Multivariate expectile-based distribution: properties, Bayesian inference, and applications

Julyan Arbel, Stéphane Girard, Hien Duy Nguyen, Antoine Usseglio-Carleve

Multivariate expectile-based distribution: properties, Bayesian inference, and applications

BY J. ARBEL, S. GIRARD

Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK, 38000 Grenoble, France

julyan.arbel@inria.fr stephane.girard@inria.fr

H. D. NGUYEN

School of Mathematics and Physics, Univ. of Queensland, St. Lucia 4067, Queensland Australia

h.nguyen7@uq.edu.au

AND A. USSEGLIO-CARLEVE

Avignon Univ., Laboratoire de Mathématiques d'Avignon EA 2151, 84000 Avignon, France

antoine.usseglio-carleve@univ-avignon.fr

SUMMARY

Expectiles form a family of risk measures that have recently gained interest over the more common value-at-risk or return levels, primarily due to their capability to be determined by the probabilities of tail values and magnitudes of realisations at once. However, a prevalent and ongoing challenge of expectile inference is the problem of uncertainty quantification, which is especially critical in sensitive applications, such as in medical, environmental or engineering tasks. We address this issue by developing a novel distribution, termed the multivariate expectile-based distribution (MED), that possesses an expectile as a closed-form parameter. Desirable properties of the distribution, such as log-concavity, make it an excellent fitting distribution in multivariate applications. Maximum likelihood estimation and Bayesian inference algorithms are described. Simulated examples and applications to expectile and mode estimation illustrate the usefulness of the MED for uncertainty quantification.

Some key words: Multivariate expectiles; Bayesian inference; Maximum likelihood estimation; Multivariate distribution.

1. INTRODUCTION

Risk assessment mainly relies on drawing inference from quantiles, often referred to as value-at-risk in finance or return levels in environmental science. A shortcoming of quantiles is that they only use the information on the frequency of events and not on the actual magnitudes. This is an issue since taking account of the magnitudes of extreme events may be crucial in many fields of application. One way to tackle this problem is to work with expectiles, introduced in Newey and Powell (1987) since, unlike quantiles, expectiles depend on both the probability of tail values and the magnitude of realisations (Kuan et al., 2009).

In univariate settings, expectiles are defined as the minimizers of an asymmetric quadratic loss and as such, they benefit from a number of desirable properties. They are well-defined and unique whenever the underlying distribution has a finite first moment; see for instance Abdous

and Remillard (1995). Furthermore, they induce the only coherent, law-invariant and elicitable risk measure (see Ziegel, 2016), and therefore benefit from a natural backtesting methodology. Expectiles are therefore a sensible risk management tool, as a complement or an alternative to quantiles.

In the last five years, there has been an increasing interest in estimating expectiles, both at fixed probability levels (see *e.g.* Holzmann and Klar, 2016; Krätschmer and Zähle, 2017) as well as at extreme levels (Daouia et al., 2018; Girard et al., 2021). Another recent line of work garnering increasing interest in mathematical finance consists of extending the expectile risk measure to the multivariate setting: see, for example, Maume-Deschamps et al. (2017, 2018); Herrmann et al. (2018); Daouia and Paindaveine (2019); Beck et al. (2021), among others.

Here, we focus on the notion of multivariate geometric expectiles, as introduced by Herrmann et al. (2018), which is the analogue of the multivariate geometric notion of quantiles proposed by Chaudhuri (1996).

1·1. *Our contribution*

We construct a multivariate distribution (Section 2), which is based on the loss function used for defining multivariate geometric expectiles (Herrmann et al., 2018). This distribution, termed the multivariate expectile-based distribution (MED), is parameterized by a location vector, scale matrix, asymmetry level and direction. We argue that inference employing the MED offers a practical way to handle uncertainty quantification in expectile estimation. Our specific contributions are summarized below.

We first prove a number of appealing properties of the MED: it is log-concave, unimodal, and the unique mode is an expectile whose parameters (asymmetry level and direction) are specified as closed-form parameters of the distribution. Doing so, we propose an elementary proof of the strict convexity of the above-mentioned loss function. We establish that the MED family is closed under linear transformation, leading to a simple standardization of the distribution.

We provide bounds on the MED density (Section 3), showing that it is sub-Gaussian. Limiting cases of the asymmetry level are studied, proving that the MED comprises the multivariate Gaussian and some asymmetric Gaussian distributions as special cases. Various closed-form moments are provided, along with marginal and conditional distributions (Section 4).

The main benefit of our proposal lies in inference. We propose a maximum likelihood approach (Section 5·1) which relies on optimization algorithms in Riemannian geometries. Consistency of the maximum likelihood estimator is established under compactness assumptions on the parameter space. We devise a Bayesian algorithm based on Hamiltonian Monte Carlo (Section 5·2). We also describe our accompanying R package (R Core Team, 2019) containing statistical functions related to the multivariate expectile-based distribution, available at `https://github.com/AntoineUC/MEDdist`, and provide a simulation study (Section 6).

We finally showcase three multivariate applications of the MED (Section 7): first, Bayesian estimation of expectiles and of the mode of a distribution, and last illustration to discriminant analysis. We show that a Bayesian approach relying on the MED yields well-calibrated credible intervals, thus addressing the uncertainty quantification shortcoming in expectile estimation.

All proofs and some additional implementation and applications results are deferred to the Appendix.

1·2. *Notations*

We denote asymptotic equivalence between positive real functions f and g by $f(y) \underset{y\to a}{\approx} g(y)$, meaning that $f(y)/g(y) \to 1$ when $y \to a$ for $a \in [-\infty, \infty]$. Vectors and matrices are denoted

with bold face, for instance $\boldsymbol{Z} = (Z_1, \ldots, Z_d)$ denotes a vector in $\mathbb{R}^d$, $\mathbf{0}$ denotes the vector of zeros, and $(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_d)$ denotes the canonical basis of $\mathbb{R}^d$. We write the last $d-1$ entries of $\boldsymbol{Z}$ as $\boldsymbol{Z}_{-1}$, in the sense that $\boldsymbol{Z} = (Z_1, \boldsymbol{Z}_{-1})$. The sphere in dimension d is denoted by $\mathcal{S}^{d-1}$. The set of $d \times d$ symmetric positive definite matrices is denoted by $\mathrm{SPD}(d)$, the identity matrix is denoted by $\boldsymbol{I}_d$. Let $\langle \cdot, \cdot \rangle$ denote the usual dot product and let $\| \cdot \|$ be the associated Euclidean norm. For any matrix $\boldsymbol{M} \in \mathrm{SPD}(d)$, we set $\langle \cdot, \cdot \rangle_{\boldsymbol{M}} \coloneqq \langle \cdot, \boldsymbol{M} \cdot \rangle$ and denote by $\| \cdot \|_{\boldsymbol{M}}$ the associated norm. The determinant of some matrix $\boldsymbol{M}$ is denoted by $|\boldsymbol{M}|$. We denote by Γ the gamma function and by ${}_2F_1$ the hypergeometric function. The d-dimensional Gaussian density with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$ is denoted by $\phi_d(\,\cdot\,; \boldsymbol{\mu}, \boldsymbol{\Sigma})$. The Gaussian, uniform, Wishart, and inverse-Wishart distributions are respectively denoted by $\mathcal{N}$, $\mathcal{U}$, $\mathcal{W}$, and $\mathcal{IW}$.

2. Multivariate expectile-based distribution

For any $\rho \in [0, 1)$, for any $\boldsymbol{\Sigma} \in \mathrm{SPD}(d)$ and $\boldsymbol{\nu} \in \mathbb{R}^d$, such that $\|\boldsymbol{\nu}\|_{\boldsymbol{\Sigma}^{-1}} = 1$, consider the loss function defined for any $\boldsymbol{y} \in \mathbb{R}^d$ by

$$\Lambda_{\rho,\boldsymbol{\nu},\boldsymbol{\Sigma}}(\boldsymbol{y}) = \|\boldsymbol{y}\|_{\boldsymbol{\Sigma}^{-1}} \left(\|\boldsymbol{y}\|_{\boldsymbol{\Sigma}^{-1}} + \rho \langle \boldsymbol{y}, \boldsymbol{\nu} \rangle_{\boldsymbol{\Sigma}^{-1}} \right).$$

In the particular case where $\boldsymbol{\Sigma} = \boldsymbol{I}_d$, this loss function has been used to define the multivariate geometric expectile (Herrmann et al., 2018) of a random vector $\boldsymbol{X}$ as

$$\operatorname*{arg\,min}_{\boldsymbol{x} \in \mathbb{R}^d} E\left(\Lambda_{\rho,\boldsymbol{\nu},\boldsymbol{I}_d}(\boldsymbol{X} - \boldsymbol{x}) \right).$$

It has been shown that this optimisation problem is well-defined and benefits from a unique solution provided that the margins of $\boldsymbol{X}$ have finite second order moments (Proposition 4.6 of Herrmann et al., 2018). We show in Proposition A1 in Appendix that this condition can be relaxed to assuming only a *finite first order moment*. This notion can be extended without difficulty to any $\boldsymbol{\Sigma} \in \mathrm{SPD}(d)$ by setting:

$$\mathbf{e}_{\rho,\boldsymbol{\nu},\boldsymbol{\Sigma}}(\boldsymbol{X}) = \operatorname*{arg\,min}_{\boldsymbol{x} \in \mathbb{R}^d} E\left(\Lambda_{\rho,\boldsymbol{\nu},\boldsymbol{\Sigma}}(\boldsymbol{X} - \boldsymbol{x}) \right). \tag{1}$$

Indeed, existence and unicity of the expectile defined in (1) stems from Proposition A2 in Appendix on the strict convexity of $\Lambda_{\rho,\boldsymbol{\nu},\boldsymbol{\Sigma}}$, which extends the result of Herrmann et al. (2018, Theorem 4.3) which deals with the particular case of $\Lambda_{\rho,\boldsymbol{\nu},\boldsymbol{I}_d}$. A proof relying on more elementary arguments than Herrmann et al. (2018)'s proof is provided in Appendix.

In view of (1), one clearly has $\mathbf{e}_{\rho,\boldsymbol{\nu},\boldsymbol{\Sigma}}(\boldsymbol{X} + \boldsymbol{b}) = \mathbf{e}_{\rho,\boldsymbol{\nu},\boldsymbol{\Sigma}}(\boldsymbol{X}) + \boldsymbol{b}$, for any $\boldsymbol{b} \in \mathbb{R}^d$ and $\mathbf{e}_{\rho,\boldsymbol{\nu},\boldsymbol{\Sigma}}(\sigma \boldsymbol{X}) = \sigma \mathbf{e}_{\rho,\boldsymbol{\nu},\boldsymbol{\Sigma}}(\boldsymbol{X})$ for any $\sigma > 0$, which are the equivariance properties of multivariate geometric expectiles with respect to translation and scale respectively. See Herrmann et al. (2018, Proposition 4.1 and Proposition 4.2) in the case where $\boldsymbol{\Sigma} = \boldsymbol{I}_d$.

Definition 1. *Let $\rho \in [0, 1)$, $\boldsymbol{\Sigma} \in \mathrm{SPD}(d)$ and $\boldsymbol{\mu}$, $\boldsymbol{\nu} \in \mathbb{R}^d$ such that $\|\boldsymbol{\nu}\|_{\boldsymbol{\Sigma}^{-1}} = 1$. The multivariate expectile-based distribution, denoted $\mathrm{MED}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\nu}, \rho)$ or $\mathrm{MED}(\boldsymbol{\theta})$ where $\boldsymbol{\theta}$ is a shorthand notation for $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, $\boldsymbol{\nu}$, and ρ, is defined on $\mathbb{R}^d$ by its density*

$$f_d(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\nu}, \rho) = \frac{c_d(\rho)}{(2\pi)^{d/2} |\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\tfrac{1}{2} \Lambda_{\rho,\boldsymbol{\nu},\boldsymbol{\Sigma}}(\boldsymbol{x} - \boldsymbol{\mu}) \right),$$

for any $\boldsymbol{x} \in \mathbb{R}^d$, where

$$c_d(\rho) = \sqrt{1 - \rho^2} \left(\frac{1 + \sqrt{1 - \rho^2}}{2} \right)^{(d-2)/2}. \tag{2}$$

See Lemma A1 in the Appendix for a proof that $f_d(\cdot; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\nu}, \rho)$ integrates to one. A straightforward application of Proposition A2 shows that the MED is log-concave, which implies in turn that it is unimodal, that its level sets are closed compact sets, and its marginal distributions remain log-concave (Brascamp and Lieb, 2002). Let us now show that MED family is closed under linear transformations. Moreover, parameter ρ is invariant with respect to such transformation. The proof is straightforward.

PROPOSITION 1 (LINEAR TRANSFORMATION). *Let $\boldsymbol{A}$ be a non-singular $d \times d$ matrix and let $\boldsymbol{b} \in \mathbb{R}^d$. If $\boldsymbol{X} \sim \mathrm{MED}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\nu}, \rho)$, then $\boldsymbol{AX} + \boldsymbol{b} \sim \mathrm{MED}(\boldsymbol{A\mu} + \boldsymbol{b}, \boldsymbol{A\Sigma A}^\top, \boldsymbol{A\nu}, \rho)$.*

As an immediate consequence of Proposition 1, we have:

COROLLARY 1 (STANDARDIZATION). *Let $\rho \in [0, 1)$, $\boldsymbol{\Sigma} \in \mathrm{SPD}(d)$ and $\boldsymbol{\mu}, \boldsymbol{\nu} \in \mathbb{R}^d$, with $\|\boldsymbol{\nu}\|_{\boldsymbol{\Sigma}^{-1}} = 1$. Consider a $d \times d$ orthogonal matrix $\boldsymbol{Q}$ such that $\boldsymbol{Q\Sigma}^{-1/2}\boldsymbol{\nu} = \boldsymbol{u}_1$. If $\boldsymbol{Z} \sim \mathrm{MED}(\boldsymbol{0}, \boldsymbol{I}_d, \boldsymbol{u}_1, \rho)$, then $\boldsymbol{X} \coloneqq \boldsymbol{\Sigma}^{1/2}\boldsymbol{Q}^\top\boldsymbol{Z} + \boldsymbol{\mu} \sim \mathrm{MED}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\nu}, \rho)$.*

In the sequel, $\mathrm{MED}(\boldsymbol{0}, \boldsymbol{I}_d, \boldsymbol{u}_1, \rho)$ is referred to as the standardized MED and, for the sake of presentation, it is denoted by $\mathrm{MED}(\rho)$, and the associated density and objective function are denoted respectively by $f_d(\,\cdot\,; \rho) \coloneqq f_d(\,\cdot\,; \boldsymbol{0}, \boldsymbol{I}_d, \boldsymbol{u}_1, \rho)$ and $\Lambda_\rho(\,\cdot\,) \coloneqq \Lambda_{\boldsymbol{I}_d, \boldsymbol{u}_1, \rho}(\,\cdot\,)$. Note that the standardization operation (or inverse standardization operation) requires obtaining a $d \times d$ orthogonal matrix $\boldsymbol{Q}$ such that $\boldsymbol{Q\Sigma}^{-1/2}\boldsymbol{\nu} = \boldsymbol{u}_1$, or such that $\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\nu} = \boldsymbol{Q}^\top\boldsymbol{u}_1$. The latter condition means that the first column of $\boldsymbol{Q}^\top$ is equal to $\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\nu}$. Completing the last $(d-1)$ columns of that matrix with the last $(d-1)$ basis vectors $\boldsymbol{u}_2, \ldots, \boldsymbol{u}_d$ provides a $d \times d$ matrix that can be transformed into a valid orthogonal matrix $\boldsymbol{Q}^\top$ by using the Gram–Schmidt process.

PROPOSITION 2 (PARTICULAR CASES). *For $\boldsymbol{X} \sim \mathrm{MED}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\nu}, \rho)$:*

(i) If $\rho = 0$, then $\boldsymbol{X}$ is multivariate Gaussian $\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.
(ii) If $\rho \to 1$, let $\boldsymbol{\Delta}_\rho = \mathrm{diag}(\sqrt{1-\rho}, 1/\sqrt{2}, \ldots, 1/\sqrt{2})$ and $\boldsymbol{Z} = \boldsymbol{Q\Sigma}^{-1/2}(\boldsymbol{X} - \boldsymbol{\mu})$ the standardized MED; then $\boldsymbol{\Delta}_\rho\boldsymbol{Z}$ converges in distribution to a half-standard Gaussian restricted to negative values of the first component.
(iii) If $d = 1$, then X is the one-dimensional asymmetric Gaussian random variable of Kato et al. (2002).

We can interpret MED as a multivariate extension of the one-dimensional asymmetric Gaussian distribution (AGD) introduced in Kato et al. (2002). Tail properties of the one-dimensional AGD are established in Chen and Huang (2014), while this distribution is used in Fu and Zhou (2016) to build a model-based classifier. An alternative multivariate extension is applied in He et al. (2018), to the problem of image denoising. We also refer to Majumdar and Paul (2016) for an AGD-based construction of a stationary spatial noise process model, whose marginal distributions have certain (one-dimensional) expectiles equal to zero.

In the following, $\rho \in [0, 1)$ is referred to as the asymmetry parameter, while $\boldsymbol{\nu}$ is called the asymmetry direction. See Figure 1 for an illustration of the possible shapes obtained by varying parameters the values of ρ, $\boldsymbol{\nu}$, and $\boldsymbol{\Sigma}$. Characteristics of the location $\boldsymbol{\mu}$ of the MED are investigated in the next result.

