Annual Review of Statistics and Its Application

Flexible Models for Complex Data with Applications

Christophe Ley,[1] Slađana Babić,[1,2] and Domien Craens[1]

[1]Department of Applied Mathematics, Computer Science and Statistics, Ghent University, Gent, B-9000 Belgium; email: christophe.ley@ugent.be

[2]Vlerick Business School, B-1210 Brussels, Belgium

Annu. Rev. Stat. Appl. 2021. 8:369–91

First published as a Review in Advance on November 13, 2020

The *Annual Review of Statistics and Its Application* is online at statistics.annualreviews.org

https://doi.org/10.1146/annurev-statistics-040720-025210

Keywords

circular statistics, copulas, directional statistics, finite mixtures, heavy tails, skewness, transformation approach

Abstract

Probability distributions are the building blocks of statistical modeling and inference. It is therefore of the utmost importance to know which distribution to use in what circumstances, as wrong choices will inevitably entail a biased analysis. In this article, we focus on circumstances involving complex data and describe the most popular flexible models for these settings. We focus on the following complex data: multivariate skew and heavy-tailed data, circular data, toroidal data, and cylindrical data. We illustrate the strength of flexible models on the basis of concrete examples and discuss major applications and challenges.

1. INTRODUCTION

Probability distributions are the building blocks of statistical modeling and inference. Perhaps more basically yet importantly, they allow us to quantify the uncertainty of random phenomena and to describe the random behavior of data by means of a mathematical formula. The latter aspect seems to be a core need for scientists—an example is the search for universal formulae explaining physical phenomena or epidemics such as the coronavirus disease 2019 (COVID-19) crisis, where researchers strive to detect whether the growth rates are of an exponential type or approach a more reassuring logistic or Gompertz curve (Jia et al. 2020). This quest for explaining randomness via a formula (ideally, a simple one) was also the driving force behind Carl Friedrich Gauss's search for a probability distribution describing the law of errors of his astronomical measurements (Gauss 1809). The resulting exponential law of errors is nowadays, of course, known as the Gaussian distribution or normal distribution. The Belgian scientist Adolphe Quetelet propagated the use of the Gaussian distribution in biology and the social sciences to model various types of data. Following the efforts of Quetelet, many researchers started to work on modeling all types of data by means of a meaningful probability distribution, thereby leading to the creation of a plethora of new models. For instance, in 1897, Galton proposed the log-normal distribution (under the impetus of finding a transformation rendering positive data normally distributed), while at the end of the nineteenth century, Edgeworth and Pearson started a fierce competition about the best nonnormal distribution. We refer the reader to Stigler (1986) for details about this fruitful period and to Ley (2015, section 2) for a brief historic account.

Describing a random phenomenon with a probability distribution is not a goal per se, but it allows, inter alia, the calculation of concrete probabilities of events and risks, and the making of predictions and defining strategies related to the data at hand (especially in our modern big data era). A historic example is the scale of intelligence quotient points, which was chosen to be a normal distribution centered at 100 with a standard deviation of 15, allowing one to derive to what upper percentile a person with, e.g., an IQ of 133 belongs (Hunt 2011). Another example, in the domain of renewable energies, is that the two-parameter Weibull distribution is generally adopted as a model for wind speed, load, and power, and thus is used to determine quantities such as capacity factor and average output of a wind turbine (Gugliani et al. 2018) as well as to identify optimal wind turbine generator parameters (Jangamshetti & Rau 2001). However, it is crucial to use the correct—or at least a reasonably fitting—distribution, since wrong choices will inevitably bias the subsequent analysis and calculations, with potentially dramatic consequences. A famous example is the 2008 financial crisis, when financial institutions had recourse to the multivariate Gaussian distribution for modeling the behavior of their assets. By nature, financial data are heavy-tailed and often also skewed because negative events are typically more extreme than positive events, and the Gaussian distribution, not accounting for extreme and skew events, led to an underestimation of risks. The danger of choosing a too-simplified model and hence missing the intrinsic nature of the data is all the more present when dealing with complex data, be they complex due to their form, topology, or volume. In this article, we focus on (potentially vast-dimensional) multivariate skew and heavy-tailed data, as well as on circular, toroidal, and cylindrical data, and present and discuss the (in our opinion) best state-of-the-art flexible models for these complex data. For univariate data on the entire real line, we refer the interested reader to the review articles of Lee et al. (2013), Jones (2015), and Ley (2015), and for data on the positive real line (so-called size or survival data), we refer readers to the monograph by Kleiber & Kotz (2003) and the book chapter by Dominicy & Sinner (2017).

A natural question at this point is how to define a flexible model, or perhaps more importantly, which properties a good flexible model ought to possess. Upon perusing the literature and from personal experience, we came up with the following list of desirable properties:

- Versatility: As the word "flexible" suggests, such a model should ideally be able to exhibit as many distinct shapes as possible and, consequently, be more robust to misspecifications than simple models.
- Tractability: The density should be of a tractable form and amenable to calculations.
- Interpretability: The number of parameters should be as small as possible, and the parameters should bear clear interpretations in order to infer conclusions about the underlying population from which the data were taken.
- Data generating mechanism: This desideratum has two goals. (*a*) We should be able to readily simulate new data from the model in order to produce large-scale stochastic simulations and predictions (e.g., about the spread of a disease), and (*b*) when it is related to a nice stochastic representation, we may be able to naturally link a particular model to data for which we can trace back their real generation process.
- Straightforward parameter estimation: A correct parameter estimation procedure is the basis for the subsequent calculation of quantities such as the risk of exceeding a certain threshold value and for statistical inference.

As an additional, but less important, property, we further mention testability and model reduction: Natural goodness-of-fit tests can be defined for the flexible model, and it should nest well-known submodels, as this permits model reduction. In what follows, we depict the, in our opinion, most useful flexible models for multivariate skew and heavy-tailed data (Section 2), circular data (Section 3), and toroidal and cylindrical data (Section 4). A discussion about major applications of and challenges related to flexible parametric distributions is given in Section 5 before we conclude in Section 6 with some final comments.

2. FLEXIBLE MODELS FOR MULTIVARIATE SKEW AND HEAVY-TAILED DATA

In this section, we present models that turned out to be useful for modeling multivariate skew and heavy-tailed data. Besides the already-mentioned field of finance, such complex data appear in various other domains such as biostatistics (Lambert & Vandenhende 2002), environmental sciences (Genton & Thompson 2004), and meteorology (Field & Genton 2006). We focus on the level of flexibility of each model in terms of degree of skewness and ability to model tails lighter and/or heavier than the normal tails. The choice of a suitable model and estimation of parameters are covered, and we see that usually it turns out that the more flexible the model is, the more challenging its parameters are to estimate, especially in higher dimensions. Our choice of versatile distributions is based on the recent comparison of Babić et al. (2019), a paper we refer the reader to for further flexible distributions. Other general popular references for multivariate distributions are the works of Kotz et al. (2004) and Balakrishnan & Lai (2009), the latter for the two-dimensional case.

2.1. Skew-Elliptical Distributions

The family of skew-elliptical distributions is a natural extension of elliptical distributions, which are themselves an extension of the multivariate Gaussian distribution. Skew-elliptical distributions are able to accommodate both skewness and heavy and/or light tails. They are obtained by introducing skewness into elliptical distributions via a skewing function. To define formally the family of skew-elliptical distributions, we need to first define elliptical distributions.

