigma^{(\ell)}(\mu-\mathbf{1})_{\setminus \ell} &=(\mu-\mathbf{1})^{\top}\tilde{\Sigma}^{(k)}(\mu-\mathbf{1})-(\mu-\mathbf{1})^{\top}\tilde{\Sigma}^{(\ell)}(\mu-\mathbf{1})\\ &=\sum_{i=1}^{d}\sum_{j=1}^{d}\left(\tilde{\Sigma}_{ij}^{(k)}-\tilde{\Sigma}_{ij}^{(\ell)}\right)(\mu_i-1)(\mu_j-1)\\ &=\frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d}(\Gamma_{ik}-\Gamma_{i\ell}+\Gamma_{jk}-\Gamma_{j\ell})(\mu_i-1)(\mu_j-1)\\ &=\sum_{i=1}^{d}(\Gamma_{ik}-\Gamma_{i\ell})(\mu_i-1)\sum_{j=1}^{d}(\mu_j-1)\\ &=(\Gamma_{k\cdot}-\Gamma_{\ell\cdot})(\mu-\mathbf{1}),\end{aligned}$$

recalling the fact that $\mu^{\top}\mathbf{1}=d+1$, or equivalently $(\mu-\mathbf{1})^{\top}\mathbf{1}=1$. This can only be zero for any $k\neq\ell$ if $\Gamma_{k\cdot}(\mu-\mathbf{1})$ has the same value for every k, that is, $\Gamma(\mu-\mathbf{1})\in\{\gamma\mathbf{1}:\gamma\in\mathbb{R}\}$. By invertibility of Γ, this forms a nonsingular system of $d-1$ linear equations. The solution set is a one-dimensional linear subspace, only one element of which, say $\mu^{*}-\mathbf{1}$, satisfies $(\mu^{*}-\mathbf{1})^{\top}\mathbf{1}=1$. Therefore, for any given parameter matrix $\Theta\in\mathcal{S}_{\mathbf{1},+}^{d\times d}$, the parameter vector μ is uniquely determined by the multivariate Pareto structure. The resulting distribution can be none other than the Hüsler–Reiss distribution with precision matrix Θ.

Note that it can be verified by simple calculations that the unique solution to the above linear system is indeed the Hüsler–Reiss distribution. Indeed, using Lemma S.5.11 of Hentschel et al. (2022), it can be confirmed that $\Gamma(\mu_{\mathrm{HR}}(\Theta)-\mathbf{1})$ is indeed the vector $\mathbf{1}$ multiplied by the scalar $d^{-2}\mathbf{1}^{\top}\Gamma(\frac{1}{2}\Theta\Gamma-I)\mathbf{1}$, where I denotes the identity matrix, and that $(\mu_{\mathrm{HR}}(\Theta)-\mathbf{1})^{\top}\mathbf{1}=1$. □

ACKNOWLEDGEMENTS

The author wishes to thank the associate editor as well as two anonymous referees for helpful comments, which helped improve the content and presentation of this paper. Funding from the Natural Sciences and Engineering Research Council of Canada is gratefully acknowledged. Open Access funding enabled and organized by Projekt DEAL.

ORCID

Michaël Lalancette https://orcid.org/0009-0005-9580-7791

REFERENCES

Asenova, S., Mazo, G., & 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*(3), 461–500. https://doi.org/10.1007/s10687-021-00407-5

Bingham, N. H., Goldie, C. M., & Teugels, J. L. (1987). *Regular variation*: Cambridge University Press.

Brown, L. D. (1986). *Fundamentals of statistical exponential families with applications in statistical decision theory*, Institute of Mathematical Statistics Lecture Notes—Monograph Series, Vol. 9: Institute of Mathematical Statistics, Hayward, CA.

Darroch, J. N., Lauritzen, S. L., & Speed, T. P. (1980). Markov fields and log-linear interaction models for contingency tables. *Annals of Statistics*, *8*(3), 522–539.

Engelke, S., & Hitz, A. S. (2020). Graphical models for extremes (with discussion). *Journal of the Royal Statistical Society: Series B Statistical Methodology*, *82*(4), 871–932.

Engelke, S., Lalancette, M., & Volgushev, S. (2022). Learning extremal graphical structures in high dimensions. arXiv preprint arXiv:2111.00840.