PROPOSITION 3 (MODE, EXPECTILE AND EXPECTATION). *Let $\boldsymbol{X} \sim \mathrm{MED}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\nu}, \rho)$. The distribution of $\boldsymbol{X}$ is unimodal and satisfies:*

(i) $\mathrm{mode}(\boldsymbol{X}) = \boldsymbol{\mu}$,
(ii) $\mathbf{e}_{\rho, \boldsymbol{\nu}, \boldsymbol{\Sigma}}(\boldsymbol{X}) = \boldsymbol{\mu}$,

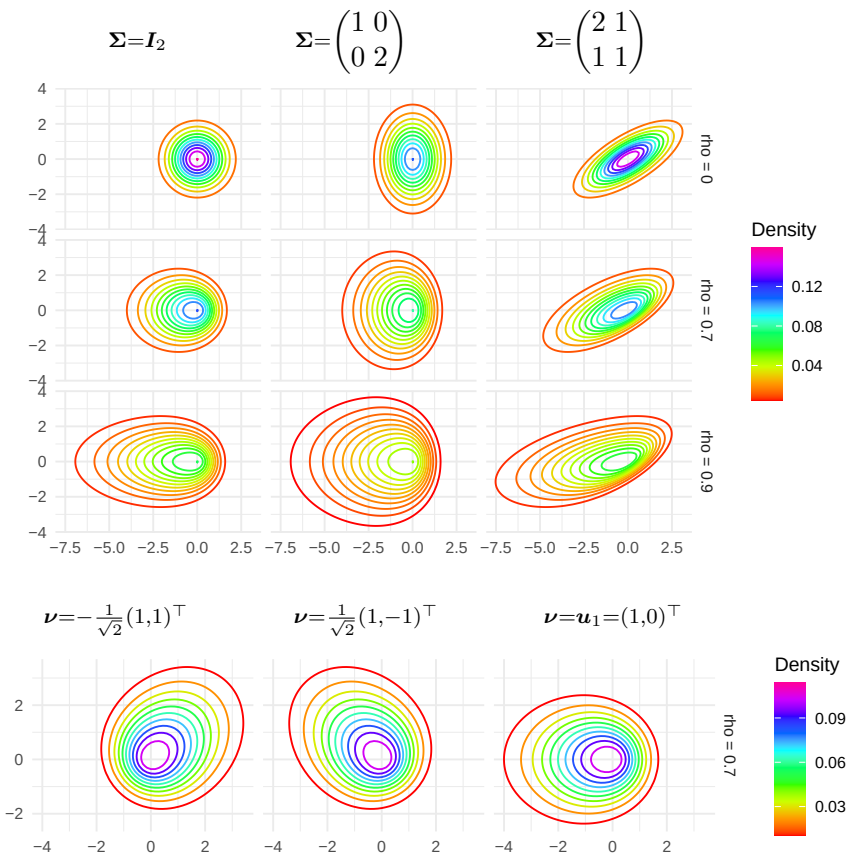

Fig. 1. Contour plots of densities $f_2(\,\cdot\,;\mathbf{0},\boldsymbol{\Sigma},\boldsymbol{\nu},\rho)$ in $\mathbb{R}^2$. Top panel: asymmetry direction $\boldsymbol{\nu}=\boldsymbol{u}_1$ fixed, varying asymmetry parameter $\rho\in\{0,0.7,0.9\}$ (row-wise) and matrices $\boldsymbol{\Sigma}$ (column-wise). Bottom panel: fixed $\boldsymbol{\Sigma}=\boldsymbol{I}_2$, $\rho=0.7$, and varying direction $\boldsymbol{\nu}$.

(iii) $E(\boldsymbol{X})=\boldsymbol{\mu}-\frac{\Gamma\left(\frac{d+3}{2}\right)}{\sqrt{2}\Gamma\left(\frac{d+2}{2}\right)}\,{}_2F_1\left(\frac{d+3}{4},\frac{d+5}{4};\frac{d+2}{2};\rho^2\right)\rho\sqrt{1-\rho^2}\left(\frac{1+\sqrt{1-\rho^2}}{2}\right)^{\frac{d}{2}-1}\boldsymbol{\nu}.$

The vector $\boldsymbol{\mu}$ is both the mode and an expectile of the MED. In contrast, the expectation of the MED also depends on the asymmetry direction: it is located in the two-dimensional subspace spanned by $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$. Specifying the above expectation in the one-dimensional case $\mathrm{MED}(\mu,\sigma,1,\rho)$, we recover the expectation of the one-dimensional asymmetric Gaussian of Kato et al. (2002): $E\left(\frac{X-\mu}{\sigma}\right)=-\frac{c_1(\rho)\sqrt{2}}{\sqrt{\pi}}\frac{\rho}{1-\rho^2}$. Second moment and skewness centered with respect to μ are given by $E\left(\frac{(X-\mu)^2}{\sigma^2}\right)=\frac{c_1(\rho)}{2}\left[\frac{1}{(1-\rho)^{3/2}}+\frac{1}{(1+\rho)^{3/2}}\right]$ and $E\left(\frac{(X-\mu)^3}{\sigma^3}\right)=-\frac{c_1(\rho)\sqrt{2}}{\sqrt{\pi}}\frac{4\rho}{(1-\rho^2)^2}$, respectively. It can be seen from Figure 5 in Appendix that asymmetry increases with ρ, corroborating the illustration of Figure 1. The role of the asymmetry parameter is further illustrated with next proposition which studies the variations of expectiles with respect to variations of the asymmetry parameter for a given MED.

PROPOSITION 4 (EXPECTILE WITH A DIFFERENT ASYMMETRY PARAMETER). *Let $\rho_1, \rho_2 \in [0,1)$. The expectile of parameters $(\rho_2, \boldsymbol{u}_1, \boldsymbol{I}_d)$ for the standardized $\boldsymbol{Z} \sim \mathrm{MED}(\rho_1)$ is supported on the first axis, and satisfies the following equivalences*

- $\langle \mathbf{e}_{\rho_2,\boldsymbol{u}_1,\boldsymbol{I}_d}(\boldsymbol{Z}), \boldsymbol{u}_1 \rangle > 0$ *if and only if* $\rho_2 > \rho_1$*;*
- $\langle \mathbf{e}_{\rho_2,\boldsymbol{u}_1,\boldsymbol{I}_d}(\boldsymbol{Z}), \boldsymbol{u}_1 \rangle = 0$ *(ie* $\mathbf{e}_{\rho_2,\boldsymbol{u}_1,\boldsymbol{I}_d}(\boldsymbol{Z}) = \mathbf{0}$*) if and only if* $\rho_2 = \rho_1$*;*
- $\langle \mathbf{e}_{\rho_2,\boldsymbol{u}_1,\boldsymbol{I}_d}(\boldsymbol{Z}), \boldsymbol{u}_1 \rangle < 0$ *if and only if* $\rho_2 < \rho_1$.

3. MOMENTS, SUB-GAUSSIANITY AND BOUNDS

The next Proposition provides both upper and lower bounds on the MED density based on the Gaussian density. This result will reveal itself to be useful to sample from the MED, using an acceptance-rejection method; see Section 5·3. To this end, let us consider, for any $\rho \in [0,1)$,

$$m_d(\rho) = c_d(\rho)(1+\rho)^{-d/2} \text{ and } M_d(\rho) = c_d(\rho)(1-\rho)^{-d/2}.$$

PROPOSITION 5 (DENSITY BOUNDS). *For any $\boldsymbol{x} \in \mathbb{R}^d$, one has*

$$m_d(\rho)\phi_d\left(\boldsymbol{x}; \boldsymbol{\mu}, \frac{\boldsymbol{\Sigma}}{1+\rho}\right) \leq f_d(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\nu}, \rho) \leq M_d(\rho)\phi_d\left(\boldsymbol{x}; \boldsymbol{\mu}, \frac{\boldsymbol{\Sigma}}{1-\rho}\right).$$

As a consequence, one can easily derive bounds on moments of a projected MED random vector.

COROLLARY 2 (BOUNDS ON PROJECTED MOMENTS). *Let $\boldsymbol{X} \sim \mathrm{MED}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\nu}, \rho)$. For any non-zero $\boldsymbol{\beta} \in \mathbb{R}^d$ and $p > 0$, one has*

$$m_d(\rho)\left(\frac{2}{1+\rho}\right)^{p/2} \leq \frac{\pi^{1/2} E\left(|\langle \boldsymbol{\beta}, \boldsymbol{X} - \boldsymbol{\mu}\rangle_{\boldsymbol{\Sigma}^{-1}}|^p\right)}{\Gamma\left((p+1)/2\right)} \leq M_d(\rho)\left(\frac{2}{1-\rho}\right)^{p/2}.$$

In the terminology of Vladimirova et al. (2020), MED random variables are thus sub-Weibull with optimal tail parameter $1/2$, that is to say they are sub-Gaussian, but no lighter than sub-Gaussian. The sub-Weibull property in a multivariate setting entails projections $\langle \boldsymbol{\beta}, \boldsymbol{Z}\rangle_{\boldsymbol{\Sigma}^{-1}}$ for any non-zero vector $\boldsymbol{\beta} \in \mathbb{R}^d$ to satisfy the (univariate) sub-Weibull property.

The next proposition provides the quadratic moment of the MED.

PROPOSITION 6 (QUADRATIC MOMENT). *Let $\boldsymbol{X} \sim \mathrm{MED}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\nu}, \rho)$. Then*

$$E\left[\|\boldsymbol{X} - \boldsymbol{\mu}\|^2_{\boldsymbol{\Sigma}^{-1}}\right] = \frac{d}{\sqrt{1-\rho^2}} + \frac{2\rho^2}{(1-\rho^2)(1+\sqrt{1-\rho^2})}.$$

We always have $E\left[\|\boldsymbol{X} - \boldsymbol{\mu}\|^2_{\boldsymbol{\Sigma}^{-1}}\right] \geq d$. If $\rho = 0$, we get $E\left[\|\boldsymbol{X} - \boldsymbol{\mu}\|^2_{\boldsymbol{\Sigma}^{-1}}\right] = d$, the mean of a χ^2 distribution with d degrees of freedom. The next proposition provides the limiting behaviour of the expectation and quadratic moment when $\rho \to 1$.

PROPOSITION 7 (ASYMPTOTICS OF MOMENTS). *Let $\boldsymbol{X} \sim \mathrm{MED}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\nu}, \rho)$. Then, when $\rho \to 1$, we have the following asymptotic equivalents*

$$(i) \quad E\left[\boldsymbol{X} - \boldsymbol{\mu}\right] \underset{\rho\to 1}{\approx} -\frac{\Gamma(\frac{d+3}{2})}{2^{d/2}\Gamma(\frac{d+3}{4})\Gamma(\frac{d+5}{4})}\frac{1}{\sqrt{1-\rho}}\boldsymbol{\nu},$$

$$(ii) \quad E\left[\|\boldsymbol{X} - \boldsymbol{\mu}\|^2_{\boldsymbol{\Sigma}^{-1}}\right] \underset{\rho\to 1}{\approx} \frac{1}{1-\rho}.$$

4. CONDITIONAL & MARGINAL DISTRIBUTIONS

Let us first focus on the margins of the standard MED. As alluded to in the Introduction, the log-concavity of the MED implies log-concavity of any of its marginal distributions. Two distributions are considered: the univariate margin in the asymmetry direction $\boldsymbol{u}_1$ and the $(d-1)$-dimensional margin in the orthogonal direction, denoted by $\boldsymbol{u}_1^{\perp}$. In the second case, the marginal distribution is elliptically contoured (Cambanis et al., 1981) or, more precisely, spherical when $\boldsymbol{\Sigma} = \boldsymbol{I}_d$.

PROPOSITION 8 (MARGINALS). *Let $\boldsymbol{Z} \sim \mathrm{MED}(\rho)$ with $d \geq 2$, and write $\boldsymbol{Z} = (Z_1, \boldsymbol{Z}_{-1})$.*

(i) In the direction $\boldsymbol{u}_1$, the marginal density of Z_1, denoted by f_{Z_1}, satisfies for any $z_1 \in \mathbb{R}$:

$$\frac{1}{(1+\rho)^{(d-1)/2}} \exp\left(-\frac{z_1^2}{2}(1+\rho)\right) \leq \frac{(2\pi)^{1/2}}{c_d(\rho)} f_{Z_1}(z_1) \leq \frac{1}{(1-\rho)^{(d-1)/2}} \exp\left(-\frac{z_1^2}{2}(1-\rho)\right),$$

and

$$f_{Z_1}(z_1) \underset{z_1 \to +\infty}{\approx} \frac{c_d(\rho)}{(2\pi)^{1/2}} \left(\frac{1}{1+\rho/2}\right)^{(d-1)/2} \exp\left(-\frac{z_1^2}{2}(1+\rho)\right), \tag{3}$$

$$f_{Z_1}(z_1) \underset{z_1 \to -\infty}{\approx} \frac{c_d(\rho)}{(2\pi)^{1/2}} \left(\frac{1}{1-\rho/2}\right)^{(d-1)/2} \exp\left(-\frac{z_1^2}{2}(1-\rho)\right). \tag{4}$$

(ii) In the hyperplane orthogonal to $\boldsymbol{u}_1$, the marginal density of $\boldsymbol{Z}_{-1}$, denoted by $f_{\boldsymbol{Z}_{-1}}$, is an elliptically contoured distribution. That is, there exists a density $f_{\perp}$ on $\mathbb{R}_+$ such that for any $\boldsymbol{z}_{-1} \in \mathbb{R}^{d-1}$:

$$f_{\boldsymbol{Z}_{-1}}(\boldsymbol{z}_{-1}) = \tilde{f}(\|\boldsymbol{z}_{-1}\|), \tag{5}$$

and

$$\tilde{f}(z) \underset{z \to \infty}{\approx} \frac{1}{(2\pi\sigma^2(\rho))^{(d-1)/2}} \exp\left(-\frac{z^2}{2\sigma^2(\rho)}\right), \tag{6}$$

where $\sigma^2(\rho) = \frac{2}{1+\sqrt{1-\rho^2}}$.

Let us note that the bounds and the asymptotic equivalents in Proposition 8 (i) all coincide in the particular case where $\rho = 0$, yielding back the standard Gaussian density. When $\rho \in (0,1)$, the gap between asymptotic equivalences (3) and (4) shows that the univariate marginal distribution of (i) has asymmetric tails. The asymptotic equivalence of (5) and (6) shows that the $(d-1)$-dimensional marginal of (ii) reduces to a centered Gaussian density on $\mathbb{R}^{d-1}$.

We conclude this section with some one-dimensional and $(d-1)$-dimensional conditional distributions.

PROPOSITION 9 (CONDITIONALS). *Let $\boldsymbol{Z} \sim \mathrm{MED}(\rho)$ with $d \geq 2$.*

(i) Let $\boldsymbol{w} \in \mathbb{R}^d$, such that $\|\boldsymbol{w}\| = 1$, and denote by $P_{\boldsymbol{w}^{\perp}}(\boldsymbol{Z})$ the projection of $\boldsymbol{Z}$ on the hyperplane orthogonal to $\boldsymbol{w}$. The one-dimensional conditional distribution of $\langle \boldsymbol{w}, \boldsymbol{Z} \rangle$ given that $P_{\boldsymbol{w}^{\perp}}(\boldsymbol{Z}) = \boldsymbol{0}$ coincides with the one-dimensional MED *with $\mu = 0$, $\Sigma = 1$, asymmetry parameter equal to $\rho\,|\langle \boldsymbol{w}, \boldsymbol{u}_1 \rangle|$, and asymmetry vector* $\mathrm{sign}(w_1)$*, where $w_1 = \langle \boldsymbol{w}, \boldsymbol{u}_1 \rangle$,*

$$\langle \boldsymbol{w}, \boldsymbol{Z} \rangle \mid P_{\boldsymbol{w}^{\perp}}(\boldsymbol{Z}) = \boldsymbol{0} \sim \mathrm{MED}(0, 1, \rho\,|\langle \boldsymbol{w}, \boldsymbol{u}_1 \rangle|, \mathrm{sign}(\langle \boldsymbol{w}, \boldsymbol{u}_1 \rangle)).$$

(ii) Write $\boldsymbol{Z} = (Z_1, \boldsymbol{Z}_{-1})$. Then, the $(d-1)$-dimensional conditional of $\boldsymbol{Z}_{-1}$, given $Z_1 = 0$, is a standard multivariate Gaussian distribution:

$$\boldsymbol{Z}_{-1} \mid Z_1 = 0 \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_{d-1}).$$

Note that in *(i)* above, the considered distribution is one-dimensional, thus the asymmetry vector is either 1 or -1, which can be written as the sign of a scalar product. According to Proposition 2 (iii), the conditional distribution given in *(i)* also corresponds to the one-dimensional asymmetric Gaussian density introduced in Kato et al. (2002). When projecting in this way, the asymmetry parameter is reduced by a factor of $|\langle \boldsymbol{w}, \boldsymbol{u}_1 \rangle| \leq 1$. In particular, it remains unchanged if the projection direction $\boldsymbol{w}$ is identical to $\boldsymbol{u}_1$, while it vanishes when $\boldsymbol{w} \perp \boldsymbol{u}_1$; in the latter case, the distribution obtained coincides with the standard Gaussian distribution.

5. STATISTICAL INFERENCE

In this section, we start by describing inference approaches for fitting the MED, either based on maximum likelihood or on a Bayesian model. We conclude by explaining how the statistical functions for MLE and Bayesian inference can be used in R via the **MEDdist** package.