A d-dimensional random vector $\mathbf{X}$ is said to be elliptically distributed if and only if there exists a vector $\boldsymbol{\mu} \in \mathbb{R}^d$, a positive semidefinite and symmetric matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{d\times d}$, and a function $\varphi : \mathbb{R} \to \mathbb{R}^+$

such that the characteristic function of $\mathbf{X}$ is given by $\mathbf{t} \mapsto \exp(i\mathbf{t}'\boldsymbol{\mu})\,\varphi\,(\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}), \mathbf{t} \in \mathbb{R}^d$. An elliptical random vector $\mathbf{X}$ can conveniently be represented by the stochastic representation (here $=_d$ stands for equality in distribution)

$$\mathbf{X} =_d \boldsymbol{\mu} + \mathcal{R}\boldsymbol{\Lambda}\mathbf{U}^{(k)}, \tag{1.}$$

where $\boldsymbol{\Lambda} \in \mathbb{R}^{d\times k}$ has maximal rank $k \leq d$ and is such that $\boldsymbol{\Lambda}\boldsymbol{\Lambda}' = \boldsymbol{\Sigma}, \mathbf{U}^{(k)}$ is a k-dimensional random vector uniformly distributed on the unit hypersphere, and $\mathcal{R}$ is a nonnegative random variable independent of $\mathbf{U}^{(k)}$. If $\mathcal{R}$ has a density, then the density of $\mathbf{X}$ is of the form

$$\mathbf{x} \mapsto c_{d,g}|\boldsymbol{\Sigma}|^{-1/2}g((\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})), \tag{2.}$$

where $g : \mathbb{R}_0^+ \to \mathbb{R}^+$ is the radial function related to the distribution of $\mathcal{R}$ and typically depends on a tail weight parameter, and $c_{d,\,g}$ is a normalizing constant. The vector $\boldsymbol{\mu}$ represents a location parameter, while $\boldsymbol{\Sigma}$ stands for a scatter parameter. The multivariate normal distribution, of course, corresponds to $g(r) = \exp(-r/2)$, and the Student's t-distribution with $\nu > 0$ degrees of freedom to $g(r) = (1 + r/\nu)^{-(\nu\,+\,d)/2}$. Elliptical distributions were introduced by Kelker (1970) and allow for both lighter-than-normal and heavier-than-normal tails while keeping the elliptical geometry of the multinormal equidensity contours. We refer the reader to Paindaveine (2012) for an overview and to Babić et al. (2019, section 2) for a concise presentation of the main properties and modeling limitations.

Skew-elliptical distributions are obtained by modulating elliptical symmetry by means of a skewing function $\Pi : \mathbb{R}^d \times \mathbb{R}^d \to [0, 1]$ satisfying $\Pi(-\mathbf{z}, \boldsymbol{\delta}) + \Pi(\mathbf{z}, \boldsymbol{\delta}) = 1, \mathbf{z}, \boldsymbol{\delta} \in \mathbb{R}^d$, and $\Pi(\mathbf{z}, \mathbf{0}) = 1/2, \mathbf{z} \in \mathbb{R}^d$. Let $\mathbf{Y}$ be an elliptically symmetric random d-vector with density given by Equation 2 and U a uniform random variable on (0,1), both independent of each other. Then,

$$\mathbf{X} =_d \begin{cases} \mathbf{Y} & \text{if } U \leq \Pi(\boldsymbol{\Sigma}^{-1/2}(\mathbf{Y}-\boldsymbol{\mu}), \boldsymbol{\delta}) \\ -\mathbf{Y} & \text{if } U > \Pi(\boldsymbol{\Sigma}^{-1/2}(\mathbf{Y}-\boldsymbol{\mu}), \boldsymbol{\delta}) \end{cases} \tag{3.}$$

follows a skew-elliptical distribution with density

$$\mathbf{x} \mapsto 2c_{d,g}|\boldsymbol{\Sigma}|^{-1/2}g((\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}))\Pi(\boldsymbol{\Sigma}^{-1/2}(\mathbf{x}-\boldsymbol{\mu}), \boldsymbol{\delta}).$$

Whenever $\boldsymbol{\delta} \neq \mathbf{0}$, the resulting density is skewed in the direction of $\boldsymbol{\delta}$, while at $\boldsymbol{\delta} = \mathbf{0}$, the original elliptical distribution is retrieved. Thus, $\boldsymbol{\delta}$ endorses the role of skewness parameter. A salient feature of skew-elliptical distributions is their simple expression since the normalizing constant is directly inherited from the elliptical base distribution. Generating skew-elliptical data is easy thanks to the stochastic representations given by Equations 1 and 3. A potential drawback lies in the fact that the tail weight parameter is typically one-dimensional since it issues from the scalar function g. Maximum likelihood estimation in skew-elliptical distributions is relatively straightforward, yet certain problems are well known, such as the fact that some members of the skew-elliptical family suffer from a Fisher information singularity in the vicinity of symmetry, due to collinearity between the scores for location and skewness (see, e.g., Ley & Paindaveine 2010b, Hallin & Ley 2012). The most famous representative suffering from this flaw is the multivariate skew-normal of Azzalini & Dalla Valle (1996). For a monograph-long treatment of skew-elliptical distributions, both from a theoretical and an applied modeling perspective, we refer the reader to Genton (2004). An interesting recent contribution of relevance for applied work is that of Adcock & Azzalini (2020). We also remark that skew-elliptical distributions are part of the wider family of skew-symmetric distributions (Wang et al. 2004), where the base density is centrally symmetric instead of elliptically symmetric.

An interesting representative of skew-elliptical distributions is the multivariate skew-t distribution proposed by Azzalini & Capitanio (2003; see also Azzalini & Genton 2008) [note that other skew versions of the multivariate t-distribution exist in the literature; see, for instance, Jones & Faddy (2003)]. It is a tractable and robust distribution with parameters that regulate both skewness and heavy tails. It is formally defined as a ratio of a multivariate skew-normal variate and an appropriate transformation of a chi-square random variable, that is, $\mathbf{X} =_d \boldsymbol{\mu} + V^{-1/2}\mathbf{Y}$, where $\mathbf{Y}$ has a multivariate skew-normal distribution and $V \sim \chi^2_\nu/\nu$ with $\nu > 0$, independent of $\mathbf{Y}$. The skew-t distribution allows for tails that are heavier than tails of the Gaussian distribution, but it cannot represent lighter tails. This limitation can be overcome by considering the multivariate extended skew-t distribution introduced by Arellano-Valle & Genton (2010).

Skew-elliptical models can be used to model vast-dimensional data since they remain tractable. The R package `fMultivar` (Wuertz et al. 2020) provides tools for fitting the multivariate skew-t distribution, while the `sn` package (Azzalini 2017) can be used for the skew-normal distribution.

2.2. Transformation Approach Distributions

The transformation approach is based on a simple idea: start from a basic random vector $\mathbf{Y}$, typically multivariate normal, and turn it into $\mathbf{X} =_d \mathbf{H}^{-1}(\mathbf{Y})$ via some diffeomorphism $\mathbf{H} : \mathbb{R}^d \to \mathbb{R}^d$. This implies a simple data generating mechanism, provided we can simulate $\mathbf{Y}$. Writing f, the density of $\mathbf{Y}$, the new random vector $\mathbf{X}$ has density $\mathbf{x} \mapsto f(\mathbf{H}(\mathbf{x}))|D\mathbf{H}(\mathbf{x})|, \mathbf{x} \in \mathbb{R}^d$, where $D\mathbf{H}(\mathbf{x})$ stands for the determinant of the Hessian matrix associated with the transformation $\mathbf{H}$. For instance, if $\mathbf{Y}$ follows a multinormal $\mathbb{N}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then a transformation $\mathbf{H}_{\mathbf{a},\mathbf{b}}$, with $\mathbf{a} \in \mathbb{R}^d$ a skewness parameter and $\mathbf{b} \in \mathbb{R}^d$ a tail weight parameter, leads to a multivariate distribution with well-identified location, scatter, skewness, and tail weight parameters, where, moreover, the tail weight parameter is d-dimensional. The arguably best-known transformation is the Box-Cox transformation (Box & Cox 1964), whose multivariate version is presented by Andrews et al. (1971). We will not delve into the Box-Cox transformation here, nor into extensions like Yeo & Johnson (2000), but rather refer the interested reader to the recent survey paper by Atkinson (2020).