Engelke, S., & Volgushev, S. (2022). Structure learning for extremal tree models. *Journal of the Royal Statistical Society: Series B Statistical Methodology*, *84*(5), 2055–2087.

Falk, M., & Guillou, A. (2008). Peaks-over-threshold stability of multivariate generalized Pareto distributions. *Journal of Multivariate Analysis*, *99*(4), 715–734.

Gumbel, E. J. (1960). Distributions des valeurs extrêmes en plusiers dimensions. *Publications de l'Institut de statistique de l'Université de Paris*, *9*, 171–173.

Hentschel, M., Engelke, S., & Segers, J. (2022). Statistical inference for Hüsler–Reiss graphical models through matrix completions. arXiv preprint arXiv: 2210.14292.

Ho, Z. W. O., & Dombry, C. (2019). Simple models for multivariate regular variation and the Hüsler–Reiß Pareto distribution. *Journal of Multivariate Analysis*, *173*, 525–550.

Hüsler, J., & Reiss, R.-D. (1989). Maxima of normal random vectors: Between independence and complete dependence. *Statistics & Probability Letters*, *7*(4), 283–286.

Inouye, D., Ravikumar, P., & Dhillon, I. (2016). Square root graphical models: Multivariate generalizations of univariate exponential families that permit positive dependencies. In *International Conference on Machine Learning*, PMLR, pp. 2445–2453.

Kestelman, H. (1947). On the functional equation $f(x+y)=f(x)+f(y)$. *Fundamenta Mathematicae*, *1*(34), 144–147.

Kiriliouk, A., Rootzén, H., Segers, J., & Wadsworth, J. L. (2019). Peaks over thresholds modeling with multivariate generalized pareto distributions. *Technometrics*, *61*(1), 123–135.

Klein, N., Orellana, J., Brincat, S. L., Miller, E. K., & Kass, R. E. (2020). Torus graphs for multivariate phase coupling analysis. *The Annals of Applied Statistics*, *14*(2), 635–660. https://doi.org/10.1214/19-AOAS1300

Lederer, J., & Oesting, M. (2023). Extremes in high dimensions: Methods and scalable algorithms. arXiv preprint arXiv:2303.04258.

Lin, L., Drton, M., & Shojaie, A. (2016). Estimation of high-dimensional graphical models using regularized score matching. *Electronic Journal of Statistics*, *10*(1), 806–854. https://doi.org/10.1214/16-EJS1126

Liu, H., Lafferty, J., & Wasserman, L. (2009). The nonparanormal: Semiparametric estimation of high dimensional undirected graphs. *Journal of Machine Learning Research*, *10*, 2295–2328.

Maathuis, M., Drton, M., Lauritzen, S., & Wainwright, M. (Eds.) (2019). *Handbook of graphical models Edited by* Maathuis, M., Drton, M., Lauritzen, S., & Wainwright, M., Chapman & Hall/CRC Handbooks of Modern Statistical Methods: CRC Press, Boca Raton, FL.

Rootzén, H., Segers, J., & Wadsworth, J. L. (2018). Multivariate peaks over thresholds models. *Extremes*, *21*(1), 115–145.

Rootzén, H., & Tajvidi, N. (2006). Multivariate generalized Pareto distributions. *Bernoulli*, *12*, 917–930.

Röttger, F., Coons, J. I., & Grosdos, A. (2023). Parametric and nonparametric symmetries in graphical models for extremes. arXiv preprint arXiv: 2306.00703.

Röttger, F., Engelke, S., & Zwiernik, P. (2023). Total positivity in multivariate extremes. *Annals of Statistics*, *51*, 962–1004.

Yang, E., Ravikumar, P., Allen, G. I., & Liu, Z. (2015). Graphical models via univariate exponential family distributions. *Journal of Machine Learning Research*, *16*, 3813–3847.

Yu, S., Drton, M., & Shojaie, A. (2022). Generalized score matching for general domains. *Information and Inference*, *11*(2), 739–780. https://doi.org/10.1093/imaiai/iaaa041

How to cite this article: Lalancette, M. (2023). On pairwise interaction multivariate Pareto models. *Stat*, *12*(1), e613. https://doi.org/10.1002/sta4.613