5·1. *Maximum likelihood estimation*

Let $\boldsymbol{\theta} = (\boldsymbol{\mu}, \boldsymbol{\Sigma}, \rho, \boldsymbol{\nu})$, where $\boldsymbol{\mu} \in \mathbb{R}^d$, $\boldsymbol{\Sigma} \in \text{SPD}\,(d)$, $\rho \in [0, 1)$, and $\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\nu} \in \mathcal{S}^{d-1}$. For estimation purposes, it is more convenient to work with the parameterization $\tilde{\boldsymbol{\theta}} = (\boldsymbol{\mu}, \boldsymbol{\Sigma}, \rho, \tilde{\boldsymbol{\nu}})$ instead, which replaces $\boldsymbol{\nu}$ in $\boldsymbol{\theta}$, by $\tilde{\boldsymbol{\nu}} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\nu}$.

Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n \overset{\text{iid}}{\sim} \text{MED}(\tilde{\boldsymbol{\theta}})$. Then, we may write the log-likelihood up to an additive constant as $\ell_n(\tilde{\boldsymbol{\theta}}) = n \log(c_d(\rho)) - \frac{n}{2} \log|\boldsymbol{\Sigma}| - \frac{1}{2}\sum_{i=1}^n \|\boldsymbol{X}_i - \boldsymbol{\mu}\|^2_{\boldsymbol{\Sigma}^{-1}} - \frac{\rho}{2}\sum_{i=1}^n \langle \boldsymbol{X}_i - \boldsymbol{\mu}, \tilde{\boldsymbol{\nu}} \rangle_{\boldsymbol{\Sigma}^{-1/2}} \|\boldsymbol{X}_i - \boldsymbol{\mu}\|_{\boldsymbol{\Sigma}^{-1}}$. We can minimize the negative log-likelihood using a coordinate descent algorithm over each of the different components of $\tilde{\boldsymbol{\theta}}$. That is, let $\tilde{\boldsymbol{\theta}}^{(t)}$ be the tth iteration of the coordinate descent algorithm. Then, when $t = 0 \bmod 4$, we set $\boldsymbol{\Sigma}^{(t)} = \boldsymbol{\Sigma}^{(t-1)}$, $\rho^{(t)} = \rho^{(t-1)}$, $\tilde{\boldsymbol{\nu}}^{(t)} = \tilde{\boldsymbol{\nu}}^{(t-1)}$, and make the update

$$\boldsymbol{\mu}^{(t)} = \underset{\boldsymbol{m} \in \mathbb{R}^d}{\arg\min}\; q_{\boldsymbol{\mu}}\left(\boldsymbol{m}; \tilde{\boldsymbol{\theta}}^{(t-1)}\right), \tag{7}$$

with $q_{\boldsymbol{\mu}}\left(\boldsymbol{m}; \tilde{\boldsymbol{\theta}}\right) = \sum_{i=1}^n \|\boldsymbol{X}_i - \boldsymbol{m}\|^2_{\boldsymbol{\Sigma}^{-1}} + \rho\sum_{i=1}^n \langle \boldsymbol{X}_i - \boldsymbol{m}, \tilde{\boldsymbol{\nu}} \rangle_{\boldsymbol{\Sigma}^{-1/2}} \|\boldsymbol{X}_i - \boldsymbol{m}\|_{\boldsymbol{\Sigma}^{-1}}$. When $t = 1 \bmod 4$, we set $\boldsymbol{\mu}^{(t)} = \boldsymbol{\mu}^{(t-1)}$, $\rho^{(t)} = \rho^{(t-1)}$, $\tilde{\boldsymbol{\nu}}^{(t)} = \tilde{\boldsymbol{\nu}}^{(t-1)}$, and make the update

$$\boldsymbol{\Sigma}^{(t)} = \underset{\boldsymbol{S} \in \text{SPD}(d)}{\arg\min}\; q_{\boldsymbol{\Sigma}}\left(\boldsymbol{S}; \tilde{\boldsymbol{\theta}}^{(t-1)}\right), \tag{8}$$

with $q_{\boldsymbol{\Sigma}}(\boldsymbol{S}; \tilde{\boldsymbol{\theta}}) = n \log|\boldsymbol{S}| + \sum_{i=1}^n \|\boldsymbol{X}_i - \boldsymbol{\mu}\|^2_{\boldsymbol{S}^{-1}} + \rho\sum_{i=1}^n \langle \boldsymbol{X}_i - \boldsymbol{\mu}, \tilde{\boldsymbol{\nu}} \rangle_{\boldsymbol{S}^{-1/2}} \|\boldsymbol{X}_i - \boldsymbol{\mu}\|_{\boldsymbol{S}^{-1}}$. When $t = 2 \bmod 4$, we set $\boldsymbol{\mu}^{(t)} = \boldsymbol{\mu}^{(t-1)}$, $\boldsymbol{\Sigma}^{(t)} = \boldsymbol{\Sigma}^{(t-1)}$, $\tilde{\boldsymbol{\nu}}^{(t)} = \tilde{\boldsymbol{\nu}}^{(t-1)}$, and make the update

$$\rho^{(t)} = \underset{r \in [0,1)}{\arg\min}\; q_{\rho}\left(r; \tilde{\boldsymbol{\theta}}^{(t-1)}\right), \tag{9}$$

with $q_{\rho}\left(r; \tilde{\boldsymbol{\theta}}\right) = -2n \log(c_d(r)) + r\bar{W}$, and $\bar{W} = \sum_{i=1}^n \langle \boldsymbol{X}_i - \boldsymbol{\mu}, \tilde{\boldsymbol{\nu}} \rangle_{\boldsymbol{\Sigma}^{-1/2}} \|\boldsymbol{X}_i - \boldsymbol{\mu}\|_{\boldsymbol{\Sigma}^{-1}}$. Lastly, when $t = 3 \bmod 4$, we set $\boldsymbol{\mu}^{(t)} = \boldsymbol{\mu}^{(t-1)}$, $\boldsymbol{\Sigma}^{(t)} = \boldsymbol{\Sigma}^{(t-1)}$, $\rho^{(t)} = \rho^{(t-1)}$, and make the update

$$\tilde{\boldsymbol{\nu}}^{(t)} = \boldsymbol{v}^{(t)}/\|\boldsymbol{v}^{(t)}\|, \tag{10}$$

where $\boldsymbol{v}^{(t)} = (\boldsymbol{\Sigma}^{(t-1)})^{-1/2} \sum_{i=1}^{n} \left\| \boldsymbol{X}_i - \boldsymbol{\mu}^{(t-1)} \right\|_{(\boldsymbol{\Sigma}^{(t-1)})^{-1}} \left(\boldsymbol{X}_i - \boldsymbol{\mu}^{(t-1)} \right)$. The updates (7)–(10) are conducted for each $t \leq T$, for some sufficiently large $T \in \mathbb{N}$. We then take the value of $\tilde{\boldsymbol{\theta}}^{(T)}$ to be the maximum likelihood estimator (MLE) that is obtained from the data $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$.

The updates for $\boldsymbol{\mu}$ and ρ are straight forward to compute using generally available optimization algorithms in Euclidean geometry and the update for $\tilde{\boldsymbol{\nu}}$ is obtained in closed-form. However, the update for $\boldsymbol{\Sigma}$ requires the use of optimization algorithms in Riemannian geometries. These algorithms have become mainstream via the works of Absil et al. (2009) and are openly implemented in software libraries such as Huang et al. (2018); Martin et al. (2020).

Other than the warranted descent of the negative log-likelihood as t increases, the algorithm that is directly defined via the updates (7)–(10) confer little in terms of theoretical guarantees. However using the theory of Boumal et al. (2019) and Tian et al. (2019), we can make the following modifications in order to confer our optimization process with a global convergence property.

Firstly, we must limit the domain of each of the optimization subproblems to compact spaces. In the cases of (7) and (9), this can be done with box constraints, whereas in the case of (8), we may impose lower and upper bounds on the eigenvalues of the SPD matrices (cf. Ritter, 2014, Sec. B.6.2). The entire set of compact constraints can be summarized by limiting our attention to the parameter space

$$\bar{\Theta} = \left\{ \tilde{\boldsymbol{\theta}} : \rho \in [0, 1-\gamma], \boldsymbol{\mu} \in [-m, m]^d, v_1 \leq \lambda_{(1)}(\boldsymbol{\Sigma}) \leq \lambda_{(d)}(\boldsymbol{\Sigma}) \leq v_d, \tilde{\boldsymbol{\nu}} \in \mathcal{S}^{d-1} \right\}, \quad (11)$$

where $\gamma \in (0,1)$, $m, v_1, v_d \in (0, \infty)$, and $\lambda_{(1)}$ and $\lambda_{(d)}$ are the smallest and largest eigenvalues of $\boldsymbol{\Sigma}$, respectively.

Secondly, we must utilize a trust-region algorithm of the variety from Absil et al. (2009, Ch. 7) to solve each of the subproblems. And thirdly, at each t, we assess each of the updates and only make the one that increases the likelihood by the largest amount.

Upon making these modifications, we may apply Theorem 3 of Tian et al. (2019) in order to obtain the convergence to zero for each of the Riemannian gradients corresponding to partial differentials with respect to the four components of $\tilde{\boldsymbol{\theta}}$.

If we define $\tilde{\boldsymbol{\theta}}_n$ as the MLE, restricted to the constraint set (11):

$$\tilde{\boldsymbol{\theta}}_n = \underset{\tilde{\boldsymbol{\theta}} \in \bar{\Theta}}{\arg\max}\, \ell_n(\tilde{\boldsymbol{\theta}}),$$

then we can establish the consistency of $\tilde{\boldsymbol{\theta}}_n$, via Lemma 4.2 of Pötscher and Prucha (2013), see Appendix A·1 for a proof.

PROPOSITION 10. *If* $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n \overset{\text{iid}}{\sim} \text{MED}(\tilde{\boldsymbol{\theta}})$*, then* $\tilde{\boldsymbol{\theta}}_n$ *is consistent in the sense that*

$$\inf_{\tilde{\boldsymbol{\theta}} \in \tilde{\Theta}_0} \|\tilde{\boldsymbol{\theta}}_n - \tilde{\boldsymbol{\theta}}\| \xrightarrow{P} 0,$$

as $n \to \infty$*, where*

$$\tilde{\Theta}_0 = \left\{ \tilde{\boldsymbol{\theta}}_0 \in \bar{\Theta} : E \log f_d(\boldsymbol{X}; \tilde{\boldsymbol{\theta}}_0) \geq \max_{\tilde{\boldsymbol{\theta}} \in \bar{\Theta}} E \log f_d(\boldsymbol{X}; \tilde{\boldsymbol{\theta}}) \right\}.$$

5·2. *Bayesian inference*

An interesting feature of Bayesian inference is that some expert knowledge can be incorporated in the prior distribution and that the posterior distribution can be used to provide uncertainty estimates with credible bands. This is complementing the previous section as well as the work of Herrmann et al. (2018).

We adopt the same parameterization as that of the previous section, which replaces $\boldsymbol{\nu}$ in $\boldsymbol{\theta}$, by $\tilde{\boldsymbol{\nu}} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\nu}$. Instead of the covariance matrix, it is more efficient from an implementation viewpoint to work with the precision matrix $\boldsymbol{\Omega} = \boldsymbol{\Sigma}^{-1} \in \mathrm{SPD}\,(d)$. A simple strategy consists in assigning independent priors on $\boldsymbol{\mu}$, $\boldsymbol{\Omega}$, ρ and $\tilde{\boldsymbol{\nu}}$. First, a normal-inverse-Wishart prior on $(\boldsymbol{\mu}, \boldsymbol{K})$ parameters, where $\boldsymbol{K}$ is the covariance matrix parameter of the normal prior for $\boldsymbol{\mu}$:

$$\begin{aligned} \boldsymbol{\mu} \mid \boldsymbol{K} &\sim \mathcal{N}(\boldsymbol{\mu}_0, \boldsymbol{K}), \\ \boldsymbol{K} &\sim \mathcal{IW}(\nu_K, \boldsymbol{S}_K), \end{aligned}$$

where $\boldsymbol{\mu}_0 \in \mathbb{R}^d$, $\nu_K \in (d-1, \infty)$, and $\boldsymbol{S}_K \in \mathrm{SPD}\,(d)$ can be fixed or be given hyperpriors. Second, a Wishart prior is placed on $\boldsymbol{\Omega}$:

$$\boldsymbol{\Omega} \sim \mathcal{W}(\nu_\Omega, \boldsymbol{S}_\Omega),$$

where $\nu_\Omega \in (d-1, \infty)$ and $\boldsymbol{S}_\Omega$ is symmetric and positive definite, and can be fixed or can be given hyperpriors. And finally, uniform priors are placed on ρ and $\tilde{\boldsymbol{\nu}}$:

$$\rho \sim \mathcal{U}(0,1), \quad \tilde{\boldsymbol{\nu}} \sim \mathcal{U}(\mathcal{S}^{d-1}).$$

Note that in the presence of prior information, more informative priors than the ones presented here could be considered in order to encode such prior knowledge.

Finally, the data generating process for observations $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ is the MED:

$$\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n \mid \boldsymbol{\mu}, \boldsymbol{\Omega}, \rho, \tilde{\boldsymbol{\nu}}, \boldsymbol{\varphi} \overset{\text{iid}}{\sim} \mathrm{MED}\big(\boldsymbol{\mu}, \boldsymbol{\Omega}^{-1}, \boldsymbol{\Omega}^{-1/2}\tilde{\boldsymbol{\nu}}, \rho\big),$$

where $\boldsymbol{\varphi}$ represents potential hyperparameters described above. We have implemented the model in the Stan probabilistic programming language (Carpenter et al., 2017), which uses Hamiltonian Monte Carlo. See Appendix B·2 for implementation details.

5·3. *The **MEDdist** R package*

We describe an R package containing the MLE and Bayesian procedures introduced above. Density, distribution function, quantile function and random generation for the MED are also documented. The package, termed **MEDdist**, is available at `https://github.com/AntoineUC/MEDdist`.

- `fitmed`: estimates the parameters $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, ρ and $\widetilde{\boldsymbol{\nu}}$ from a sample (input) using the MLE procedure of Section 5·1. The starting parameters $\boldsymbol{\mu}^{(0)}$, $\boldsymbol{\Sigma}^{(0)}$, $\rho^{(0)}$ and $\widetilde{\boldsymbol{\nu}}^{(0)}$ may be chosen as inputs, as well as the maximum number of iterations `maxiter`.
- `dmed`: returns the value of the p.d.f. of the MED with parameters $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, ρ and $\boldsymbol{\nu}$ (inputs) at point $\boldsymbol{x} \in \mathbb{R}^d$ (input) using Equation (2).
- `pmed`: returns the value of the c.d.f. (only in dimension $d = 1$) of the MED with parameters μ, Σ, ρ and ν (inputs) at point $x \in \mathbb{R}$ (input).
- `qmed`: returns the quantiles (only in dimension $d = 1$) of the MED with parameters μ, Σ, ρ and ν (inputs) of level $\tau \in (0,1)$ (input).
- `rmed`: returns a sample of size n (input) from the MED with parameters $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, ρ and $\boldsymbol{\nu}$ (inputs) using the bounds of Section 3. Indeed, the upper bound of Proposition 5 leads to a simple rejection algorithm, where the proposal distribution is multivariate Gaussian. The acceptance probability, $M_d(\rho)^{-1}$, deteriorates as ρ increases to 1 and as the dimension d increases. Some potential refinements for the proposal distribution taking the form of mixtures of truncated multivariate Gaussian distributions are not investigated here, but are left to future work.

Table 1. *Means (standard font) and standard deviations (italic) of the elements from a random sample of* 100 *MLEs, for each of the simulation scenarios S1–S8. Note that the circular means and standard deviations are reported for* $\phi^{(T)}$

n		$\mu_1^{(T)}$	$\mu_2^{(T)}$	$\Sigma_{11}^{(T)}$	$\Sigma_{12}^{(T)}$	$\Sigma_{22}^{(T)}$	$\rho^{(T)}$	$\phi^{(T)}$
100	**S1**	0.067	0.060	0.919	-0.023	0.906	0.418	0.688
		0.342	*0.309*	*0.196*	*0.108*	*0.173*	*0.215*	*1.183*
	S2	0.096	0.136	0.946	0.424	0.890	0.402	0.955
		0.270	*0.268*	*0.142*	*0.114*	*0.147*	*0.172*	*0.979*
	S3	0.279	0.225	1.090	-0.208	1.077	0.851	0.765
		0.266	*0.373*	*0.246*	*0.169*	*0.342*	*0.071*	*0.244*
	S4	0.262	0.186	1.046	0.391	1.097	0.823	0.725
		0.210	*0.273*	*0.194*	*0.185*	*0.258*	*0.041*	*0.238*
1,000	**S5**	0.088	0.074	1.010	-0.027	1.012	0.288	0.771
		0.110	*0.110*	*0.059*	*0.035*	*0.051*	*0.089*	*0.475*
	S6	0.110	0.097	1.002	0.475	0.998	0.282	0.722
		0.089	*0.100*	*0.055*	*0.042*	*0.056*	*0.083*	*0.365*
	S7	0.290	0.294	1.097	-0.249	1.089	0.856	0.787
		0.094	*0.116*	*0.082*	*0.056*	*0.116*	*0.026*	*0.071*
	S8	0.275	0.266	1.082	0.385	1.084	0.827	0.784
		0.070	*0.079*	*0.071*	*0.067*	*0.093*	*0.019*	*0.056*

- `bayexpectile` and `baymode`: return posterior mean estimators for multivariate expectile and mode, respectively, under a sample from the MED. These functions, containing Stan codes including prior specification, are provided in Appendix B·2, and detailed in Section 7.