A well-known modern example obtained via the transformation approach is the multivariate $\mathbf{g}$-and-$\mathbf{h}$ distribution of Field & Genton (2006), extending the well-known scalar g-and-h distribution of Tukey (1977). A random vector $\mathbf{X} \in \mathbb{R}^d$ is said to have a standard multivariate $\mathbf{g}$-and-$\mathbf{h}$ distribution, where $\mathbf{g} = (g_1, \ldots, g_d)' \in \mathbb{R}^d$ controls the skewness and $\mathbf{h} = (h_1, \ldots, h_d)' \in \mathbb{R}^d_+$ controls the tail weight, if it can be represented as

$$\mathbf{X} =_d (\tau_{g_1,h_1}(Z_1), \ldots, \tau_{g_d,h_d}(Z_d))' = \boldsymbol{\tau}_{\mathbf{g},\mathbf{h}}(\mathbf{Z}),$$

where $\mathbf{Z} = (Z_1, \ldots, Z_d)'$ has a standard multivariate normal distribution and the univariate functions $\tau_{g_i,h_i}, i = 1, \ldots, d$, are defined as

$$\tau_{g_i,h_i}(z) = \left(\frac{\exp(g_i z) - 1}{g_i}\right) \exp\left(\frac{h_i}{2} z^2\right), \quad z \in \mathbb{R}.$$

Location and scatter parameters can be introduced as usual, leading to the stochastic representation

$$\mathbf{X} =_d \boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2} \tau_{\mathbf{g},\mathbf{h}}(\mathbf{Z}).$$

By construction, the marginals of $\boldsymbol{\Sigma}^{-1/2}(\mathbf{X} - \boldsymbol{\mu})$ follow scalar g-and-h distributions. The multivariate $\mathbf{g}$-and-$\mathbf{h}$ distribution does not have a closed-form density unless $\mathbf{g} = \mathbf{0}$, and one has to resort to some definition of quantiles to estimate its parameters. Namely, affine-equivariant quantiles of the data set are required. One issue of the multivariate $\mathbf{g}$-and-$\mathbf{h}$ distribution in higher dimensions is that the number of directions in which to compute the quantiles grows exponentially. Unfortunately, to the best of our knowledge, there is no R package that provides tools for fitting this distribution, but more details can be found in Field & Genton (2006).

Finally, we also refer the reader to Jones & Pewsey (2009) for the multivariate sinh-arcsinh normal distribution, another representative of this approach, and to the general study on transformation approach distributions as skewing mechanisms by Ley & Paindaveine (2010a).

2.3. Copulas

A function $C : [0, 1]^d \to [0, 1]$ is a d-dimensional copula if C is a joint cumulative distribution function (cdf) of a d-dimensional random vector on the unit cube $[0, 1]^d$ with uniform marginals. A key result concerning copulas is Sklar's theorem (Sklar 1959), which enables linking multivariate cdfs to their univariate marginals through the concept of copulas. It states that any multivariate cdf F with marginal cdfs $F_1, \ldots, F_d$ can be expressed under the form $F(x_1, \ldots, x_d) = C(F_1(x_1), \ldots, F_d(x_d))$, where C is a suitable copula. Conversely, any combination of such a copula C and marginal distributions leads to a multivariate random vector with cdf $C(F_1(x_1), \ldots, F_d(x_d))$. The first appearance of copulas in the statistical literature is often traced back to Fréchet (1951), and cornerstone references about copulas are the monographs by Joe (1997) and Nelsen (2006).

Copulas are a powerful tool for constructing multivariate distributions with separately specified marginals and dependence structure. Moreover, they are very useful for representing dependence structures between random variables. Many measures of dependence, such as Spearman's rho or Kendall's tau, can be expressed in terms of copulas. Despite the fact that copulas allow for various types of dependence structures, they face some limitations when it comes to higher dimensions (some of them can be overcome via vine copulas; see below). In what follows, we focus on several families of copulas that are useful in statistical modeling.

Copulas generated by elliptical distributions represent a simple and widely used class of copula distributions also known as meta-elliptical distributions (Fang et al. 2002). An elliptical copula is of the form $C(u) = F(F_g^{-1}(u_1), \ldots, F_g^{-1}(u_d))$, where F is a multivariate elliptical cdf and F_g is the same symmetric marginal cdf for every component. It is convenient to work with elliptical copulas because they inherit the nice stochastic properties of elliptical distributions. They are symmetric, the upper and lower tail dependence coefficients are the same, they can be simply computed, they have a low number of parameters, and lower-dimensional margins are elliptical copulas again. The most commonly used elliptical copulas are the Gaussian and Student's t copulas.

Archimedean copulas are another commonly used class of copulas. A d-variate Archimedean copula is defined as

$$C(u_1, \ldots, u_d) = \Phi^{-1}(\Phi(u_1) + \cdots + \Phi(u_d)), \quad (u_1, \ldots, u_d)' \in [0, 1]^d,$$

where the generator Φ: $[0, 1] \to [0, \infty]$ is continuous, decreasing, convex, and such that $\Phi(1) = 0$. An important source of generators consists of the inverses of the Laplace transforms of cdfs. Examples of Archimedean copulas are the Gumbel (Gumbel 1960), Clayton (Clayton 1978), and Frank (Frank 1979) copulas. The Gumbel and Clayton copulas are asymmetric copulas, the former exhibiting greater dependence in the positive tail and the latter exhibiting greater dependence in the negative tail. The Frank copula is a symmetric copula with both lower and upper tail dependence coefficients equal to 0. Most Archimedean copulas admit an explicit formula, something not possible, for instance, for the Gaussian copula. They allow modeling dependence in arbitrarily high dimensions but with only one parameter governing the strength of dependence. Archimedean copulas are widely used, especially in finance and insurance. Nelsen (2003), in his survey of properties and applications of copulas, mentions some open problems in the field of Archimedean copulas. For example, are there any statistical properties of two random variables that assure that their copula is Archimedean? The choice of the suitable copula family is more difficult, in general, than the choice of the proper marginal distribution families. While there is no general procedure for selecting the copula class, Genest & Rivest (1993) suggest some strategies for selecting the parametric family of Archimedean copulas that provides the best possible fit to a given set of data.

An important class of copulas is vine copulas. Most multivariate copulas become inflexible in higher dimensions and/or they do not allow for different dependence structures between pairs of variables. For many data applications this is too restrictive; therefore, Joe (1996) suggested constructing multivariate copulas using only blocks of bivariate copulas (the pair-copula approach), and Bedford & Cooke (2002) introduced a new graphical model called a vine. We do not enter into the details here but refer the reader to Aas et al. (2009) and the monograph by Kurowicka & Joe (2010). In terms of flexibility, a vine copula on d variables is composed of $d(d-1)/2$ pair-copulas, and each pair-copula can have multiple parameters. The `VineCopula` package by Nagler et al. (2020) provides tools for parameter estimation, model selection, simulation, goodness-of-fit tests, and visualization of vine copula models. The advantage is, of course, the flexibility, but the drawback is that this approach can be computationally unaffordable in terms of estimation, and there is always a possibility of overfitting. Recently, Müller & Czado (2019) proposed a novel three-step approach that overcomes the computational limitation. First, Gaussian methods are applied to split data sets into feasibly small subsets, then parsimonious and flexible vine copulas are applied, and finally, these submodels are reconciled into one joint model. We refer the reader to Müller & Czado (2019) for more details and for an example of the use of vine copulas in high dimensions. A modified Bayesian information criterion (BIC) tailored to sparse vine copula models by Nagler et al. (2019) represents another approach for dealing with the computational cost and overfitting in vine copula models for high dimensions.

Constructing copulas and sampling from them is extremely simple and fast, yet the estimation can become challenging. The copula concept allows us to estimate the marginal distributions and the copula separately, and much work in this field is focused on the estimation of the copula function. The natural approach is the classical maximum likelihood estimation for all parameters and the entire distribution, but it is often computationally too demanding. The most widely used approach is a two-stage estimation process, which is also called inference functions for margins (IFM). First, the margins are estimated, followed by the copula parameters. In both steps, maximum likelihood is used. Alternatively, pseudomaximum likelihood, a two-step procedure in which the marginals are estimated nonparametrically, can be applied. We refer the reader to Embrechts & Hofert (2013) for a discussion of estimation in high dimensions and Joe (1997) for the asymptotics of the two-step approach. For the R platform, there is the well-known `copula` package by Hofert et al. (2017).

2.4. Finite Mixtures

Many multivariate distributions arise from the Gaussian distribution by applying some transformation to a multivariate normal random vector or considering mixtures of multinormal distributions, for instance. Interesting constructions are the so-called scale and location-scale mixtures (see, e.g., McNeil et al. 2005), even more so the subsequent multiple scale (Forbes & Wraith 2014) and multiple location-scale mixtures (Wraith & Forbes 2015) of multinormals. High modeling flexibility, such as different tail weights in every dimension and a simple generating mechanism, are mitigated by a nontractable density and, consequently, serious computational difficulties for parameter estimation even in low-dimensional cases (we had trouble in dimension five, for instance). We refer the reader to Babić et al. (2019) for more details. Instead, we focus here on finite mixtures.

When modeling complex data, it is often unrealistic to assume that our data are coming from the same unimodal distribution. Therefore, one needs flexible distributions that can accommodate multimodality, for instance. Finite mixtures, which are widely used and very powerful for complex data sets and problems that involve data clustering, are a good option here. They can be formally written as

$$f(\mathbf{x}, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_k) = \sum_{i=1}^{k} \pi_i f_i(\mathbf{x}; \boldsymbol{\theta}_i), \mathbf{x} \in \mathbb{R}^d,$$

where the mixing proportion π_i represents the probability that an observation belongs to the ith subpopulation with corresponding component density $f_i(\mathbf{x}; \boldsymbol{\theta}_i)$, with $\boldsymbol{\theta}_i, i = 1, \ldots, k$, the set of parameters of every component (their dimensions need not coincide). The mixing proportions are nonnegative, with $\sum_{i=1}^{k} \pi_i = 1$. Usually the component densities $f_i(\mathbf{x}; \boldsymbol{\theta}_i)$ belong to the same parametric family—for example, the multivariate normal. There is a vast literature concerning mixture models, including works by Everitt & Hand (1981), McLachlan & Basford (1988), Böhning (1999), and Mengersen et al. (2011), to cite but a few. We particularly refer the interested reader to the recent article by McLachlan et al. (2019) for more details and references about finite mixtures (of multinormals).

A common approach for parameter estimation by means of maximum likelihood in finite mixtures is the EM algorithm. The `mclust` package by Scrucca et al. (2016) can be used for Gaussian finite mixture models.

2.5. An Empirical Analysis of Financial Returns Data

We consider a two-dimensional data set consisting of 18 years of daily returns from two major financial indexes from the United States, the Nasdaq and the S&P 500 (Standard and Poor's 500). The sample consists of 4,619 observations from January 7, 2000, through September 20, 2017. Those observations, of course, are serially dependent. In order to neutralize conditional heteroskedasticity, following the suggestion of Lombardi & Veredas (2009) for elliptical and possibly heavy-tailed data, they were adjusted via AR(2)-GARCH(1,1) filtering. We fitted the following models on these adjusted data: the Gaussian, *t*, Gumbel, Clayton, and Frank copulas, all of them combined with normal, Student's *t*, and SAS (sinh-arcsinh) marginals, and the skew-*t* distribution. For the sake of completeness, we mention the density of the SAS-normal distribution (Jones & Pewsey 2009):

$$z \mapsto \frac{h/\sigma}{\sqrt{2\pi(1 + (z-\mu)^2/\sigma^2)}} \left(1 + S_{g,h}^2\left(\frac{z-\mu}{\sigma}\right)\right)^{1/2} \exp\left(-\frac{S_{g,h}^2\left(\frac{z-\mu}{\sigma}\right)}{2}\right),$$

with $z \mapsto S_{g,h}(z) := \sinh(h \sinh^{-1}(z) - g), z \in \mathbb{R}$, and where $\mu \in \mathbb{R}$ is a location, $\sigma > 0$ is a scale, $g \in \mathbb{R}$ is a skewness, and $h > 0$ is a tail weight parameter. We performed a test of unimodality using the `folding.test` function from the `Rfolding` package (Siffer 2018). It turned out that the data are unimodal; therefore, we opted not to include the Gaussian mixture model. The metric comparison was the BIC.

The BIC scores are shown in **Table 1** and indicate that the *t* copula combined with SAS marginals outperforms its competitors. **Figure 1** shows the contour lines of the *t*-SAS model with the estimated parameters and the scatter plot of Nasdaq and S&P 500 data. In general, almost all copula models exhibit the best performance when they are combined with SAS marginals. This is in line with the fact that the SAS distribution can capture skewness and heavy tails. Indeed, the skewness and kurtosis values of Nasdaq are −0.36 and 1.15, respectively, while for the S&P 500 they are −0.47 and 1.75. The values of skewness indicate that both indices are skewed to the left, and the values of kurtosis show that both indices are heavy-tailed. This is reflected by the parameters of the SAS distribution: for Nasdaq, the location was estimated to be 0.089, the scale 0.59, the

Table 1 BIC scores for copula models

Copula	Marginals		
	Normal	**Student's *t***	**SAS**
Gaussian	17,345.56	17,056.73	16,940.23
t	17,159.94	16,738.99	16,623.79
Gumbel	17,365.73	17,030.32	17,103.57
Clayton	19,912.03	18,884.83	18,431.11
Frank	17,650.65	17,640.43	17,640.05

Abbreviation: BIC, Bayesian information criterion; SAS, sinh-arcsinh.

skewness −0.091, and the tail weight 0.72, while the values for the S&P 500 were slightly smaller: 0.066, 0.51, −0.078 and 0.67, respectively. We draw the reader's attention to the fact that values of the tail weight parameter of the SAS distribution that are smaller than 1 indicate heavy tails. The parameters ρ (dependence) and ν (degrees of freedom) of the Student's t copula are estimated to be 0.92 and 4.68, respectively. The value of ρ indicates strong linear correlation.

The last model that we considered is the skew-t distribution, fitted by using the well-known `sn` package (Azzalini 2017). The BIC value that we obtained was 16,697.42, which is second best behind the t-SAS.

3. FLEXIBLE MODELS FOR CIRCULAR DATA

In this section, we give an introduction to circular data and present models that turned out to be useful for modeling univariate circular data. First, we briefly explain the complexity of circular data and illustrate the difference with classical data on the line via a concrete example. Next, we

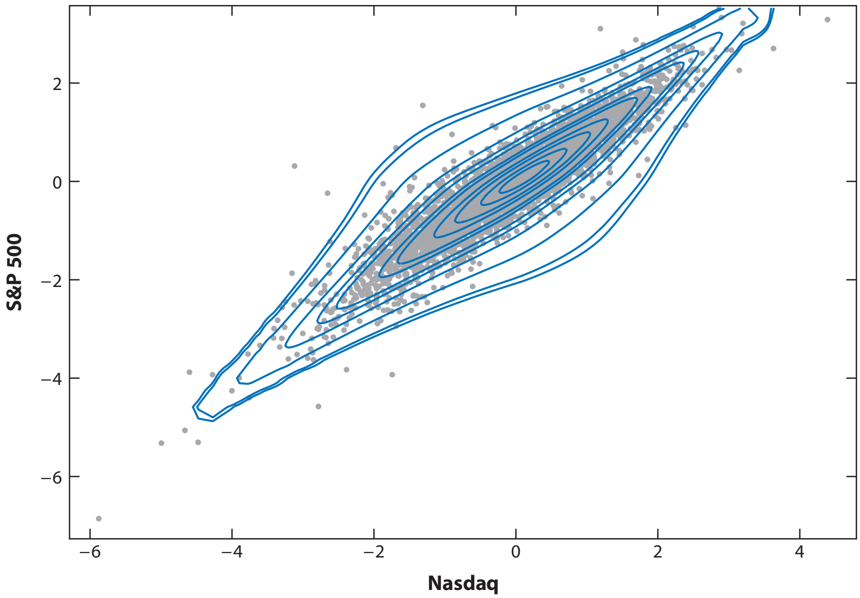

Figure 1

The contour lines of the t-SAS model with the estimated parameters and the scatter plot of Nasdaq and S&P 500. Abbreviations: S&P, Standard and Poor's; SAS, sinh-arcsinh.