6. MLE SIMULATION STUDY

In order to assess the effectiveness of the MLE procedure, we conducted a number of simulation studies in R (R Core Team, 2019). Within these studies, we used the manifold optimization routines provided in the **ManifoldOptim** (Martin et al., 2020) package.

In total, eight simulation scenarios were considered. In the first scenario (S1), we computed the MLE $\tilde{\boldsymbol{\theta}}^{(T)}$ from a data set of $n = 100$ observations generated from the MED, where $d = 2$, $\boldsymbol{\mu}^\top = (0, 0)$, $\boldsymbol{\Sigma} = \boldsymbol{I}_2$, $\rho = 0.2$, and $\tilde{\boldsymbol{\nu}}^\top = (1, 1)/\sqrt{2}$. In the second scenario (S2), the data are generated in the same manner as in S1, except that $\boldsymbol{\Sigma} = (\boldsymbol{I}_2 + \boldsymbol{J}_2)/2$, where $\boldsymbol{J}_2$ is the 2×2 matrix with a value of 1 in all entries.

Scenarios S3 and S4 are identical to scenarios S1 and S2, respectively, except that $\rho = 0.8$. Scenarios S5–S8 are identical to scenarios S1–S4, respectively, except that $n = 1,000$. We repeated each scenario 100 times and present the mean and standard deviation of the sample distribution for each of the elements of $\big(\mu_1^{(T)}, \mu_2^{(T)}, \Sigma_{11}^{(T)}, \Sigma_{12}^{(T)}, \Sigma_{22}^{(T)}, \rho^{(T)}, \phi^{(T)}\big)$, for each case, in Table 1, where $\phi^{(T)} = \arctan(\tilde{\nu}_2^{(T)}/\tilde{\nu}_1^{(T)})$ is the polar angle of the point $\tilde{\boldsymbol{\nu}}^{(T)}$. Here, the mean and standard deviation of $\phi^{(T)}$ are reported as circular mean and standard deviation. Note that the polar angle of $\tilde{\boldsymbol{\nu}}$ is $\phi = \pi/4 \approx 0.785$ (cf. Mardia and Jupp, 2000, Ch. 2).

From Table 1, we observe that in each scenario, the true value of the parameter elements are well within a standard deviation of the mean, of each of the sampling distributions. Furthermore, the variances from Scenarios S5–S8 are much smaller than those from scenarios S1–S4. These

two aspects provide evidence towards the fact that the MLE is consistent; see Proposition 10. Thus the MLE appears to be performing as expected.

In the specific cases of the simulations that were assessed, it appears that for small sample sizes, the estimates for a small population value of ρ may be severely upwardly biased, and that the direction ϕ may also be biased in such cases (see, the results for S1 and S2). We note that these biases are somewhat reduced when the population value of ρ is larger, as in S3 and S4.

Furthermore, we note that the bias is not nearly as severe when we increase the sample size from $n = 100$ to $n = 1,000$ (see, the results of S5 and S6). However, a smaller value of ρ still appears to lead to higher variation of the MLE for ρ and for ϕ, as can be seen when comparing to the results of S7 and S8. Finally, we note that the size of ρ also appears to influence the variability of the MLE of the elements of $\boldsymbol{\Sigma}$, although the MLE of $\boldsymbol{\mu}$ appears to be generally unaffected.

7. MULTIVARIATE APPLICATIONS

7·1. *Bayesian estimation of expectiles*

In this section, we estimate geometric expectiles in the Bayesian framework described in Section 5·2. We adopt the bivariate simulation experiments of Herrmann et al. (2018). More specifically, we consider four bivariate random variables $\boldsymbol{X}^{(1)}, \ldots, \boldsymbol{X}^{(4)}$ (See Table 5 in the Appendix for a description of the data distributions).

We are interested in geometric expectiles $\mathbf{e}_{\rho,\boldsymbol{\nu}(\phi),\boldsymbol{I}_d}(\boldsymbol{X}^{(i)})$, where $\rho = 0.98$ and the whole spectrum $\boldsymbol{\nu}(\phi) = (\cos(\phi), \sin(\phi))$ with $\phi \in [0, 2\pi)$ is considered. According to the Bayesian model of Section 5·2, estimation of such expectiles is carried out by simply setting the parameters $(\rho, \boldsymbol{\nu}, \boldsymbol{\Sigma})$ to fixed desired values $(0.98, \boldsymbol{\nu}(\phi), \boldsymbol{I}_d)$ (i.e. no prior is set) and by estimating the expectile through parameter $\boldsymbol{\mu}$.

We sampled $n = 1,000$ observations for each of the four distribution settings. For each value of ϕ in the grid $\{\phi_\ell = 2\pi(\ell - 1)/100, \ell = 1, \ldots, 100\}$, we used Stan (code described in Appendix B·2) with $2,000$ HMC iterations after a $1,000$-iteration burn-in. Stan provides a $\hat{R}$ value of 1 for all of the chains, thus not indicating convergence issues. Expectiles were estimated by the posterior mean. For each value of ϕ, credible bands could be computed as those regions containing a prespecified proportion of iterations from the HMC chain (see, for instance, Section 7·2). We have followed a different approach in order to better visualize the credible regions as bands around the estimated expectiles. More specifically, we used multivariate quantiles (in the sense of Chaudhuri, 1996) of orders 2.5% and 97.5%, to get pointwise credible bounds for each value of ϕ_ℓ, and then connected them with a curve visible as blue dashed lines on Figure 2. Expectiles were also estimated using the maximum likelihood approach of Herrmann et al. (2018). For each of the four distribution settings, Figure 2 displays contour plots of the true distribution, and the Bayesian and MLE estimates in solid blue and red lines, respectively. Eight values for ϕ defined as $\varphi_j = j\pi/8$, $j = 0, \ldots, 7$, are marked along the expectile curves.

We can see from Figure 2 that Bayesian estimates of the expectiles are almost indistinguishable from MLE estimates. The main asset of our Bayesian approach is the availability of uncertainty quantification of these estimates.

7·2. *Bayesian estimation of the mode of a distribution*

We take advantage of the property that the MED mode coincides with its location parameter $\boldsymbol{\mu}$ (Proposition 3 (i)) to propose a procedure for estimating the mode of a distribution.

Consider the four distribution settings of the previous section. For each of them, we generated n observations, $n \in \{100; 500; 1,000\}$. We used the Bayesian model of Section 5·2 implemented in Stan (code described in Appendix B·2) with $2,000$ HMC iterations after a $1,000$-

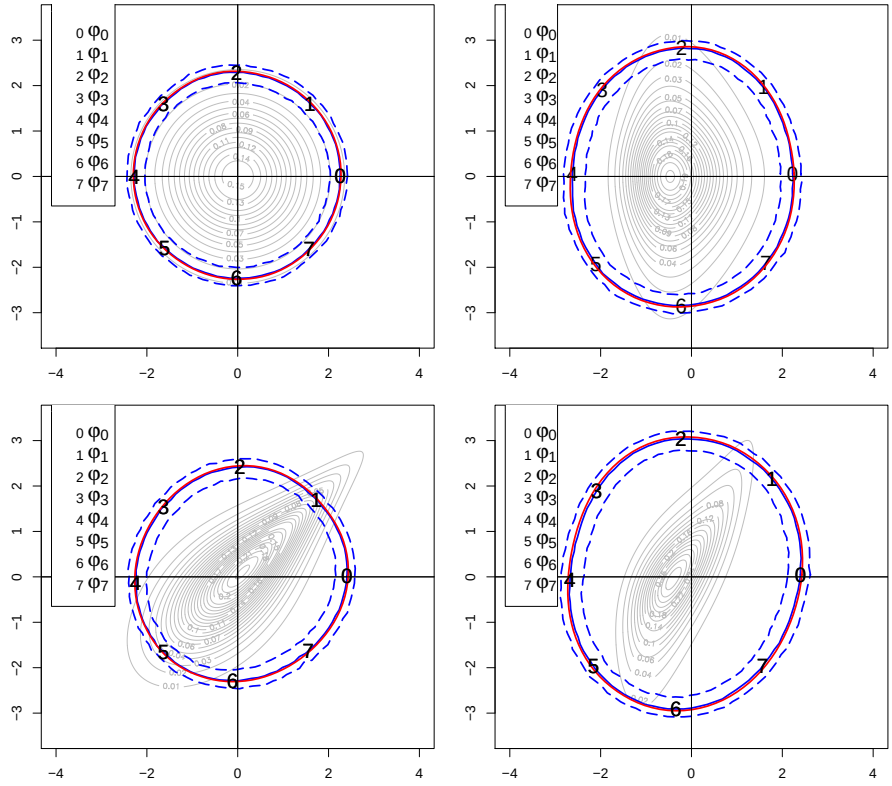

Fig. 2. Red: Empirical estimates of expectiles. Blue: Bayesian expectile estimates (solid lines, superimposed with red curves) and credible bands (dashed lines). Grey: level curves of the densities.

iteration burn-in. Stan provides a $\hat{R}$ value of 1 for all of the chains, thus non indicating any convergence issue. As a competitor to the MED, we used a multivariate Gaussian distribution as a fitting distribution. The Gaussian distribution can be simply obtained from the MED by fixing the ρ parameter to 0, see Proposition 2 (i) (this automatically makes the $\boldsymbol{\nu}$ parameter deprecated). The Gaussian distribution also satisfies the property that the mode is available in closed-form (as its mean). For both cases, expectiles were estimated by the posterior mean. In contrast to Section 7·1, 95% credible regions were provided by sets containing 95% of the $2,000$ HMC iterations. These were constructed by using the `dataEllipse` R function from package **car**, which provides the smallest ellipse enveloping the desired proportion of points (95%). The whole procedure was repeated 100 times, leading to Table 2 which displays the mean squared error of the estimates over the 100 repetitions (MED / Gaussian distribution).

Figure 3 represents the results for one of the 100 repetitions obtained with $n = 1,000$ data points. Each panel corresponds to one of the four distributions, the red cross indicates the true mode, a blue (resp. green) cross the estimate from the MED (Gaussian distribution). The ellipses correspond to 95% credible intervals. The results are in accordance to those of Table 2, illustrating that the approach based on the MED provides better calibrated credible intervals than those based on the Gaussian distribution.

Table 2. *Mean squared errors computed on* 100 *repetitions of the Bayesian mode estimator (based on the MED / Gaussian distribution) for different sample sizes* n

n	$\boldsymbol{X}^{(1)}$	$\boldsymbol{X}^{(2)}$	$\boldsymbol{X}^{(3)}$	$\boldsymbol{X}^{(4)}$
100	0.0588 / 0.0204	0.1054 / 0.0549	0.0553 / 0.0358	0.1139 / 0.0321
500	0.0161 / 0.0037	0.0391 / 0.0408	0.0215 / 0.0253	0.0498 / 0.0188
1,000	0.0086 / 0.0020	0.0347 / 0.0362	0.0109 / 0.0218	0.0352 / 0.0175

Table 3. *Empirical coverage probabilities (i.e. the posterior probability that true mode belongs to the credible 95% ellipse) mesured on* 100 *repetitions of the Bayesian mode estimator (based on the MED / Gaussian distribution) for different sample sizes* n

n	$\boldsymbol{X}^{(1)}$	$\boldsymbol{X}^{(2)}$	$\boldsymbol{X}^{(3)}$	$\boldsymbol{X}^{(4)}$
100	0.99 / 0.95	0.97 / 0.35	1.00 / 0.89	0.94 / 0.61
500	1.00 / 0.96	0.87 / 0.00	0.98 / 0.40	0.79 / 0.01
1,000	1.00 / 0.97	0.74 / 0.00	0.99 / 0.08	0.61 / 0.00

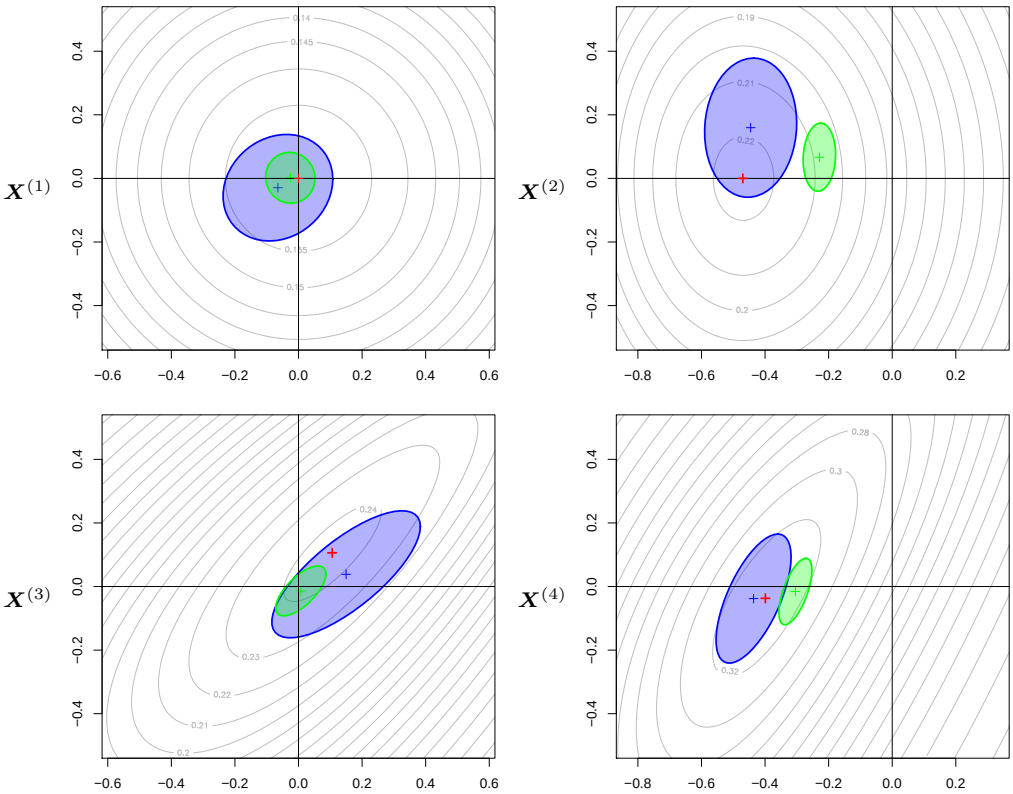

Fig. 3. Red, blue and green crosses: theoretical mode, and mode estimates based on the MED and Gaussian distributions, respectively. Blue and green areas: credible ellipses of the mode based on the MED and Gaussian distributions, respectively. Grey: level curves of the densities.

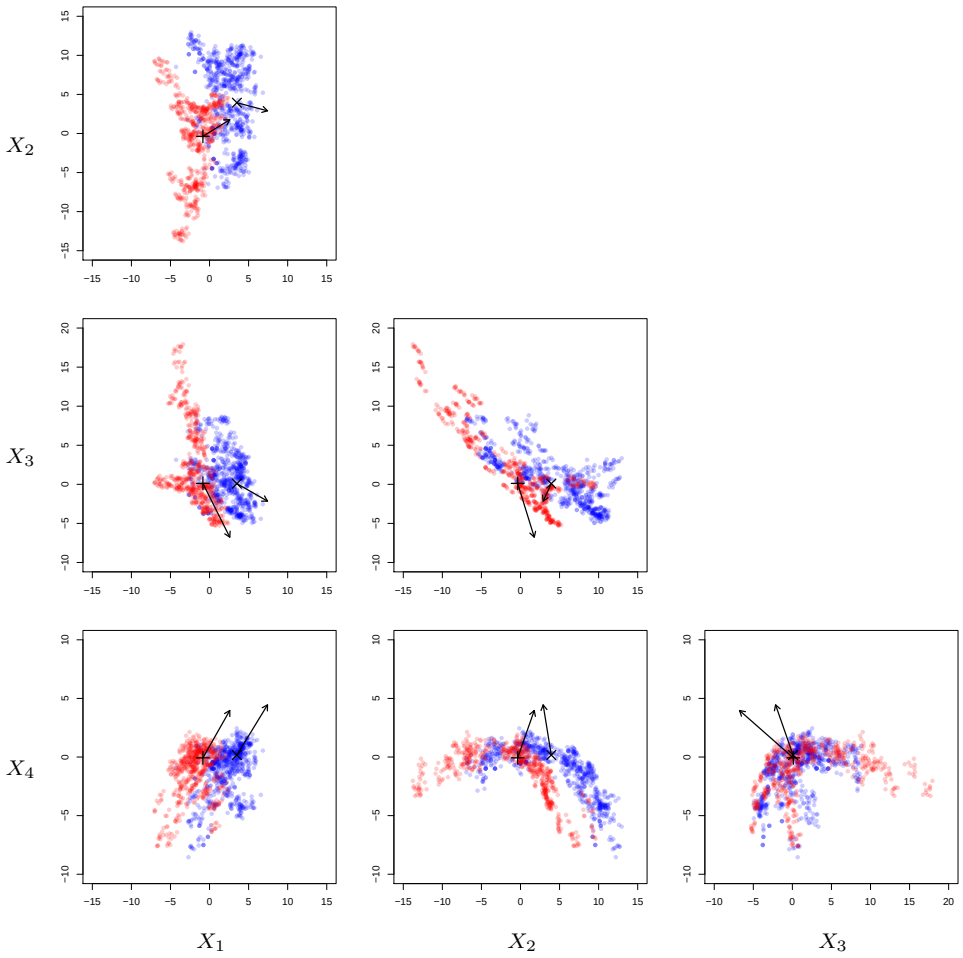

Fig. 4. Banknote authentication data set. Covariates $(X_1, \ldots, X_4)$ in blue (genuine) and in red (forged) as bivariate scatterplots. Expectile $\boldsymbol{\mu}$ estimated with the MED is reported as "$\times$" (genuine) and "$+$" (forged); direction $\boldsymbol{\nu}$ is represented as an arrow originating from the estimated expectile.

7·3. *Discriminant analysis*

We propose a discriminant analysis on real data, by considering the Banknote authentication data set, available at `https://archive.ics.uci.edu/ml/datasets/banknote+authentication`. The data of size $n = 1,372$ contain four covariates ($X_1 =$ variance, $X_2 =$ skewness, $X_3 =$ kurtosis and $X_4 =$ entropy of wavelet transformed banknote images), and a binary variable indicating whether the banknote is genuine (0) or forged (1). As illustrated in Figure 4, the genuine (blue) vs forged (red) balance in the data set is 56% vs 44%.

We randomly split the data set in a train sample of size $n = 1,000$ and a test sample of size 372, and repeat 100 times the experiment. For each experiment, we thus have two classes of records. The aim of discriminant analysis is to attribute a class for each record of the test sample. A usual and widespread method is to fit the train sample with Gaussian distributions (one for each class), and attributing observations of the test sample according to the maximum a posteriori rule

Table 4. *Median classification errors computed on the test sample, their standard deviations (SD), their median absolute deviations (MAD) and average AIC and BIC for genuine / forged classes*

	% of error	SD	MAD	AIC	BIC
Gaussian	1.34	0.693	0.538	9633 / 7633	9694 / 7691
Skew normal	1.34	0.676	0.538	9501 / **7347**	9579 / **7421**
MED	1.34	**0.611**	**0.269**	**9483** / 7509	**9565** / 7587

(see Hastie and Tibshirani, 1996). In addition to resorting to the usual Gaussian distribution, we also considered here the MED and the skew normal ($\mathcal{SN}$) distribution (Azzalini and Capitanio, 1999), using the R function `msn.mle` from package **sn**. Since the Gaussian distribution is a special case of the MED, the latter is supposed to be more flexible, and provide a better fit. Maximum likelihood estimators of the MED parameters $(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \rho, \boldsymbol{\nu})$ are computed for both classes and reported in Table 6 in Appendix B·3. Estimated expectiles $\boldsymbol{\mu}$ and asymmetry directions $\boldsymbol{\nu}$ are also reported in Figure 4.

The results comparing the three approaches are reported in Table 4. We note that, in comparison with the Gaussian (and skew normal) mixture approaches, the MED has the same error rate, with lower standard deviations and median absolute deviations, and seems to better fit (in terms of AIC) the distribution of genuine and forged banknotes, illustrating its flexibility.

ACKNOWLEDGEMENTS

We would like to thank Mario Beraha for improving the Stan implementation of our Bayesian model.

REFERENCES

Abdous, B. and Remillard, B. (1995). Relating quantiles and expectiles under weighted-symmetry. *Annals of the Institute of Statistical Mathematics*, 47:371–384.

Absil, P.-A., Mahony, R., and Sepulchre, R. (2009). *Optimization algorithms on matrix manifolds*. Princeton University Press.

Andrews, D. W. (1992). Generic uniform convergence. *Econometric Theory*, 8:241–257.

Azzalini, A. and Capitanio, A. (1999). Statistical applications of the multivariate skew normal distribution. *Journal of the Royal Statistical Society: Series B*, 61(3):579–602.

Beck, N., Di Bernardino, E., and Mailhot, M. (2021). Semi-parametric estimation of multivariate extreme expectiles. *Journal of Multivariate Analysis*, 184:104758.

Boumal, N., Absil, P.-A., and Cartis, C. (2019). Global rates of convergence for nonconvex optimization on manifolds. *IMA Journal of Numerical Analysis*, 39(1):1–33.

Brascamp, H. J. and Lieb, E. H. (2002). On extensions of the Brunn-Minkowski and Prékopa-Leindler theorems, including inequalities for log concave functions, and with an application to the diffusion equation. In *Inequalities*, pages 441–464. Springer.

Cambanis, S., Huang, S., and Simons, G. (1981). On the theory of elliptically contoured distributions. *Journal of Multivariate Analysis*, 11(3):368–385.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., and Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, 76(1):1–32.

Chaudhuri, P. (1996). On a geometric notion of quantiles for multivariate data. *Journal of the American Statistical Association*, 91(434):862–872.

Chen, S. and Huang, J. (2014). Rates of convergence of extreme for asymmetric normal distribution. *Statistics & Probability Letters*, 84:158–168.

Daouia, A., Girard, S., and Stupfler, G. (2018). Estimation of tail risk based on extreme expectiles. *Journal of the Royal Statistical Society: Series B*, 80:263–292.

Daouia, A. and Paindaveine, D. (2019). From halfspace m-depth to multiple-output expectile regression. *arXiv preprint arXiv:1905.12718*.

Fu, M. and Zhou, W. (2016). Non-rigid point set registration via mixture of asymmetric Gaussians with integrated local structures. In *2016 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, pages 999–1004.

Girard, S., Stupfler, G., and Usseglio-Carleve, A. (2021). Extreme conditional expectile estimation in heavy-tailed heteroscedastic regression models. *The Annals of Statistics*. to appear.

Gradshteyn, I. S. and Ryzhik, I. M. (2014). *Table of integrals, series, and products*. Academic press.

Hastie, T. and Tibshirani, R. (1996). Discriminant analysis by Gaussian mixtures. *Journal of the Royal Statistical Society: Series B*, 58(1):155–176.

He, W., Yu, R., Zheng, Y., and Jiang, T. (2018). Image denoising using asymmetric Gaussian mixture models. In *2018 International Symposium in Sensing and Instrumentation in IoT Era (ISSI)*, pages 1–4.

Herrmann, K., Hofert, M., and Mailhot, M. (2018). Multivariate geometric expectiles. *Scandinavian Actuarial Journal*, 2018(7):629–659.

Holzmann, H. and Klar, B. (2016). Expectile asymptotics. *Electronic Journal of Statistics*, 10:2355–2371.

Huang, W., Absil, P.-A., Gallivan, K. A., and Hand, P. (2018). ROPTLIB: an object-oriented C++ library for optimization on Riemannian manifolds. *ACM Transactions on Mathematical Software (TOMS)*, 44(4):1–21.

Kato, T., Omachi, S., and Aso, H. (2002). Asymmetric Gaussian and its application to pattern recognition. In *Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition (SPR) and Structural and Syntactic Pattern Recognition (SSPR)*, pages 405–413. Springer.

Krätschmer, V. and Zähle, H. (2017). Statistical inference for expectile-based risk measures. *Scandinavian Journal of Statistics*, 44:425–454.

Kuan, C.-M., Yeh, J. H., and Hsu, Y.-C. (2009). Assessing value at risk with CARE, the Conditional Autoregressive Expectile models. *Journal of Econometrics*, 150:261–270.

Kume, A. and Wood, A. T. (2005). Saddlepoint approximations for the Bingham and Fisher–Bingham normalising constants. *Biometrika*, 92(2):465–476.

Majumdar, A. and Paul, D. (2016). Zero expectile processes and Bayesian spatial regression. *Journal of Computational and Graphical Statistics*, 25(3):727–747.

Mardia, K. V. and Jupp, P. E. (2000). *Directional Statistics*. John Wiley and Sons, New York.

Martin, S., Raim, A., Huang, W., and Adragni, K. (2020). ManifoldOptim: An R Interface to the ROPTLIB Library for Riemannian Manifold Optimization. *Journal of Statistical Software*, 93(1):1–32.

Maume-Deschamps, V., Rullière, D., and Said, K. (2017). Multivariate extensions of expectiles risk measures. *Dependence Modeling*, 5(1):20–44.

Maume-Deschamps, V., Rullière, D., and Said, K. (2018). Extremes for multivariate expectiles. *Statistics & Risk Modeling*, 35(3-4):111–140.

Nelsen, R. (2007). *An introduction to copulas*. Springer Science & Business Media.

Newey, W. K. and Powell, J. L. (1987). Asymmetric least squares estimation and testing. *Econometrica*, 55:819–847.

Pötscher, B. M. and Prucha, I. R. (2013). *Dynamic nonlinear econometric models: Asymptotic theory*. Springer Science & Business Media.

R Core Team (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Ritter, G. (2014). *Robust cluster analysis and variable selection*. Chapman and Hall/CRC.

Tian, Y., Khosoussi, K., and How, J. P. (2019). Block-Coordinate Descent on the Riemannian Staircase for Certifiably Correct Distributed Rotation and Pose Synchronization. *arXiv preprint arXiv:1911.03721*.

Vladimirova, M., Girard, S., Nguyen, H. D., and Arbel, J. (2020). Sub-Weibull distributions: generalizing sub-Gaussian and sub-Exponential properties to heavier-tailed distributions. *Stat*, 9:e318.

Ziegel, J. F. (2016). Coherence and elicitability. *Mathematical Finance*, 26:901–918.

A. Appendix: Proofs

A·1. *Proof of main results*

Proof of results from Section 2

Remarking that the optimization problem $\min_{\boldsymbol{x}\in\mathbb{R}^d} E\left(\Lambda_{\rho,\boldsymbol{\nu},\boldsymbol{\Sigma}}(\boldsymbol{X}-\boldsymbol{x})\right)$ is equivalent to $\min_{\boldsymbol{x}\in\mathbb{R}^d} E\left(\Lambda_{\rho,\boldsymbol{\nu},\boldsymbol{\Sigma}}(\boldsymbol{X}-\boldsymbol{x})-\Lambda_{\rho,\boldsymbol{\nu},\boldsymbol{\Sigma}}(\boldsymbol{X})\right)$, the following Proposition provides a sufficient condition for the well-posedness of (1).

Proposition A1. *Let $\rho\in[0,1)$, $\boldsymbol{\Sigma}\in\mathrm{SPD}(d)$ and $\boldsymbol{\nu}\in\mathbb{R}^d$ such that $\|\boldsymbol{\nu}\|_{\boldsymbol{\Sigma}^{-1}}=1$.*
If $E(\|\boldsymbol{X}\|)_{\boldsymbol{\Sigma}^{-1}}<\infty$, then $E(|\Lambda_{\rho,\boldsymbol{\nu},\boldsymbol{\Sigma}}(\boldsymbol{X}-\boldsymbol{x})-\Lambda_{\rho,\boldsymbol{\nu},\boldsymbol{\Sigma}}(\boldsymbol{X})|)<\infty$.

Proof of Proposition A1. Let us consider the expansion

$$\begin{aligned}\Lambda_{\rho,\boldsymbol{\nu},\boldsymbol{\Sigma}}(\boldsymbol{X}-\boldsymbol{x}) - \Lambda_{\rho,\boldsymbol{\nu},\boldsymbol{\Sigma}}(\boldsymbol{X}) &= \|\boldsymbol{X}-\boldsymbol{x}\|^2_{\boldsymbol{\Sigma}^{-1}} - \|\boldsymbol{X}\|^2_{\boldsymbol{\Sigma}^{-1}} \\ &+ \rho\|\boldsymbol{X}-\boldsymbol{x}\|_{\boldsymbol{\Sigma}^{-1}}\left(\langle \boldsymbol{X}-\boldsymbol{x},\boldsymbol{\nu}\rangle_{\boldsymbol{\Sigma}^{-1}} - \langle \boldsymbol{X},\boldsymbol{\nu}\rangle_{\boldsymbol{\Sigma}^{-1}}\right) \\ &+ \rho\langle \boldsymbol{X},\boldsymbol{\nu}\rangle_{\boldsymbol{\Sigma}^{-1}}\left(\|\boldsymbol{X}-\boldsymbol{x}\|_{\boldsymbol{\Sigma}^{-1}} - \|\boldsymbol{X}\|_{\boldsymbol{\Sigma}^{-1}}\right) \\ &:= T_1 + T_2 + T_3,\end{aligned}$$

and consider the three terms separately:

$$\begin{aligned}|T_1| &= |\langle \boldsymbol{x}, 2\boldsymbol{X}-\boldsymbol{x}\rangle_{\boldsymbol{\Sigma}^{-1}}| \le \|\boldsymbol{x}\|_{\boldsymbol{\Sigma}^{-1}}(2\|\boldsymbol{X}\|_{\boldsymbol{\Sigma}^{-1}} + \|\boldsymbol{x}\|_{\boldsymbol{\Sigma}^{-1}}),\\ |T_2| &= \rho\|\boldsymbol{X}-\boldsymbol{x}\|_{\boldsymbol{\Sigma}^{-1}}|\langle \boldsymbol{x},\boldsymbol{\nu}\rangle_{\boldsymbol{\Sigma}^{-1}}| \le (\|\boldsymbol{X}\|_{\boldsymbol{\Sigma}^{-1}} + \|\boldsymbol{x}\|_{\boldsymbol{\Sigma}^{-1}})\|\boldsymbol{x}\|_{\boldsymbol{\Sigma}^{-1}},\\ |T_3| &\le \|\boldsymbol{X}\|_{\boldsymbol{\Sigma}^{-1}}\|\boldsymbol{x}\|_{\boldsymbol{\Sigma}^{-1}},\end{aligned}$$

by repeated uses of Cauchy–Schwarz and triangle inequalities. The conclusion follows. □

PROPOSITION A2 (STRICT CONVEXITY OF $\Lambda_{\rho,\boldsymbol{\nu},\boldsymbol{\Sigma}}$). *For any $\rho \in [0,1)$, for any matrix $\boldsymbol{\Sigma} \in \mathrm{SPD}(d)$ and $\boldsymbol{\nu} \in \mathbb{R}^d$ such that $\|\boldsymbol{\nu}\|_{\boldsymbol{\Sigma}^{-1}} = 1$, $\Lambda_{\rho,\boldsymbol{\nu},\boldsymbol{\Sigma}}$ is strictly convex on $\mathbb{R}^d$.*

Proof of Proposition A2. Note that convexity is a property that is preserved by translation and rotation, so that it suffices to prove the result for the standardized MED. Denoting by $x_1 = \langle \boldsymbol{x}, \boldsymbol{u}_1\rangle$ the first component of $\boldsymbol{x}$, and by $\boldsymbol{x}_{-1}$ the other $d-1$ entries, we have

$$\Lambda_\rho(\boldsymbol{x}) = \|\boldsymbol{x}\|^2 + \rho x_1\|\boldsymbol{x}\| = x_1^2 + \|\boldsymbol{x}_{-1}\|^2 + \rho x_1\sqrt{x_1^2+\|\boldsymbol{x}_{-1}\|^2} = \tilde{\Lambda}_\rho(x_1, \|\boldsymbol{x}_{-1}\|), \tag{A1}$$

where function $\tilde{\Lambda}_\rho : \mathbb{R}^2 \to \mathbb{R}_+$, $(x,y) \mapsto x^2+y^2+\rho x\sqrt{x^2+y^2}$ coincides with Λ_ρ in the bivariate case ($d=2$). Note that it is strictly convex as the Hessian determinant and trace are both positive, equal respectively to $3(\rho x/\|\boldsymbol{x}\|+1)^2 + 1 - \rho^2$ and $3\rho x/\|\boldsymbol{x}\| + 4$, with $\boldsymbol{x} = (x,y)$. Also, the function of the second component $y \mapsto \tilde{\Lambda}_\rho(x,y)$ has partial derivative $y \mapsto y(\rho x/\|\boldsymbol{x}\|+1)$, meaning that it is non-decreasing on $\mathbb{R}_+$ for any $x \in \mathbb{R}$. As a result, we have that for any $\lambda \in (0,1)$, and for any $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$,

$$\begin{aligned}\Lambda_\rho(\lambda\boldsymbol{x} + (1-\lambda)\boldsymbol{y}) &\overset{(a)}{=} \tilde{\Lambda}_\rho\left(\lambda x_1 + (1-\lambda)y_1, \|\lambda\boldsymbol{x}_{-1} + (1-\lambda)\boldsymbol{y}_{-1}\|\right)\\ &\overset{(b)}{\le} \tilde{\Lambda}_\rho(\lambda x_1 + (1-\lambda)y_1, \lambda\|\boldsymbol{x}_{-1}\| + (1-\lambda)\|\boldsymbol{y}_{-1}\|)\\ &\overset{(c)}{<} \lambda\tilde{\Lambda}_\rho(x_1, \|\boldsymbol{x}_{-1}\|) + (1-\lambda)\tilde{\Lambda}_\rho(y_1, \|\boldsymbol{y}_{-1}\|)\\ &\overset{(a)}{=} \lambda\Lambda_\rho(\boldsymbol{x}) + (1-\lambda)\Lambda_\rho(\boldsymbol{y}),\end{aligned}$$

where both equalities (a) are by (A1), (b) is by combining the above non-decreasing property of $y \in \mathbb{R}_+ \mapsto \tilde{\Lambda}_\rho(x_1, y)$ with the convexity of the Euclidean norm $\|\cdot\|$, and (c) is by the strict convexity of $\tilde{\Lambda}$. This concludes the proof. □

Proof of Proposition 2. Part (i) is obvious. For part (ii), the standard MED random variable $\boldsymbol{Z} = (Z_1, \boldsymbol{Z}_{-1})$ has density

$$f_d(\boldsymbol{z};\rho) = \frac{c_d(\rho)}{(2\pi)^{d/2}}\exp\left(-\frac{z_1^2}{2} - \frac{\|\boldsymbol{z}_{-1}\|^2}{2} - \frac{\rho z_1}{2}\sqrt{z_1^2+\|\boldsymbol{z}_{-1}\|^2}\right)$$