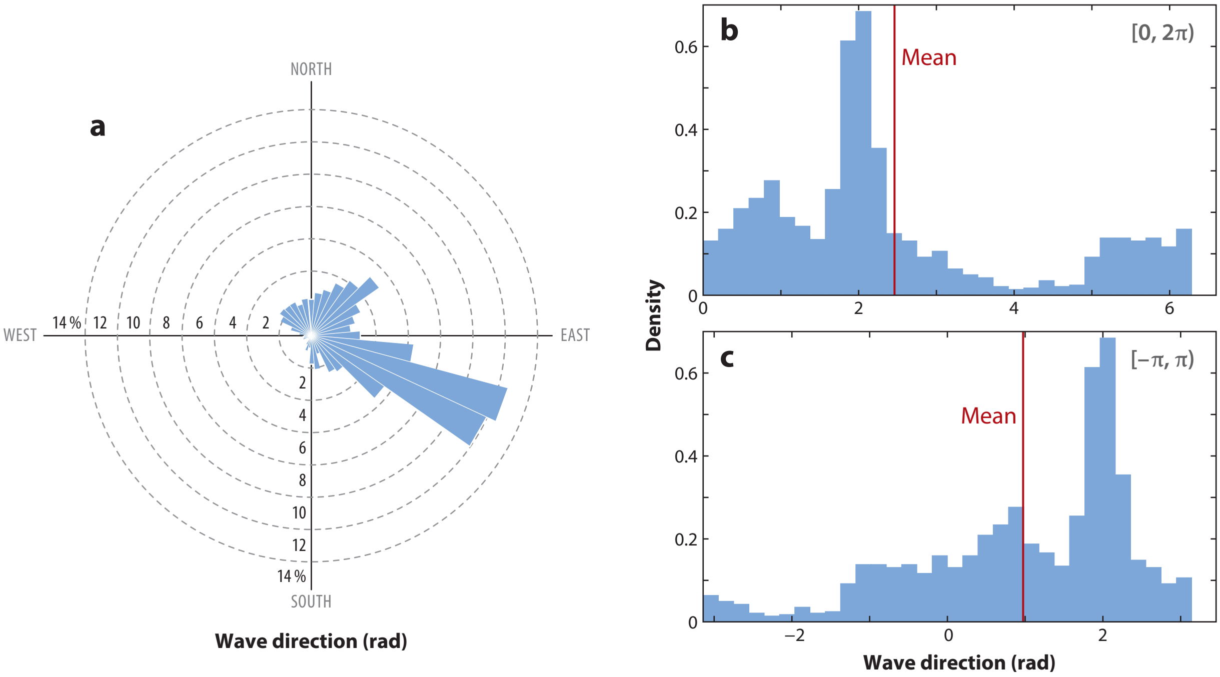

Figure 2

Direction of waves presented in (*a*) a Rose diagram from Lagona et al. (2015) and represented as angles in (*b*). $[0, 2\pi)$ and (*c*) $[-\pi, \pi)$. The red line in panels *b* and *c* indicates the classical mean of the data.

show popular methods to generate distributions on the circle, highlighting some classical models. Then we turn our attention to the, in our opinion, most useful flexible distributions. For a broad overview on circular distributions, we refer the reader to Mardia & Jupp (2000, section 3.5), Jammalamadaka & SenGupta (2001, chapter 2), Pewsey et al. (2013, section 4.3), Ley & Verdebout (2017, sections 2.2, 2.5) and Mardia & Ley (2018).

3.1. Specificities of Circular Data and Circular Densities

The rose diagram in **Figure 2*a*** displays a time series of semihourly wave directions, recorded in the period from February 15 to March 16, 2010, by the buoy of Ancona, located in the Adriatic Sea approximately 30 km from the coast. Lagona et al. (2015) provide a more detailed description of the study of these sea currents. Suppose that we are interested in the average direction of the waves. One could identify each direction with a value in radians: For example, start with north at 0 radians, east at $\frac{\pi}{2}$ radians, and so on, until we are back at north as 2π radians. A histogram of these data is given in **Figure 2*b***. Calculating the classical mean of the data gives 2.46 radians, i.e., around southeast. Instead, we might just as well decide to start labeling south with $-\pi$ and end labeling (clockwise) at π. The histogram of these labelings is shown in **Figure 2*c***, with a resulting classical mean of 0.97 radians, i.e., around northeast. These wildly differing results will convince the reader that one cannot just cut the circle at an arbitrary point and ignore that the start and end points are connected. Thus, a notion as simple as the mean needs to be (re)defined carefully (e.g., as a point on the unit circle of the complex field) when dealing with circular data, and the same goes for nearly all statistical concepts, including densities. We refer the reader to Mardia & Jupp (2000) and Jammalamadaka & SenGupta (2001) for book-long treatments of the field of circular statistics and to Craens & Ley (2018) for a very succinct description of how to deal with circular data.

Besides being a positive function over the real line, a circular density $f(\theta)$ needs to satisfy $f(\theta+2k\pi)=f(\theta)$ for every integer k (2π periodicity) and $\int_{\alpha}^{\alpha+2\pi} f(\theta)\mathrm{d}\theta = 1$ for all $\alpha \in \mathbb{R}$. The expected value of a function g is then well defined if g is 2π periodic: $\mathrm{E}[g(\Theta)] := \int_0^{2\pi} g(\theta)f(\theta)\mathrm{d}\theta$ for $\Theta \sim f$. The trigonometric moment of order p is defined as $m_\Theta(p) = \mathrm{E}[e^{ip\Theta}]$ (the circular equivalent of classical moments for densities on $\mathbb{R}$), and contrary to densities on $\mathbb{R}$, all trigonometric moments exist and are finite.

3.2. Classical Circular Models and How to Construct Them

We now describe the classical approaches to building circular densities.

- The conditioning approach: Express a distribution on $\mathbb{R}^2$ as the joint distribution of the polar coordinates' length and angle, (r, Θ), and consider then the distribution of the angle conditionally on the restriction $r = 1$. A well-known example is the von Mises distribution, obtained by conditioning on a bivariate normal distribution with mean $(\cos(\mu), \sin(\mu))$ and diagonal covariance matrix with both diagonal elements equal to κ^{-1}, $\mu \in [-\pi, \pi)$, $\kappa > 0$. The resulting von Mises density reads

$$\theta \mapsto \frac{1}{2\pi \mathrm{I}_0(\kappa)} e^{\kappa \cos(\theta-\mu)},$$

  where $\mathrm{I}_0(\kappa) = \int_0^{2\pi} e^{\kappa\cos(x)}\mathrm{d}x$ is the modified Bessel function of the first kind and of order zero. Here, μ plays the role of circular mean direction, and the mean resultant length or circular concentration is controlled by κ.
- The projection approach: Instead of conditioning on $r = 1$, one can integrate the length part out to obtain the marginal density of the angular part. A popular case of this approach is the projected normal distribution (see Mardia & Jupp 2000, section 3.5.6).
- The wrapping approach: Let X be a continuous variable on the real line with density f_X, then $\Theta := X \mod 2\pi$ is a circular variable with density

$$f_\Theta(\theta) = \sum_{k=-\infty}^{+\infty} f_X(\theta + 2\pi k).$$

  Most wrapped distributions like the wrapped normal suffer from a major drawback, namely, that the density function does not have a simplified form. A notable exception is the wrapped Cauchy distribution with density

$$\theta \mapsto \frac{1}{2\pi}\frac{1-\rho^2}{1+\rho^2-2\rho\cos(\theta-\mu)},$$

  where $\rho \in [0, 1)$ regulates concentration. A particularly flexible four-parameter example of this approach is the wrapped stable distribution investigated by Pewsey (2008).
- The perturbation approach: An existing circular density is multiplied by a function under the constraint that the resulting product remains a proper circular density. A well-known example of this approach is the cardioid distribution, with density $\frac{1}{2\pi}(1+2\rho\cos(\theta-\mu))$ with concentration parameter $\rho \in [0, 0.5]$. It is obtained from the uniform density $1/(2\pi)$ over the circle.
- The transformation approach: There are various ways to construct new circular distributions using transformations of a stochastic variable, which can be located either on the real line or

on the circle. For a variable X on the real line, a natural choice is $2\arctan(X)$. A well-known transformation on the circle is the Möbius transformation,

$$\Theta \mapsto \mu + \nu + 2\arctan\left[\frac{1-r}{1+r}\tan\left(\frac{1}{2}(\Theta-\nu)\right)\right],$$

with parameters $\mu \in [-\pi, \pi)$, $\nu \in [0, 2\pi)$ and $0 \leq r < 1$. Kato & Jones (2010) applied this transformation to the von Mises distribution.

3.3. Parsimonious Flexible Unimodal Models

We now describe two highly flexible unimodal proposals from the literature: first, a three-parameter symmetric model, and then, a four-parameter model that is able to capture asymmetry.

3.3.1. A flexible symmetric three-parameter model: the Jones–Pewsey family. For modeling symmetric data on the circle, a good choice is the Jones–Pewsey model (Jones & Pewsey 2005) with density given by

$$\theta \mapsto \frac{(\cosh(\kappa\psi))^{1/\psi}}{2\pi P_{1/\psi}(\cosh(\kappa\psi))}\left\{1+\tanh(\kappa\psi)\cos(\theta-\mu)\right\}^{1/\psi}, \qquad 4.$$

where P_α is the associated Legendre function of the first kind of degree $\alpha \in \mathbb{R}$ and order zero, $\mu \in [-\pi, \pi)$ is a location parameter, $\kappa \geq 0$ is a concentration parameter, and $\psi \in \mathbb{R}$ is a shape parameter called index. This model includes the popular circular distributions mentioned in Section 3.2: the von Mises when $\psi \to 0$, the cardioid and wrapped Cauchy distribution for $\psi = 1$ and $\psi = -1$, respectively, and the uniform when $\kappa = 0$ or $\psi \to \pm\infty$ (with κ finite). The Jones–Pewsey density is unimodal for any values of ψ and $\kappa > 0$. Besides its versatility, it allows for stochastic representations as conditional distribution of spherically/elliptically symmetric distributions on $\mathbb{R}^2$, and the trigonometric moments can be expressed in terms of known functions.