The density of the transformed random variable $\boldsymbol{\Delta}_\rho\boldsymbol{Z} = (\sqrt{1-\rho}Z_1, \boldsymbol{Z}_{-1}/\sqrt{2})$ is given by $2^{(d-1)/2}(1-\rho)^{-1/2}f_d(z_1/\sqrt{1-\rho}, \sqrt{2}\boldsymbol{z}_{-1};\rho)$; that is:

$$\begin{aligned}&\frac{2^{(d-1)/2}c_d(\rho)}{(2\pi)^{d/2}(1-\rho)^{1/2}}\exp\left(-\frac{z_1^2}{2(1-\rho)} - \|\boldsymbol{z}_{-1}\|^2 - \frac{\rho z_1}{2\sqrt{1-\rho}}\sqrt{\frac{z_1^2}{1-\rho}+2\|\boldsymbol{z}_{-1}\|^2}\right)\\ &= \frac{2^{(d-1)/2}c_d(\rho)}{(2\pi)^{d/2}(1-\rho)^{1/2}}\exp\left(-|\boldsymbol{z}_{-1}\|^2 - \frac{z_1^2}{2(1-\rho)}\left[1+\rho\mathrm{sign}(z_1)\sqrt{1+\frac{2(1-\rho)\|\boldsymbol{z}_{-1}\|^2}{z_1^2}}\right]\right)\end{aligned}$$

$$= \frac{2^{(d-1)/2} c_d(\rho)}{(2\pi)^{d/2}(1-\rho)^{1/2}} \exp\left(-\|\boldsymbol{z}_{-1}\|^2 - \frac{z_1^2}{2(1-\rho)}\left[1+\rho\text{sign}(z_1)\left(1+\frac{(1-\rho)\|\boldsymbol{z}_{-1}\|^2}{z_1^2}+o(1-\rho)\right)\right]\right)$$
$$\underset{\rho\to 1}{\approx} \frac{2}{(2\pi)^{d/2}} \exp\left(-\frac{\|\boldsymbol{z}_{-1}\|^2}{2} - \frac{1+\rho\text{sign}(z_1)}{(1-\rho)}\frac{z_1^2}{2}\right)$$
$$\underset{\rho\to 1}{\longrightarrow} \frac{2}{(2\pi)^{d/2}} \exp\left(-\frac{\|\boldsymbol{z}_{-1}\|^2}{2} - \frac{z_1^2}{2}\right)\mathbb{I}\{z_1<0\}.$$

This corresponds to a half-standard Gaussian restricted to negative values of the first component. Lastly, for part (iii) with $d=1$, consider without loss of generality, the $\text{MED}(\mu,\sigma^2,1,\rho)$. Its density is made of two half-Gaussians on $(-\infty,\mu]$ and on $[\mu,\infty)$ with respective variances $\sigma_-^2 = \frac{\sigma^2}{1-\rho}$ and $\sigma_+^2 = \frac{\sigma^2}{1+\rho}$. This corresponds to the one-dimensional skew-Gaussian distribution of Kato et al. (2002), shifted by μ, with variance σ_+^2 (for the right tail) and ratio of variances $r^2 = \frac{1+\rho}{1-\rho}$. □

Proof of Proposition 3. (i) The mode(s) of the $\text{MED}(\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\nu},\rho)$ is (are) equivalently defined by

$$\underset{\boldsymbol{x}\in\mathbb{R}^d}{\arg\min}\,\Lambda_{\rho,\boldsymbol{\nu},\boldsymbol{\Sigma}}(\boldsymbol{x}-\boldsymbol{\mu}) = \boldsymbol{\mu} + \underset{\boldsymbol{y}\in\mathbb{R}^d}{\arg\min}\,\Lambda_{\rho,\boldsymbol{\nu},\boldsymbol{\Sigma}}(\boldsymbol{y}).$$

From Proposition A2, $\Lambda_{\rho,\boldsymbol{\nu},\boldsymbol{\Sigma}}$ is strictly convex on $\mathbb{R}^d$ and therefore the mode is unique. Moreover, $\Lambda_{\rho,\boldsymbol{\nu},\boldsymbol{\Sigma}}(\mathbf{0}) = 0$ and $\Lambda_{\rho,\boldsymbol{\nu},\boldsymbol{\Sigma}}(\boldsymbol{y}) \geq 0$, for any $\boldsymbol{y}\in\mathbb{R}^d$, in view of Cauchy–Schwarz inequality. It follows that

$$\underset{\boldsymbol{y}\in\mathbb{R}^d}{\arg\min}\,\Lambda_{\rho,\boldsymbol{\nu},\boldsymbol{\Sigma}}(\boldsymbol{y}) = \mathbf{0}$$

and the first result is proved.

(ii) By definition, the unique multivariate geometric expectile is given by

$$\begin{aligned}\mathbf{e}_{\rho,\boldsymbol{\nu},\boldsymbol{\Sigma}}(\boldsymbol{X}) &= \underset{\boldsymbol{x}\in\mathbb{R}^d}{\arg\min}\, E\left(\Lambda_{\rho,\boldsymbol{\nu},\boldsymbol{\Sigma}}(\boldsymbol{X}-\boldsymbol{x})\right)\\ &= \underset{\boldsymbol{x}\in\mathbb{R}^d}{\arg\min}\int_{\mathbb{R}^d}\Lambda_{\rho,\boldsymbol{\nu},\boldsymbol{\Sigma}}(\boldsymbol{t}-\boldsymbol{x})\exp(-\tfrac{1}{2}\Lambda_{\rho,\boldsymbol{\nu},\boldsymbol{\Sigma}}(\boldsymbol{t}-\boldsymbol{\mu}))\mathrm{d}\boldsymbol{t} =: \underset{\boldsymbol{x}\in\mathbb{R}^d}{\arg\min}\, J_{\rho,\boldsymbol{\nu},\boldsymbol{\Sigma},\boldsymbol{\mu}}(\boldsymbol{x}).\end{aligned}$$

To prove that $\mathbf{e}_{\rho,\boldsymbol{\nu},\boldsymbol{\Sigma}}(\boldsymbol{X}) = \boldsymbol{\mu}$, it is then sufficient to check that the gradient of $J_{\rho,\boldsymbol{\nu},\boldsymbol{\Sigma},\boldsymbol{\mu}}$ vanishes at $\boldsymbol{x}=\boldsymbol{\mu}$:

$$\begin{aligned}\nabla J_{\rho,\boldsymbol{\nu},\boldsymbol{\Sigma},\boldsymbol{\mu}}|_{\boldsymbol{x}=\boldsymbol{\mu}} &= -\int_{\mathbb{R}^d}\nabla\Lambda_{\rho,\boldsymbol{\nu},\boldsymbol{\Sigma}}(\boldsymbol{t}-\boldsymbol{\mu})\exp(-\tfrac{1}{2}\Lambda_{\rho,\boldsymbol{\nu},\boldsymbol{\Sigma}}(\boldsymbol{t}-\boldsymbol{\mu}))\mathrm{d}\boldsymbol{t}\\ &= -\int_{\mathbb{R}^d}\nabla\Lambda_{\rho,\boldsymbol{\nu},\boldsymbol{\Sigma}}(\boldsymbol{y})\exp(-\tfrac{1}{2}\Lambda_{\rho,\boldsymbol{\nu},\boldsymbol{\Sigma}}(\boldsymbol{y}))\mathrm{d}\boldsymbol{y} = 2\int_{\mathbb{R}^d}\nabla\exp(-\tfrac{1}{2}\Lambda_{\rho,\boldsymbol{\nu},\boldsymbol{\Sigma}}(\boldsymbol{y}))\mathrm{d}\boldsymbol{y}\\ &= 2\int_{\mathbb{R}^d}\nabla\exp(-\tfrac{1}{2}\Lambda_{\rho,\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\nu},\boldsymbol{I}_d}(\boldsymbol{\Sigma}^{-1/2}\boldsymbol{y}))\mathrm{d}\boldsymbol{y} \propto \int_{\mathbb{R}^d}\nabla\exp(-\tfrac{1}{2}\Lambda_{\rho,\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\nu},\boldsymbol{I}_d}(\boldsymbol{z}))\mathrm{d}\boldsymbol{z}.\end{aligned}$$

For any $j=1,\ldots,d$, integrating along the jth component provides the difference of the limits of $\exp(-\frac{1}{2}\Lambda_{\rho,\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\nu},\boldsymbol{I}_d}(\boldsymbol{z}))$ as z_j, the jth coordinate of $\boldsymbol{z}$, goes to $+\infty$ and $-\infty$, which are both zero since $\Lambda_{\rho,\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\nu},\boldsymbol{I}_d}$ is coercive, see Herrmann et al. (2018, Theorem 4.5). As a consequence, $\nabla J_{\rho,\boldsymbol{\nu},\boldsymbol{\Sigma},\boldsymbol{\mu}}|_{\boldsymbol{x}=\boldsymbol{\mu}} = \mathbf{0}$ and the result follows.

(iii) The proof is similar to that of Lemma A1 in Appendix A·2. Recalling that $\boldsymbol{Q}$ is an orthogonal matrix such that $\boldsymbol{Q}\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\nu} = \boldsymbol{u}_1$, one can write:

$$E\left[\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{X}-\boldsymbol{\mu})\right] = \frac{c_d(\rho)}{(2\pi)^{d/2}}2^{\frac{d-1}{2}}\Gamma\left(\frac{d+1}{2}\right)\boldsymbol{Q}^\top\int_{\mathcal{S}^{d-1}}\frac{\boldsymbol{\theta}\mathrm{d}\mathcal{S}^{d-1}(\boldsymbol{\theta})}{(1+\rho\langle\boldsymbol{\theta},\boldsymbol{u}_1\rangle)^{(d+1)/2}}.$$

Inspired by the proof of Lemma A1, changes of variables lead to:

$$E\left[\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{X}-\boldsymbol{\mu})\right] = \frac{c_d(\rho)}{(2\pi)^{d/2}}2^{\frac{d-1}{2}}\Gamma\left(\frac{d+1}{2}\right)\mathcal{A}^{d-2}\boldsymbol{Q}^\top\boldsymbol{u}_1\int_0^\pi\frac{\cos(\psi)\left(\sin(\psi)\right)^{d-2}}{\left(1+\rho\cos(\psi)\right)^{\frac{d+1}{2}}}\mathrm{d}\psi.$$

Since $\boldsymbol{Q}^\top \boldsymbol{u}_1 = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\nu}$, and

$$\frac{1}{\Gamma\left(\frac{d-1}{2}\right)} \int_0^\pi \frac{\cos(\psi)\left(\sin(\psi)\right)^{d-2}}{\left(1+\rho\cos(\psi)\right)^{\frac{d+1}{2}}} \mathrm{d}\psi = -\frac{d+1}{4}\sqrt{\pi}\rho \frac{{}_2F_1\left(\frac{d+3}{4}, \frac{d+5}{4}; \frac{d+2}{2}; \rho^2\right)}{\Gamma\left(\frac{d+2}{2}\right)},$$

we get the expected result. □

Proof of Proposition 4. Let $\boldsymbol{X} \sim \mathrm{MED}(\rho_1)$. By definition, $\mathbf{e}_{\rho_2}(\boldsymbol{X})$ is the solution of a slightly modified Equation (1) involving two asymmetry parameters ρ_1 and ρ_2:

$$\mathbf{e}_{\rho_2}(\boldsymbol{X}) = \operatorname*{arg\,min}_{\boldsymbol{x}\in\mathbb{R}^d} E\left(\Lambda_{\rho_2}(\boldsymbol{X}-\boldsymbol{x})\right), \tag{A2}$$

where expectation is taken with respect to the $\mathrm{MED}(\rho_1)$. In other words, the expectation is taken with respect to $\boldsymbol{X} \sim \mathrm{MED}(\rho_1)$, but the order of the expectile is ρ_2. Since the objective function in the above optimization task is strictly convex by Proposition A2, it has a unique solution. By cylindrical symmetry of the distribution of $\boldsymbol{X}$ along the first axis (the direction of $\boldsymbol{u}_1$ through the origin), the expectile $\mathbf{e}_{\rho_2}(\boldsymbol{X})$ is also supported on this axis, so there exists some $m \in \mathbb{R}$ such that $\mathbf{e}_{\rho_2}(\boldsymbol{X}) = m\boldsymbol{u}_1$. Let ψ be the function with arguments $(m, \rho_2) \in \mathbb{R} \times [0,1)$ defined from the objective function above as:

$$\psi(m,\rho_2) \coloneqq E\left[\Lambda_{\rho_2}(\boldsymbol{X}-m\boldsymbol{u}_1)\right] = E\left[\|\boldsymbol{X}-m\boldsymbol{u}_1\|(\|\boldsymbol{X}-m\boldsymbol{u}_1\| + \rho_2(X_1-m)\right].$$

Differentiating with respect to m yields

$$\frac{\partial}{\partial m}\psi(m,\rho_2) = E\left[2(m-X_1) - \rho_2\left(\frac{(m-X_1)^2}{\|\boldsymbol{X}-m\boldsymbol{u}_1\|} + \|\boldsymbol{X}-m\boldsymbol{u}_1\|\right)\right].$$

The value in $m=0$ is equal to

$$\frac{\partial}{\partial m}\psi(m=0,\rho_2) = E\left[-2X_1 - \rho_2\left(\frac{X_1^2}{\|\boldsymbol{X}\|} + \|\boldsymbol{X}\|\right)\right], \tag{A3}$$

which is an affine function with negative slope. Additionally, this function vanishes when ρ_2 equals ρ_1, by definition of the expectile of order ρ_1 being the minimizer of (1). Hence, the equivalence of the statement: if $\rho_2 > \rho_1$, then the derivative in (A3) is negative, therefore going along the first axis in the positive direction decreases the objective function, which is necessarily the direction of $\mathbf{e}_{\rho_2}(\boldsymbol{X})$ by convexity; so $m > 0$. The other equivalences are derived similarly. □

Proof of results from Section 3

Proof of Proposition 5. One can write the density of $\boldsymbol{Z} \sim \mathrm{MED}(\rho)$ as:

$$f_d(\boldsymbol{z};\rho) = \frac{c_d(\rho)}{(2\pi)^{d/2}} \exp\left(-\frac{\|\boldsymbol{z}\|^2}{2}\left[1+\rho\left\langle\frac{\boldsymbol{z}}{\|\boldsymbol{z}\|}, \boldsymbol{u}_1\right\rangle\right]\right),$$

and thus, noting that the dot product $\langle\,\cdot\,,\,\cdot\,\rangle$ lies in $[-1,1]$ provides the simple bounds

$$\frac{c_d(\rho)}{(2\pi)^{d/2}} \exp\left(-\frac{\|\boldsymbol{z}\|^2}{2}(1+\rho)\right) \le f_d(\boldsymbol{z};\rho) \le \frac{c_d(\rho)}{(2\pi)^{d/2}} \exp\left(-\frac{\|\boldsymbol{z}\|^2}{2}(1-\rho)\right). \tag{A4}$$

Let $\boldsymbol{X} = \boldsymbol{\Sigma}^{1/2}\boldsymbol{Q}^\top\boldsymbol{Z} + \boldsymbol{\mu} \sim \mathrm{MED}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\nu}, \rho)$ from Corollary 1. By change of variable, the density of $\boldsymbol{X}$ is given by

$$f_d(\boldsymbol{x};\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\nu},\rho) = \frac{1}{|\boldsymbol{\Sigma}|^{1/2}} f_d(\boldsymbol{Q}\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{x}-\boldsymbol{\mu});\rho).$$

By replacement in (A4), the result is proved. □

Proof of Corollary 2. Let $\boldsymbol{Z} = \boldsymbol{Q}\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{X} - \boldsymbol{\mu})$ be the standardized version of $\boldsymbol{X}$; see Corollary 1. For any non-zero $\boldsymbol{\beta} \in \mathbb{R}^d$ and $p > 0$, one has

$$E\left(|\langle \boldsymbol{\beta}, \boldsymbol{Z}\rangle|^p\right) = \int_{\mathbb{R}^d} |\langle \boldsymbol{\beta}, \boldsymbol{z}\rangle|^p f_d(\boldsymbol{z};\rho)\mathrm{d}\boldsymbol{z},$$

and thus, in view of Proposition 5,

$$m_d(\rho)E\left(|Y_+|^p\right) \leq E\left(|\langle \boldsymbol{\beta}, \boldsymbol{Z}\rangle|^p\right) \leq M_d(\rho)E\left(|Y_-|^p\right),$$

where Y_- and Y_+ are univariate centered Gaussian random variables with variance $1/(1-\rho)$ and $1/(1+\rho)$, respectively. It follows that

$$m_d(\rho)(1+\rho)^{-p/2}E\left(|Y_0|^p\right) \leq E\left(|\langle \boldsymbol{\beta}, \boldsymbol{Z}\rangle|^p\right) \leq M_d(\rho)(1-\rho)^{-p/2}E\left(|Y_0|^p\right), \tag{A5}$$

where Y_0 is a standard Gaussian random variable whose moments are given by $E\left(|Y_0|^p\right) = 2^{p/2}\pi^{-1/2}\Gamma((p+1)/2)$. By $E\left(|\langle \boldsymbol{\beta}, \boldsymbol{Z}\rangle|^p\right) = E\left(|\langle \boldsymbol{\Sigma}^{1/2}\boldsymbol{\beta}, \boldsymbol{X} - \boldsymbol{\mu}\rangle_{\boldsymbol{\Sigma}^{-1}}|^p\right)$, and noting that the bounds of (A5) do not depend on $\boldsymbol{\beta} \neq 0$, the conclusion follows. □

Proof of Proposition 6. A similar calculation as that of Proposition 3 leads to

$$E\left[\|\boldsymbol{X} - \boldsymbol{\mu}\|^2_{\boldsymbol{\Sigma}^{-1}}\right] = d\frac{(1-\rho)^{d/2}\left(1+\sqrt{\frac{1+\rho}{1-\rho}}\right)^d\left(\rho^2-1\right)}{2^{d-1}\left(\rho^2-1-\sqrt{1-\rho^2}\right)}\frac{\Gamma\left(\frac{d}{2}\right)}{\Gamma\left(\frac{d-1}{2}\right)}\pi^{-1/2}\int_{-1}^{1}\frac{\left(1-x^2\right)^{\frac{d-3}{2}}}{(1+\rho x)^{1+d/2}}\mathrm{d}x.$$

The result is deduced via the identity

$$\frac{1}{\Gamma\left(\frac{d-1}{2}\right)}\int_{-1}^{1}\frac{\left(1-x^2\right)^{(d-3)/2}}{(1+\rho x)^{1+d/2}}\mathrm{d}x = \frac{1}{\Gamma\left(1+\frac{d}{2}\right)}\frac{2^{d/2-2}\sqrt{\pi}\left(d-d\rho^2+2\rho^2+d\sqrt{1-\rho^2}\right)}{\left(1+\sqrt{1-\rho^2}\right)^{d/2}(1-\rho^2)^{3/2}}.$$

Proof of Proposition 7. *(i)* is a direct application of the asymptotic equivalent for the ${}_2F_1$ function from Lemma A2 in Appendix A·2, while *(ii)* comes from the fact that the second term in the squared norm expectation in Proposition 6 dominates the first. □

Proof of results from Section 4

Proof of Proposition 8. Let H_{d-1} be the hyperplane orthogonal to $\boldsymbol{u}_1$. (i) Consider the one-dimensional margin Z_1 of $\boldsymbol{Z}$:

$$f_{Z_1}(z_1) = \int_{H_{d-1}} f_d(z_1, \boldsymbol{z}_{-1})\mathrm{d}\boldsymbol{z}_{-1} = \frac{c_d\left(\rho\right)}{(2\pi)^{d/2}}\int_{H_{d-1}}\exp\left(-\frac{z_1^2}{2} - \frac{\|\boldsymbol{z}_{-1}\|^2}{2} - \frac{\rho z_1}{2}\sqrt{z_1^2+\|\boldsymbol{z}_{-1}\|^2}\right)\mathrm{d}\boldsymbol{z}_{-1}.$$

The first result is obtained by integrating the bounds provided by Proposition 5 on H_{d-1}. Secondly, let us assume $z_1 > 0$, the case $z_1 < 0$ being obtained by replacing ρ by $-\rho$. Consider the change of variable $\boldsymbol{z}_{-1} = z_1\boldsymbol{x}$:

$$f_{Z_1}(z_1) = \frac{c_d\left(\rho\right)}{(2\pi)^{d/2}}z_1^{d-1}\int_{H_{d-1}}\exp\left(-\frac{z_1^2}{2}(1+\|\boldsymbol{x}\|^2 + \rho\ \mathrm{sign}(z_1)\sqrt{1+\|\boldsymbol{x}\|^2})\right)\mathrm{d}\boldsymbol{x}.$$

Switching to polar coordinates yields:

$$f_{Z_1}(z_1) = \mathcal{A}^{d-2}\frac{c_d\left(\rho\right)}{(2\pi)^{d/2}}z_1^{d-1}\int_0^\infty\exp\left(-\frac{z_1^2}{2}(1+r^2+\rho\sqrt{1+r^2})\right)r^{d-2}\mathrm{d}r,$$

where $\mathcal{A}^{d-2} = 2\pi^{(d-1)/2}/\Gamma((d-1)/2)$ is the surface of the unit sphere $\mathcal{S}^{d-2}$. Introducing

$$I_d(z_1) = \int_0^\infty\exp\left(-\frac{z_1^2}{2}\varphi(r)\right)r^{d-2}\mathrm{d}r$$

with $\varphi(r) = 1 + r^2 + \rho\sqrt{1+r^2}$, and applying Lemma A3 with $\varphi(0) = 1+\rho$ and $\varphi''(0) = 2+\rho$, it follows that,

$$f_{Z_1}(z_1) \underset{z_1 \to +\infty}{\approx} \frac{c_d(\rho)}{(2\pi)^{1/2}} \left(\frac{1}{1+\rho/2}\right)^{(d-1)/2} \exp\left(-\frac{z_1^2}{2}(1+\rho)\right),$$

while, in view of the above remarks,

$$f_{Z_1}(z_1) \underset{z_1 \to -\infty}{\approx} \frac{c_d(\rho)}{(2\pi)^{1/2}} \left(\frac{1}{1-\rho/2}\right)^{(d-1)/2} \exp\left(-\frac{z_1^2}{2}(1-\rho)\right).$$

(ii) For any $\boldsymbol{z} \in \mathbb{R}^d$, write $\boldsymbol{z} = (z_1, \boldsymbol{z}_{-1})$ so that

$$\begin{aligned} f_{\boldsymbol{Z}_{-1}}(\boldsymbol{z}_{-1}) &= \int_{\mathbb{R}} f_d(z_1 \boldsymbol{u}_1 + \boldsymbol{z}_{-1}) \mathrm{d}z_1 \\ &= \frac{c_d(\rho)}{(2\pi)^{d/2}} \exp\left(-\frac{\|\boldsymbol{z}_{-1}\|^2}{2}\right) \int_{\mathbb{R}} \exp\left(-\frac{z_1^2}{2} - \frac{\rho z_1}{2}\sqrt{z_1^2 + \|\boldsymbol{z}_{-1}\|^2}\right) \mathrm{d}z_1 \\ &=: \tilde{f}(\|\boldsymbol{z}_{-1}\|), \end{aligned}$$

which proves that the marginal distribution of $\boldsymbol{Z}_{-1}$ is an elliptically contoured distribution. Let $z = \|\boldsymbol{z}_{-1}\|$. A simple linear change of variables decouples z and z_1 in the integrand above

$$\int_{\mathbb{R}} \exp\left(-\frac{z_1^2}{2} - \frac{\rho z_1}{2}\sqrt{z_1^2 + \|\boldsymbol{z}_{-1}\|^2}\right) \mathrm{d}z_1 = z \int_{\mathbb{R}} \exp\left(-\frac{z^2}{2}\left(z_1^2 + \rho z_1 \sqrt{z_1^2+1}\right)\right) \mathrm{d}z_1.$$

Let function h be defined on $\mathbb{R}$ by $h(z_1) = z_1^2 + \rho z_1\sqrt{z_1^2+1}$. We know from Proposition A2 that h is strictly convex, thus it admits a single minimum on $\mathbb{R}$. Simple algebra yields that the arg min and minimum are $t_\rho = -\left(\frac{1}{2}\left(\frac{1}{\sqrt{1-\rho^2}} - 1\right)\right)^{1/2}$ and $h(t_\rho) = \frac{1}{2}(\sqrt{1-\rho^2} - 1)$. As a result, Laplace's method provides the following large z equivalent to the integral

$$\int_{\mathbb{R}} \exp\left(-\frac{z^2}{2}\left(z_1^2 + \rho z_1\sqrt{z_1^2+1}\right)\right) \mathrm{d}z_1 \underset{z_1 \to +\infty}{\approx} \sqrt{\frac{2\pi}{z^2 h''(t_\rho)/2}} \exp\left(-\frac{z^2}{2} h(t_\rho)\right),$$

where

$$h''(t_\rho) = 2 - \frac{1}{\sqrt{2}}\rho\sqrt{\frac{1}{\sqrt{1-\rho^2}} - 1}\left(\frac{1}{\sqrt{1-\rho^2}} + 2\right)\left(1/2\left(\frac{1}{\sqrt{1-\rho^2}} - 1\right) + 1\right)^{-3/2}.$$

Together with the above identities, we get the large $\|\boldsymbol{z}_{-1}\|$ distribution

$$\begin{aligned} f_{\boldsymbol{Z}_{-1}}(\boldsymbol{z}_{-1}) &= \frac{c_d(\rho)}{(2\pi)^{d/2}} \|\boldsymbol{z}_{-1}\| \exp\left(-\frac{\|\boldsymbol{z}_{-1}\|^2}{2}\right) \exp\left(-\frac{\|\boldsymbol{z}_{-1}\|^2}{2}\frac{1}{2}(\sqrt{1-\rho^2} - 1)\right) \\ &\underset{\|\boldsymbol{z}_{-1}\| \to +\infty}{\approx} C_d(\rho) \frac{c_d(\rho)}{(2\pi)^{d/2}} \exp\left(-\frac{\|\boldsymbol{z}_{-1}\|^2}{2}\frac{1}{2}(\sqrt{1-\rho^2} + 1)\right), \end{aligned}$$

where

$$C_d(\rho) = \sqrt{\frac{4\pi}{h''(t_\rho)}} = \frac{\sqrt{\pi}}{2^{d-2}} \frac{(1+\sqrt{1-\rho^2})^{d-3/2}}{\sqrt{1-\rho^2}}.$$

Additionally, $C_d(0) = \sqrt{2\pi}$, since $h''(t_\rho) = 2$ when $\rho = 0$. □

Proof of Proposition 9. *(i)* For any $z \in \mathbb{R}$, the density $f_{\langle \boldsymbol{w}, \boldsymbol{Z}\rangle | P_{\boldsymbol{w}^\perp}(\boldsymbol{Z}) = \boldsymbol{0}}(z)$ is equal to:

$$\frac{f_d(z\boldsymbol{w};\rho)}{f_{P_{\boldsymbol{w}^\perp}(\boldsymbol{Z})}(\boldsymbol{0})} \propto f_d(z\boldsymbol{w};\rho) \propto \exp\left(-\frac{1}{2}z^2 - \frac{\rho}{2}\langle \boldsymbol{w}, \boldsymbol{u}_1\rangle z|z|\right) = \exp\left(-\frac{1}{2}(1 + \rho\langle \boldsymbol{w}, \boldsymbol{u}_1\rangle \mathrm{sign}(z)) z^2\right),$$

which is the one-dimensional MED of the statement.

(ii) Conditioning on $Z_1 = 0$ cancels the non-quadratic term in the exponential term of the density. As a consequence, this conditional density is proportional to $\exp\left(-\frac{1}{2}\|\boldsymbol{z}_{-1}\|^2\right)$, which identifies the asserted multivariate Gaussian distribution. □

Proofs of results from Section 5

Proof of Proposition 10. The proof is a direct application of Lemma 4.2 of Pötscher and Prucha (2013). The only non-trivial criterion to be checked is that

$$\sup_{\tilde{\boldsymbol{\theta}}\in\bar{\Theta}} \left| n^{-1}\ell_n(\tilde{\boldsymbol{\theta}}) - E\log f_d(\boldsymbol{X};\tilde{\boldsymbol{\theta}})\right| \xrightarrow{P} 0,$$

as $n \to \infty$. We can apply Theorem 5 of Andrews (1992) to verify the condition. To do so, we verify that $n^{-1}\ell_n(\tilde{\boldsymbol{\theta}}) \xrightarrow{P} E\log f_d(\boldsymbol{X};\tilde{\boldsymbol{\theta}})$, pointwise. This is possible via a trivial application of the weak law of large numbers combined with Proposition 5 and Corollary 2. To complete the proof, we must demonstrate the existence of a function $\alpha(\boldsymbol{x})$, such that $\alpha(\boldsymbol{x}) \geq \sup_{\tilde{\boldsymbol{\theta}}\in\bar{\Theta}} |\log f_d(\boldsymbol{x};\tilde{\boldsymbol{\theta}})|$ and $E(\alpha(\boldsymbol{X})) < \infty$. Such a function can be obtained as follows. Use Proposition 5 to obtain the inequalities:

$$m_d(\rho)\,\phi_d\left(\boldsymbol{x};\boldsymbol{\mu},\frac{\boldsymbol{\Sigma}}{1+\rho}\right) \leq f_d\left(\boldsymbol{x};\tilde{\boldsymbol{\theta}}\right) \leq M_d(\rho)\,\phi_d\left(\boldsymbol{x};\boldsymbol{\mu},\frac{\boldsymbol{\Sigma}}{1-\rho}\right),$$

and thus

$$\left|\log f_d\left(\boldsymbol{x};\tilde{\boldsymbol{\theta}}\right)\right| \leq |\log m_d(\rho)| + |\log M_d(\rho)| + \left|\log\phi_d\left(\boldsymbol{x};\boldsymbol{\mu},\frac{\boldsymbol{\Sigma}}{1+\rho}\right)\right| + \left|\log\phi_d\left(\boldsymbol{x};\boldsymbol{\mu},\frac{\boldsymbol{\Sigma}}{1-\rho}\right)\right|,$$

from which we set

$$\alpha(\boldsymbol{x}) = \sup_{\rho\in[0,\gamma]}\{|\log m_d(\rho)| + |\log M_d(\rho)|\} + \sup_{\tilde{\boldsymbol{\theta}}\in\bar{\Theta}}\left|\log\phi_d\left(\boldsymbol{x};\boldsymbol{\mu},\frac{\boldsymbol{\Sigma}}{1+\rho}\right)\right| + \sup_{\tilde{\boldsymbol{\theta}}\in\bar{\Theta}}\left|\log\phi_d\left(\boldsymbol{x};\boldsymbol{\mu},\frac{\boldsymbol{\Sigma}}{1-\rho}\right)\right|,$$

where γ above is defined in $\bar{\Theta}$ in (11). Function α is well-defined (and non-infinite) by construction of the compact set $\bar{\Theta}$. We apply Proposition 5, again, to obtain

$$\begin{aligned} E\sup_{\tilde{\boldsymbol{\theta}}\in\bar{\Theta}}\left|\log\phi_d\left(\boldsymbol{X};\boldsymbol{\mu},\frac{\boldsymbol{\Sigma}}{1-\rho}\right)\right| &= \int \sup_{\tilde{\boldsymbol{\theta}}\in\bar{\Theta}}\left|\log\phi_d\left(\boldsymbol{x};\boldsymbol{\mu},\frac{\boldsymbol{\Sigma}}{1-\rho}\right)\right| f\left(\boldsymbol{x};\tilde{\boldsymbol{\theta}}\right)\mathrm{d}\boldsymbol{x} \\ &\leq M_d(\rho)\int \sup_{\tilde{\boldsymbol{\theta}}\in\bar{\Theta}}\left|\log\phi_d\left(\boldsymbol{x};\boldsymbol{\mu},\frac{\boldsymbol{\Sigma}}{1-\rho}\right)\right| \phi_d\left(\boldsymbol{x};\boldsymbol{\mu},\frac{\boldsymbol{\Sigma}}{1-\rho}\right)\mathrm{d}\boldsymbol{x}, \end{aligned}$$

where the last integral is known to be finite. The analogous result holds for $E\sup_{\tilde{\boldsymbol{\theta}}\in\bar{\Theta}}\left|\log\phi_d\left(\boldsymbol{X};\boldsymbol{\mu},\frac{\boldsymbol{\Sigma}}{1+\rho}\right)\right|$, and thus we have verified that $\alpha(\boldsymbol{x})$ satisfies the necessary properties.