Concerning parameter estimation, no closed-form expressions exist for the maximum likelihood estimators of μ, κ, and ψ, implying that numerical optimization methods are required. Numerical maximization converges rapidly as long as $|\kappa\psi| < 6$, while the optimization becomes unstable when $|\kappa\psi|$ is big.

A skewed version of the Jones–Pewsey family has been proposed by Abe & Pewsey (2011a), who apply the perturbation approach by multiplying Equation 4 with $(1 + \lambda\sin(\theta - \mu))$ with $\lambda \in (-1, 1)$ taking the role of skewness parameter. Indeed, only at $\lambda = 0$ is the original symmetric family retrieved, and all other values of λ render a skew distribution. Abe & Pewsey (2011a) speak of a sine-skewed Jones–Pewsey model.

3.3.2. Flexible four-parameter model. Even more flexibility under unimodality is ensured by the four-parameter Kato–Jones distribution (Kato & Jones 2015), which, in our modest opinion, represents an ideal flexible model for this purpose. It is constructed by first defining trigonometric moments and, from there, deriving the related density. The chosen expression for trigonometric moments of order $p \geq 1$ is

$$m(p) = \gamma\left(\rho e^{i\eta}\right)^{-1}\left(\rho e^{i(\mu+\eta)}\right)^{p},$$

where $\mu, \eta \in [-\pi, \pi)$, $\gamma \geq 0$, and $\rho \in [0, 1)$. Associated with these trigonometric moments is the absolutely continuous circular density

$$\theta \mapsto \frac{1}{2\pi}\left\{1 + 2\gamma\frac{\cos(\theta-\mu)-\rho\cos(\eta)}{1+\rho^2-2\rho\cos(\theta-\mu-\eta)}\right\},$$

under the condition

$$(\rho\cos(\eta)-\gamma)^2+(\rho\sin(\eta))^2\leq(1-\gamma)^2. \qquad 5.$$

This family includes both the wrapped Cauchy ($\gamma = \rho$ and $\eta = 0$) and the cardioid density ($\rho = 0$) as special cases. If $\gamma > 0$, the distribution is always unimodal, and it can incorporate diverse shapes: symmetric or asymmetric, flat-topped or sharply peaked. The parameters bear clear interpretations: μ is the mean direction, γ is the mean resultant length, and the circular skewness cS and kurtosis cK of Batschelet (1981) are conveniently given by $\gamma\rho\sin(\lambda)$ and $\gamma\rho\cos(\lambda)$. It is thus possible to reparameterize the density in terms of the parameters cS and cK so as to have pleasing parameter interpretability in terms of location, concentration, skewness, and kurtosis. Random variable generation follows via acceptance/rejection algorithms from wrapped Cauchy variables.

Parameter estimation is possible via both the method of moments and maximum likelihood. The simple trigonometric moments enable easy method of moments estimators, and the resulting values can be used as starting values for the numerical maximization of the log-likelihood function. A difficulty represents the constraint given by Equation 5, which can be circumvented by a reparameterization suggested in Kato & Jones (2015, section 5.2).

3.4. Flexible Models for Multimodal Data

Various circular data sets such as the wave data presented in **Figure 2** are not unimodal, and hence alternatives to the abovementioned flexible unimodal distributions are required. When the data happen to be bipolar symmetric, meaning that they are bimodal and symmetric about two antipodally placed modes (which can typically occur in studies of animal behavior), then we recommend the approach of Abe & Pewsey (2011b), which consists of two steps: Duplicate some unimodal symmetric density $f(\theta-\mu)$ about $\mu\in[-\pi,\pi)$ into $f(2(\theta-\mu))$ and then perturb the latter by multiplication with $(1+\lambda\cos(\theta-\mu))$ for $\lambda\in[0,1]$. A popular yet not very tractable model for bimodal data that are not necessarily bipolar nor symmetric is the generalized von Mises distribution (see Gatto & Jammalamadaka 2007 for a rather recent description).

Our main recommendation, however, for bi- or multimodal data are finite mixtures. Since we describe finite mixtures for multivariate data in Section 2.4, we only briefly touch upon the topic here. Not surprisingly, the most frequently used and thoroughly studied mixtures are mixtures of von Mises distributions (see Mardia & Sutton 1975, Mooney et al. 2003, but also Mardia & Jupp 2000, section 5.5). A good practice to capture all possible features of highly complex data is to use mixtures of flexible unimodal distributions such as the Kato–Jones model, as these are likely to lead to fewer mixture components and better interpretability of the individually detected modes. Estimation typically is operated by means of the EM algorithm, with criteria such as the Akaike information criterion (AIC) or BIC determining the number of modes.

4. FLEXIBLE MODELS FOR TOROIDAL AND CYLINDRICAL DATA

In this section, we consider data with two circular components or with a circular and a linear component, which are respectively called toroidal (or circular-circular) and cylindrical (or circular-linear) data. Modeling such data has gained increasing interest in recent years due to the demands from emerging scientific disciplines such as bioinformatics. Besides the desiderata formulated in the Introduction, the following two properties are also particularly important for this type of models: (*a*) tractable marginal and conditional distributions, ideally of well-known forms, and (*b*) a sound dependence structure. After presenting the (in our opinion) best state-of-the-art flexible models for these data, we conclude this section with an overview of domains dealing with toroidal and cylindrical data and by discussing selected applications. For a detailed overview of the entire literature on models for toroidal and cylindrical data, we refer readers to Ley & Verdebout (2017, sections 2.4 and 2.5).

4.1. Flexible Models on the Torus

The two-dimensional unit torus $\mathcal{S}^1 \times \mathcal{S}^1$ is a natural two-dimensional extension of the unit circle $\mathcal{S}^1$. In that vein, Mardia (1975) introduced the bivariate von Mises density,

$$(\theta_1, \theta_2) \mapsto C \exp\left(\kappa_1 \cos(\theta_1 - \mu_1) + \kappa_2 \cos(\theta_2 - \mu_2)\right. \\ \left. + \ (\cos(\theta_1 - \mu_1), \sin(\theta_1 - \mu_1))\mathbf{A}(\cos(\theta_2 - \mu_2), \sin(\theta_2 - \mu_2))'\right),$$

with normalizing constant C; circular location parameters $\mu_1, \mu_2 \in [-\pi, \pi)$; concentration parameters $\kappa_1, \kappa_2 \geq 0$; and circular-circular dependence parameter $\mathbf{A}$, a 2×2 matrix. With a total of 8 parameters, the bivariate von Mises is overparameterized, and more parameter-parsimonious submodels have been proposed over the years. We here highlight two such models, namely the Sine and the Cosine models. Singh et al. (2002) defined the Sine model with density

$$(\theta_1, \theta_2) \mapsto C \exp\left(\kappa_1 \cos(\theta_1 - \mu_1) + \kappa_2 \cos(\theta_2 - \mu_2) + \beta \sin(\theta_1 - \mu_1) \sin(\theta_2 - \mu_2)\right),$$

with dependence parameter $\beta \in \mathbb{R}$ and the normalizing constant C given by

$$C^{-1} = 4\pi^2 \sum_{i=0}^{\infty} \binom{2i}{i} \left(\frac{\beta^2}{4\kappa_1\kappa_2}\right)^i I_i(\kappa_1) I_i(\kappa_2),$$

where I_p is the modified Bessel function of the first kind and is of order p. Mardia et al. (2007) investigated the Cosine model with density

$$(\theta_1, \theta_2) \mapsto C \exp\left(\kappa_1 \cos(\theta_1 - \mu_1) + \kappa_2 \cos(\theta_2 - \mu_2) - \beta \cos(\theta_1 - \mu_1 - \theta_2 + \mu_2)\right),$$

with dependence parameter $\beta \in \mathbb{R}$ and normalizing constant

$$C^{-1} = 4\pi^2 \left[I_0(\kappa_1) I_0(\kappa_2) I_0(\beta) + 2 \sum_{i=1}^{\infty} I_i(\kappa_1) I_i(\kappa_2) I_i(\beta) \right].$$

For an insightful review on the various variants of bivariate von Mises distributions, we refer the reader to Hamelryck et al. (2012, chapter 6). Multivariate extensions of the Sine and Cosine models for p angles $\theta_1, \ldots, \theta_p$ are respectively introduced in Mardia et al. (2008) and Mardia & Patrangenaru (2005).

The Sine and Cosine models have conditional von Mises densities and marginal densities proportional to expressions of the form $I_0(h(\theta - \mu))\exp(\kappa \cos(\theta - \mu))$ for some function h, some location $\mu \in [-\pi, \pi)$, and some concentration $\kappa \geq 0$. The marginal densities are symmetric, and their number of modes (1 or 2) depends on conditions involving the nonlocation parameters, while the conditions for uni- or bimodality of the joint distributions are of a simpler form (see Mardia et al. 2007).

A drawback of the bivariate von Mises and its submodels is that they cannot model any departure from symmetry. Therefore, Ameijeiras-Alonso & Ley (2019) proposed a general technique to skew symmetric models on the torus. Starting from the base density $f(\theta_1 - \mu_1, \theta_2 - \mu_2; \boldsymbol{\vartheta})$ where $\boldsymbol{\vartheta}$ covers all nonlocation parameters, their approach consists in transforming it into the sine-skewed version,

$$(\theta_1, \theta_2) \mapsto f(\theta_1 - \mu_1, \theta_2 - \mu_2; \boldsymbol{\vartheta}) \left(1 + \sum_{s=1}^{2} \lambda_s \sin(\theta_s - \mu_s)\right), \qquad 6.$$

where $\boldsymbol{\lambda} \in [-1, 1]^2$ plays the role of skewness parameter and satisfies $\sum_{s=1}^{2} |\lambda_s| \leq 1$. For the sake of presentation, we restrict our attention here to the two-dimensional case, but the construction actually holds in any dimension d, replacing $\sum_{s=1}^{2}$ with $\sum_{s=1}^{d}$. As for skew-elliptical distributions, this symmetry modulation does not imply changing the normalizing constant, many properties are inherited from the base symmetric density, and random number generation follows along a similar argument based on a separate uniform random variable.

Parameter estimation for the symmetric toroidal densities and their sine-skewed counterparts can be done by means of maximum likelihood, via the solver proposed by Ye (1987) implemented in the `Rsolnp` package (Ghalanos & Theussl 2015). Ameijeiras-Alonso & Ley (2019) suggest using distinct initial points to avoid having a local maximum as a solution. Submodel testing (in particular, testing for symmetry) is straightforward with a likelihood ratio test.

4.2. Flexible Models on the Cylinder

Two types of cylindrical data exist: those on $\mathcal{S}^1 \times \mathbb{R}$ and those on $\mathcal{S}^1 \times \mathbb{R}^+$. For each type, we present an attractive distribution.

Mardia & Sutton (1978) wished to have a cylindrical model with a simple dependence structure and, in the case of independence, with a normal linear and a von Mises circular part. Thus, they proposed the density

$$(\theta, z) \mapsto \frac{1}{\sigma(2\pi)^{3/2} I_0(\kappa)} \exp\left\{-\frac{(z - (\mu' + \lambda \cos(\theta - \nu)))^2}{2\sigma^2} + \kappa \cos(\theta - \mu)\right\}, \qquad 7.$$

with circular location $\mu \in [-\pi, \pi)$ and concentration $\kappa \geq 0$, linear location $\mu' \in \mathbb{R}$ and dispersion $\sigma > 0$, and circular-linear parameters $\nu \in [-\pi, \pi)$ and $\lambda \geq 0$. The latter parameter regulates the dependence structure, $\lambda = 0$, yielding the product of independent von Mises and normal densities. The Mardia-Sutton distribution can be derived by conditioning on a trivariate normal distribution whose first component is the linear part in Equation 7. The agreeable conditional distributions pave the way to regression analysis, the marginal circular density is again a von Mises, but the form of the marginal linear part is complicated. Maximum likelihood estimation is straightforward with closed-form expressions for the estimators. A more complicated, generalized version of the Mardia-Sutton distribution was proposed by Kato & Shimizu (2008), where the main change is that the conditional and marginal circular distributions are generalized von Mises and hence possibly asymmetric and bimodal.

Extending a too-simple previous proposal by Johnson & Wehrly (1978), Abe & Ley (2017) introduced the density

$$(\theta, z) \mapsto \frac{\alpha\beta^\alpha}{2\pi \cosh(\kappa)} (1 + \lambda \sin(\theta - \mu))\, z^{\alpha-1} \exp\left(-(\beta z)^\alpha (1 - \tanh(\kappa) \cos(\theta - \mu))\right),$$

with circular location $\mu \in [-\pi, \pi)$ and skewness $\lambda \in (-1, 1)$, with linear dispersion $\beta > 0$ and shape parameter $\alpha > 0$, and where $\kappa \geq 0$ takes on the roles of both circular concentration and cylindrical dependence parameter. The Johnson-Wehrly model is retrieved when $\lambda = 0$ and $\alpha = 1$. Contrary to most toroidal and cylindrical models, the Abe-Ley density has a very simple normalizing constant, rendering, for instance, moment calculations simple. The conditional circular and linear laws are sine-skewed von Mises and Weibull, respectively, and we remark that the circular concentration increases with the linear part. The circular marginal distribution is sine-skewed wrapped Cauchy, and the linear marginal density is proportional to $I_0(z^\alpha \beta^\alpha \tanh(\kappa)) z^{\alpha-1} \exp(-(\beta z)^\alpha)$. Parameter estimation works very well via maximum likelihood, and submodel testing (e.g., for the Johnson-Wehrly distribution) is easy.

4.3. Copula Models for the Torus and Cylinder

The very nature of toroidal and cylindrical data makes copulas an appealing structure for building flexible distributions. Thanks to the descriptions of Section 2.3, we can readily write out copula-based densities under their most general form for our purposes here:

$$(\theta, y) \mapsto c\,(F_1(\theta), F_2(y))\, f_1(\theta) f_2(y), \tag{8.}$$

where y is either a linear or a circular component, f_1 and F_1 (respectively, f_2 and F_2) are a circular (respectively, a linear or a circular) density and its associated cdf, and c is a bivariate copula density regulating the dependence. In the toroidal setting, c is a density on the torus with uniform marginals, for which Jones et al. (2015) coined the term circula. This most general form of copulas has not yet been thoroughly studied in the literature, and we hope to motivate researchers to dig more into that appealing structure upon reading these lines. Instead, a particular case of Equation 8 has been popularized by Johnson & Wehrly (1978) and Wehrly & Johnson (1980) without proper mention of the term copulas. Its density corresponds to

$$(\theta, y) \mapsto 2\pi c_b\big(2\pi\,(F_1(\theta) - qF_2(y))\big) f_1(\theta) f_2(y), \tag{9.}$$

with $q = 1$ in the cylindrical and $q = 1$ or -1 in the toroidal setting, and with c_b a circular density called binding function. Jones et al. (2015) gave an in-depth study of the copula structure given by Equation 9, while Pewsey & Kato (2016) developed parametric bootstrap goodness-of-fit tests for the toroidal case.

We wish to mention a particularly interesting copula-based density on the torus, namely the bivariate wrapped Cauchy model of Kato & Pewsey (2015). They chose both marginals as well as the binding function to be wrapped Cauchy, leading to a density proportional to

$$\begin{aligned}&\big(c_0 - c_1 \cos(\theta_1 - \mu_1) - c_2 \cos(\theta_2 - \mu_2) - c_3 \cos(\theta_1 - \mu_1)\\ &\times \cos(\theta_2 - \mu_2) - c_4 \sin(\theta_1 - \mu_1) \sin(\theta_2 - \mu_2)\big)^{-1},\end{aligned}$$

for constants c_1, c_2, c_3, and c_4 depending on two circular concentration parameters and a dependence parameter. The bivariate wrapped Cauchy distribution enjoys various appealing properties such as clear parameter interpretability, a simple (yet long) normalizing constant, unimodality, circular conditional distributions that are wrapped Cauchy, closed-form expressions for the trigonometric moments, and consequently, fast method of moments estimation of its parameters. The sine-skewed bivariate wrapped Cauchy distribution, discussed by Ameijeiras-Alonso & Ley (2019), is, thanks to its added skewness and potential bimodality, perhaps the most promising flexible model on the torus, as further supported by the real data analysis of Ameijeiras-Alonso & Ley (2019).