A·2. *Auxiliary results*

LEMMA A1. *For any $\rho \in [0,1)$, for any $\boldsymbol{\Sigma} \in \mathrm{SPD}(d)$ and $\boldsymbol{\mu}$, $\boldsymbol{\nu} \in \mathbb{R}^d$ such that $\|\boldsymbol{\nu}\|_{\boldsymbol{\Sigma}^{-1}} = 1$,*

$$\frac{c_d(\rho)}{(2\pi)^{d/2}|\boldsymbol{\Sigma}|^{1/2}}\int_{\mathbb{R}^d}\exp(-\Lambda_{\rho,\boldsymbol{\nu},\boldsymbol{\Sigma}}(\boldsymbol{x}-\boldsymbol{\mu}))\mathrm{d}\boldsymbol{x} = 1,$$

where $c_d(\rho)$ is given in (2)*.*

Proof. Let us calculate the integral

$$J_d(\rho) = |\boldsymbol{\Sigma}|^{-1/2}\int_{\mathbb{R}^d}\exp\left(-\frac{1}{2}\|\boldsymbol{z}_{-1}-\boldsymbol{\mu}\|^2_{\boldsymbol{\Sigma}^{-1}}\right)\exp\left(-\frac{\rho}{2}\langle\boldsymbol{z}_{-1}-\boldsymbol{\mu},\boldsymbol{\nu}\rangle_{\boldsymbol{\Sigma}^{-1}}\|\boldsymbol{z}_{-1}-\boldsymbol{\mu}\|_{\boldsymbol{\Sigma}^{-1}}\right)\mathrm{d}\boldsymbol{z}_{-1}.$$

The change of variable $\boldsymbol{z} = \boldsymbol{\Sigma}^{-1/2}(\boldsymbol{z}_{-1} - \boldsymbol{\mu})$ yields

$$J_d(\rho) = \int_{\mathbb{R}^d} \exp\left(-\frac{1}{2}\|\boldsymbol{z}\|^2\right) \exp\left(-\frac{\rho}{2}\langle \boldsymbol{z}, \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\nu}\rangle \|\boldsymbol{z}\|\right) \mathrm{d}\boldsymbol{z}.$$

By assumption, $\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\nu}$ lies on the unit sphere, so there exists an orthogonal matrix, denoted $\boldsymbol{Q}$, such that $\boldsymbol{Q}\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\nu} = \boldsymbol{u}_1$, the first unit axis vector. Then, the change of variable $\boldsymbol{x} = \boldsymbol{Q}\boldsymbol{z}$ leads to

$$J_d(\rho) = \int_{\mathbb{R}^d} \exp\left(-\frac{1}{2}\|\boldsymbol{x}\|^2\right) \exp\left(-\frac{\rho}{2}\langle \boldsymbol{x}, \boldsymbol{u}_1\rangle \|\boldsymbol{x}\|\right) \mathrm{d}\boldsymbol{x}.$$

We use polar coordinates: $\boldsymbol{x} = r\boldsymbol{\theta}$, $r \in \mathbb{R}_+$ and $\boldsymbol{\theta} \in \mathcal{S}^{d-1}$ leading to $\mathrm{d}\boldsymbol{x} = r^{d-1}\mathrm{d}r\,\mathrm{d}\mathcal{S}^{d-1}(\boldsymbol{\theta})$ and

$$\begin{aligned} J_d(\rho) &= \int_{\mathcal{S}^{d-1}} \int_0^\infty \exp\left(-\frac{r^2}{2}(1 + \rho\langle \boldsymbol{\theta}, \boldsymbol{u}_1\rangle)\right) r^{d-1}\mathrm{d}r\,\mathrm{d}\mathcal{S}^{d-1}(\boldsymbol{\theta}) \\ &= \int_0^\infty \exp\left(-\frac{u^2}{2}\right) u^{d-1}\mathrm{d}u \int_{\mathcal{S}^{d-1}} \frac{\mathrm{d}\mathcal{S}^{d-1}(\boldsymbol{\theta})}{(1 + \rho\langle \boldsymbol{\theta}, \boldsymbol{u}_1\rangle)^{d/2}} =: J_d^{(1)} J_d^{(2)}(\rho). \end{aligned}$$

Clearly, the first integral $J_d^{(1)}$ equals $2^{d/2-1}\Gamma(d/2)$ by definition of the gamma function. To deal with the second integral $J_d^{(2)}(\rho)$, write $\boldsymbol{\theta}$ as $(\theta_1, \boldsymbol{\theta}')$. Introduce now polar coordinates for both θ_1 and $\boldsymbol{\theta}'$: $\theta_1 = r_1\eta$ and $\boldsymbol{\theta}' = r_2\boldsymbol{\zeta}$, where $\eta \in \mathcal{S}^0 = \{-1, 1\}$ and $\boldsymbol{\zeta}' \in \mathcal{S}^{d-2}$. Since $r_1^2 + r_2^2 = 1$ and both r_1 and r_2 are non-negative, introduce $\psi \in [0, \pi/2]$ such that $r_1 = \cos\psi$ and $r_2 = \sin\psi$. Then Equation (A5) of Kume and Wood (2005) yields

$$\mathrm{d}\mathcal{S}^{d-1}(\boldsymbol{\theta}) = (\sin\psi)^{d-2}\mathrm{d}\psi\,\mathrm{d}\mathcal{S}^0(\eta)\,\mathrm{d}\mathcal{S}^{d-2}(\boldsymbol{\zeta}).$$

Noting that $x_1 = \langle \boldsymbol{x}, \boldsymbol{u}_1\rangle = r\langle \boldsymbol{\theta}, \boldsymbol{u}_1\rangle = r\theta_1$ on one side, and $\theta_1 = x_1/r_1 \cos\psi = \eta\cos\psi$ on the other side, thus $\langle \boldsymbol{\theta}, \boldsymbol{u}_1\rangle = \eta\cos\psi$, and

$$\begin{aligned} J_d^{(2)}(\rho) &= \int_{\mathcal{S}^{d-1}} \frac{\mathrm{d}\mathcal{S}^{d-1}(\boldsymbol{\theta})}{(1 + \rho\theta_1)^{d/2}} = \int_{\mathcal{S}^{d-2}} \int_{\mathcal{S}^0} \int_0^{\pi/2} \frac{(\sin\psi)^{d-2}}{(1 + \rho\eta\cos\psi)^{d/2}} \mathrm{d}\psi\,\mathrm{d}\mathcal{S}^0(\eta)\,\mathrm{d}\mathcal{S}^{d-2}(\boldsymbol{\zeta}) \\ &= \mathcal{A}^{d-2} \int_{\mathcal{S}^0} \int_0^{\pi/2} \frac{(\sin\psi)^{d-2}}{(1 + \rho\eta\cos\psi)^{d/2}} \mathrm{d}\psi\,\mathrm{d}\mathcal{S}^0(\eta) \\ &= \mathcal{A}^{d-2} \int_0^{\pi/2} \frac{(\sin\psi)^{d-2}}{(1 + \rho\cos\psi)^{d/2}} + \frac{(\sin\psi)^{d-2}}{(1 - \rho\cos\psi)^{d/2}} \mathrm{d}\psi = \mathcal{A}^{d-2} \int_0^{\pi} \frac{(\sin\psi)^{d-2}}{(1 + \rho\cos\psi)^{d/2}} \mathrm{d}\psi, \end{aligned}$$

where we have successively integrated out, with respect to $\boldsymbol{\zeta}$, leading to the surface of $\mathcal{S}^{d-2}$ denoted by $\mathcal{A}^{d-2} = 2\pi^{(d-1)/2}/\Gamma((d-1)/2)$, then to η. As a conclusion,

$$J_d(\rho) = 2^{d/2}\pi^{(d-1)/2} \frac{\Gamma(d/2)}{\Gamma((d-1)/2)} \int_0^\pi \frac{(\sin\psi)^{d-2}}{(1 + \rho\cos\psi)^{d/2}} \mathrm{d}\psi.$$

Further, tedious algebra yields:

$$\frac{1}{\Gamma((d-1)/2)} \int_0^\pi \frac{(\sin\psi)^{d-2}}{(1 + \rho\cos\psi)^{d/2}} \mathrm{d}\psi = \frac{\sqrt{\pi}}{\Gamma(d/2)\sqrt{1-\rho^2}} \left(\frac{2}{1 + \sqrt{1-\rho^2}}\right)^{\frac{d}{2}-1},$$

and finally $J_d(\rho) = (2\pi)^{d/2}/c_d(\rho)$, which concludes the proof. □

LEMMA A2. *If a, b, c are positive reals such that $a + b > c$, then the hypergeometric function satisfies the following asymptotic equivalent when $z \to 1$:*

$${}_2F_1(a, b; c; z) \underset{z\to 1}{\approx} \frac{\Gamma(a + b - c)\Gamma(c)}{\Gamma(a)\Gamma(b)} \frac{1}{(1 - z)^{a+b-c}}.$$

Proof. The hypergeometric function admits the following linear combination (Gradshteyn and Ryzhik, 2014, Chapter 7.5):

$$
\begin{aligned}
{}_2F_1(a,b;c,z) = {} & \frac{\Gamma(c)\Gamma(c-a-b)}{\Gamma(c-a)\Gamma(c-b)}{}_2F_1(a,b;a+b+1-c;1-z) \\
& + \frac{\Gamma(a+b-c)\Gamma(c)}{\Gamma(a)\Gamma(b)}(1-z)^{c-a-b}{}_2F_1(c-a,c-b;1+c-a-b;1-z).
\end{aligned}
$$

The result is proved by observing that in the $z \to 1$ limit, both hypergeometric functions in the r.h.s. linear combination tend to one. □

LEMMA A3. *Let $\varphi : \mathbb{R}_+ \to \mathbb{R}$ be a twice continuously differentiable function such that $\varphi'(0) = 0$ and $\varphi''(x) \geq m > 0$ for any $x \in \mathbb{R}_+$. Introduce, for any $t \in \mathbb{R}$ and $d \geq 2$, $I_d(t) = \int_0^\infty \exp\left(-\frac{t^2}{2}\varphi(x)\right) x^{d-2}\mathrm{d}x$. Then,*

$$
I_d(t) \underset{|t|\to\infty}{\approx} 2^{d-2}\Gamma((d-1)/2)\left(\frac{1}{\varphi''(0)}\right)^{(d-1)/2}\frac{1}{|t|^{d-1}}\exp\left(-\frac{t^2}{2}\varphi(0)\right).
$$

Proof. Clearly, $I_d(\cdot)$ is an even function, we thus focus on the situation where $t > 0$. Consider for any $x \geq 0$ the Taylor expansion:

$$
\varphi(x) = \varphi(0) + x\varphi'(0) + \frac{x^2}{2}R(x) = \varphi(0) + \frac{x^2}{2}R(x),
$$

where $R(x) = \varphi''(\theta x)$ for some $\theta = \theta(x) \in (0,1)$. By replacement, we get

$$
\begin{aligned}
I_d(t) &= \exp\left(-\frac{t^2}{2}\varphi(0)\right)\int_0^\infty x^{d-2}\exp\left(-\frac{t^2x^2}{4}R(x)\right)\mathrm{d}x \\
&= \left(\frac{2}{t}\right)^{d-1}\exp\left(-\frac{t^2}{2}\varphi(0)\right)\int_0^\infty v^{d-2}\exp\left(-v^2R(2v/t)\right)\mathrm{d}v,
\end{aligned}
$$

using the change of variable $v = tx/2$. Since $R(x) \to \varphi''(0)$ as $x \to 0$ and φ'' is lower bounded on $\mathbb{R}_+$, Lebesgue dominated convergence theorem thus entails

$$
\begin{aligned}
I_d(t) &\underset{t\to+\infty}{\approx} \left(\frac{2}{t}\right)^{d-1}\exp\left(-\frac{t^2}{2}\varphi(0)\right)\int_0^\infty v^{d-2}\exp\left(-v^2\varphi''(0)\right)\mathrm{d}v \\
&= \left(\frac{2}{\varphi''(0)}\right)^{(d-1)/2}\frac{1}{t^{d-1}}\exp\left(-\frac{t^2}{2}\varphi(0)\right)\int_0^\infty u^{d-2}\exp\left(-u^2/2\right)\mathrm{d}u.
\end{aligned}
$$

Finally, classical results regarding the standard Gaussian distribution yield

$$
\int_0^\infty u^{d-2}\exp\left(-u^2/2\right)\mathrm{d}u = 2^{(d-3)/2}\Gamma((d-1)/2)
$$

and the result is proved. □

B. APPENDIX: IMPLEMENTATION DETAILS AND ADDITIONAL APPLICATION RESULTS

B·1. *First three moments in dimension $d = 1$*

Figure 5 represents the first three moments centered with respect to the μ parameter for the 1-dimension MED. Expressions are provided in Section 2.

B·2. *Bayesian implementation in Stan*

In this section, we describe an implementation of the Bayesian model of Section 5·2 in Stan probabilistic programming language (Carpenter et al., 2017). Stan starts with describing the data:

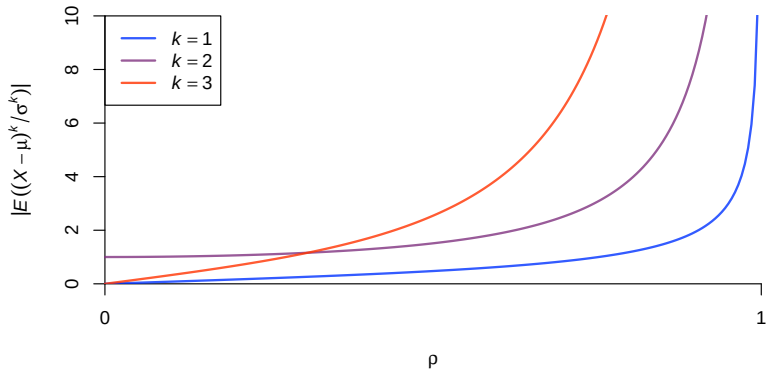

Fig. 5. Absolute value of first three moments centered with respect to parameter μ for the 1-dimension MED.

```
data {
  int<lower=1> d;          // data dimension
  int<lower=0> n;          // number of observations
  vector[d] Y[n];          // observations
}
```

Then parameters are declared. We also adopt the parameterization of Section 5·1, which replaces $\boldsymbol{\nu}$ in $\boldsymbol{\theta}$, by $\tilde{\boldsymbol{\nu}} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\nu}$. Instead of the covariance matrix, it is more efficient to work with the precision matrix $\boldsymbol{\Omega} = \boldsymbol{\Sigma}^{-1} \in \mathrm{SPD}\,(d)$.

```
parameters {
  vector[d] mu;                     // mean
  cov_matrix[d] Sigma_inv;          // precision matrix
  real<lower=0,upper=1> rho;        // asymmetry parameter
  unit_vector[d] nu_tilde;          // direction (tilde version)
  vector<lower=0>[d] diag_sigma_sq; // mu hyperprior cov mat
}

transformed parameters {
  matrix[d, d] sigma_inv_sqrt = square_root(Sigma_inv, d);
}
```

Specific functions used in order to compute the square root of a symmetric and positive definite matrix, to compute the log normalizing constant, and to define the log p.d.f. of the MED density are then declared:

```
functions {
   // assumes A is SPD
    matrix square_root(matrix A, int R) {
        matrix[R, R] eigvecs = eigenvectors_sym(A);
        vector[R] eigvals = eigenvalues_sym(A);
        return eigvecs * diag_matrix(sqrt(eigvals)) / eigvecs;
    }

    real log_norm_const(int d, real rho) {
      real num = (d-1)*log(2) + log(1+inv_sqrt(1-pow(rho, 2)));
      real den = d * log(sqrt(1 - rho) + sqrt(1 + rho));
      return num - den;
    }

    real mymed_lpdf(vector x, vector m, matrix sigma_inv,
                    matrix sigma_inv_sqrt, vector nu, real r, int d) {
      real out = multi_normal_prec_lpdf(x | m, sigma_inv);
      out -= (log_norm_const(d, r));
      out -= r / 2* (x-m)'* sigma_inv_sqrt * nu * sqrt(quad_form(sigma_inv, x-m));
```

```
      return out;
    }
}
```

The Bayesian model is finally complemented with the prior distributions and the likelihood. In the case where no specific expert information is available, the prior of Section 5·2 can be adapted as follows:

```
model {
  mu ~ normal(rep_vector(0, d), diag_sigma_sq);
  for (j in 1:d) {
    diag_sigma_sq[j] ~ inv_gamma(1, 1);
  }
  rho ~ uniform(0, 1);
  Sigma_inv ~ wishart(d, diag_matrix(rep_vector(1, d)));
  for (i in 1:n) {
    Y[i] ~ mymed(mu, Sigma_inv, sigma_inv_sqrt, nu_tilde, rho, d);
  }
}
```

B·3. *Bayesian applications: miscellanea*

The data distributions used in the Bayesian applications of Section 7·1 and Section 7·2 are described in Table 5. Let us recall that Gumbel copula is defined for all $(u, v) \in [0,1]^2$ and $\theta \geq 1$ by $C_\theta(u,v) = \exp(-[(-\log u)^\theta + (-\log v)^\theta]^{1/\theta})$. In particular, $C_1(u,v) = uv$ is the independence copula, see Nelsen (2007) for further details. Besides, $\mathcal{SN}$ denotes the skew normal distribution (Azzalini and Capitanio, 1999). Finally, the MLE results of the discriminant analysis of Section 7·3 are provided in Table 6.

Table 5. *Underlying data distributions for* $\boldsymbol{X}^{(1)}, \ldots, \boldsymbol{X}^{(4)}$ *in Section* 7·1 *and* 7·2

Vector	Copula	$\boldsymbol{X}_1^{(i)} \sim$	$\boldsymbol{X}_2^{(i)} \sim$
$\boldsymbol{X}^{(1)} = (X_1^{(1)}, X_2^{(1)})$	Independence	$\mathcal{N}(0,1)$	$\mathcal{N}(0,1)$
$\boldsymbol{X}^{(2)} = (X_1^{(2)}, X_2^{(2)})$	Independence	$\mathcal{SN}(-1,1,2)$	t_4
$\boldsymbol{X}^{(3)} = (X_1^{(3)}, X_2^{(3)})$	Gumbel, $\theta = 2$	$\mathcal{N}(0,1)$	$\mathcal{N}(0,1)$
$\boldsymbol{X}^{(4)} = (X_1^{(4)}, X_2^{(4)})$	Gumbel, $\theta = 2$	$\mathcal{SN}(-1,1,2)$	t_4

Table 6. *Maximum likelihood estimators computed with the genuine and forged records averaged over* 100 *train samples*

	$\boldsymbol{\mu}$	$\boldsymbol{\Sigma}$	$\boldsymbol{\nu}$	ρ
Genuine	$\begin{pmatrix} 3.50 \\ 3.94 \\ 0.09 \\ 0.17 \end{pmatrix}$	$\begin{pmatrix} 3.01 & -1.92 & -1.76 & 0.56 \\ -1.92 & 22.21 & -10.46 & -5.76 \\ -1.76 & -10.46 & 8.92 & 2.97 \\ 0.56 & -5.76 & 2.97 & 2.93 \end{pmatrix}$	$\begin{pmatrix} 0.98 \\ -0.26 \\ -0.56 \\ 1.07 \end{pmatrix}$	0.74
Forged	$\begin{pmatrix} -0.85 \\ -0.36 \\ 0.12 \\ -0.07 \end{pmatrix}$	$\begin{pmatrix} 2.73 & 0.94 & -3.53 & 0.07 \\ 0.94 & 23.94 & -20.65 & -5.39 \\ -3.53 & -20.65 & 21.57 & 4.26 \\ 0.07 & -5.39 & 4.26 & 3.05 \end{pmatrix}$	$\begin{pmatrix} 0.86 \\ 0.53 \\ -1.72 \\ 1.01 \end{pmatrix}$	0.73