4.4. Selected Domains of Application

Typical examples of toroidal data are wind directions measured at two distinct moments of the day or peak systolic blood pressure times, converted to angles, during two separate time periods. However, currently, the most influential domain of application has arguably been structural bioinformatics. Researchers have noted that dihedral angles of amino acids can be much better modeled by viewing them as data on the torus (see Hamelryck et al. 2012), an observation that has led to crucial contributions in the protein structure prediction problem. Devising probabilistic models for these couples of angles improves on the analyses previously based on scatter plots called Ramachandran plots (Ramachandran 1963). Given the structure of the data, mixtures of flexible toroidal models are recommended, whose parameters can typically be estimated by means of the Expectation–Maximization (EM) approach.

Environmental, ecological, and biological sciences provide a large amount of cylindrical data—for example, wind direction and other climatological variables, such as wind speed and air temperature, wave direction and wave height, wildfire orientation and burnt area, and the direction in which an animal moves and the distance moved. We wish to mention, in particular, the study of correlated cylindrical data, where the correlation occurs along time or space, leading to temporal, spatial, or spatio-temporal cylindrical models. Typically a mixture of flexible cylindrical models is combined with a hidden Markov chain or Markov random field according to which the parameters of the cylindrical model vary across time or space. The resulting final model is chosen in a parameter-parsimonious way, rendering estimation possible via the EM algorithm, especially in combination with a composite likelihood approach (see, e.g., Ranalli et al. 2018, Lagona 2019). These models, among others, have been used to segment wildfire occurrences and sea regimes.

5. APPLICATIONS AND CHALLENGES

So far the reader has seen what properties a flexible distribution should possess and which models we recommend to use in combination with diverse, complex data. In this section, we briefly consider two further important aspects: major applications of and challenges related to flexible parametric models.

5.1. Major Applications

We now highlight essential applications of flexible distributions, with the added goal of explaining why and when to use flexible parametric models instead of nonparametric approaches such as kernel density estimates (which are, of course, also highly useful but pursue different aims).

- Data analysis: As our real data example illustrates, flexible models able to fit complicated data sets have the asset of interpretable parameters. Very often, the parameters or combinations of parameters provide information about the location, scale, skewness, kurtosis, dependence, or other aspects of data shapes, and this gives the data analyst a more concrete way to describe and investigate the data at hand.
- Calculation of relevant quantities: Especially with tractable densities, it is straightforward to calculate quantities of interest for our data analysis. These may be risks of exceeding a certain threshold, correlation measures such as Kendall's tau, survival functions, peak and duration of epidemics, or fan charts used by banks to quantify uncertainty (Kowalczyk 2013), to name a few examples. Interpretable parameters can also pave the way for further uses, such as, for instance, the ranking of soccer teams on the basis of strength parameters appearing in a suitable probability distribution for match outcomes (Ley et al. 2019).
- Enrichment of other statistical and machine learning techniques: Flexible distributions can serve as a versatile basis for other statistical methods such as (quantile) regression, time series analysis (in both cases as error distribution), Bayesian statistics (as choices for priors), robust inference (flexible models are by nature more robust to model misspecification), and supervised learning (by the use of estimated parameters of a flexible model as information-rich covariate), among many others.
- Stochastic modeling: In situations where it is impossible to calculate probabilities of certain events (e.g., spread of a disease, winner of a tournament, rainfall, development of ecological systems), it is crucial to be able to simulate them repeatedly in order to approximate the true unknown probabilities. Flexible models with simple generating mechanisms are highly useful in such situations.

5.2. Major Challenges

This overview of flexible models would not be complete without an objective discussion on pitfalls and challenges. We identify the following issues that should be dealt with carefully.

- Too many cooks spoil the broth. The same holds true for flexible models: The more models there are, the more difficult it is for practitioners or researchers from other domains to find their way in the jungle of distributions. Charemza et al. (2013) discuss precisely this difficulty. Our main message here is to let quality prevail over quantity. A true contribution from the literature should make the field advance significantly because of a clear need, and not for the sake of proposing yet another variant of a certain distribution. We hope that this article, in combination with other critical reviews like those of Jones (2015) and Babić et al. (2019), sheds light on what makes a flexible distribution a good choice.
- The latter point raises the next issue: Sometimes there is a trade-off between fit and interpretability. It may happen that one parametric model yields a better fit to the data but the parameter interpretation is less obvious than for another model, whose overall fit is inferior (not too inferior though, as otherwise the choice would be clear) but which yields a better story of data generation. There is no global rule for this problem. The answer will be case-by-case and depends ultimately on what the main goal is.
- Which tools should be used to compare various flexible models? In cases where we have good reasons to assume a certain data generating mechanism, the corresponding distribution is a natural choice. If this is not given, and several distributions remain after eliminating those that are, by their properties, not an option, then there are several ways to compare the models. In this article, we used information criteria like the AIC and BIC. Alternatively, and possibly also in combination with the information criteria, goodness-of-fit tests can be used. The latter, however, are mainly means to validate or disqualify models, since one should not compare the sizes of large p-values.
- In cases when we hesitate between several models and perform goodness-of-fit tests, we run into problems of postselection inference (Tibshirani et al. 2015). This is a crucial issue because such pretests can invalidate the classical inference. Therefore, we recommend the reader make an informed decision of which path to take: Either choose a flexible distribution beforehand on the basis of solid stochastic arguments, and perform classical inference, or, if no obvious choice can be made, compare the models and have recourse to postselection inference.

We reckon that there are further challenges, not to mention computational troubles in cases of absent software packages for practical implementation. An influential paper reflecting generally on the culture of data modeling via a stochastic model versus algorithmic modeling is that of Breiman (2001).

6. FINAL COMMENTS

In this article, we have described what desirable properties a flexible distribution ought to possess and presented popular models for complex data, which we judge extremely influential and useful in the light of these desiderata. For the sake of space restrictions, we of course had to make choices on the types of data we wished to consider, and different readers may find different complex data missing. Therefore, we now briefly mention some omitted but important data types and references: (hyper-)spherical data (Ley & Verdebout 2017, sections 2.3 and 2.5), data on high-dimensional spheres and shape spaces (Dryden 2005), rotation data (Arnold & Jupp 2018), spatial data (Gelfand & Banerjee 2017), over- and underdispersed count data (Sellers et al. 2017), and heaped count data (Bermúdez et al. 2017), among others. While the number of probability distributions on the real line has been exploding recently (leading to a suboptimal trend of "generalized modified extended...distribution," with often nonsignificant improvements over existing distributions), the same does not hold true for complex data. We hope to have conveyed an incentive for the use of meaningful flexible distributions to model complex data and for the development of new distributions adhering to the principles stated in the Introduction and bearing in mind the challenges mentioned in the previous section.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

Christophe Ley's research has been supported by the Research Foundation–Flanders (FWO) Krediet aan Navorsers grant with reference number 1510391N. Slađana Babić was supported by a grant (165880) as a PhD Fellow of the FWO. All three authors wish to thank an anonymous reviewer for interesting comments that led to improvements to this article.

LITERATURE CITED

Aas K, Czado C, Frigessi A, Bakken H. 2009. Pair-copula constructions of multiple dependence. *Insur. Math. Econ.* 44:182–98

Abe T, Ley C. 2017. A tractable, parsimonious and highly flexible model for cylindrical data, with applications. *Econom. Stat.* 4:91–104

Abe T, Pewsey A. 2011a. Sine-skewed circular distributions. *Stat. Pap.* 52:683–707

Abe T, Pewsey A. 2011b. Symmetric circular models through duplication and cosine perturbation. *Comp. Stat. Data Anal.* 55:3271–82

Adcock C, Azzalini A. 2020. A selective overview of skew-elliptical and related distributions and of their applications. *Symmetry* 12:118

Ameijeiras-Alonso J, Ley C. 2019. Sine-skewed toroidal distributions and their application in protein bioinformatics. arXiv:1910.13293 [stat.ME]

Andrews D, Gnanadesikan R, Warner J. 1971. Transformations of multivariate data. *Biometrics* 27:825–40

Arellano-Valle RB, Genton MG. 2010. Multivariate extended skew-*t* distributions and related families. *Metron* 68:201–34

Arnold R, Jupp PE. 2018. Statistics of orientations of symmetrical objects. In *Applied Directional Statistics: Modern Methods and Case Studies*, ed. C Ley, T Verdebout, pp. 25–44. Boca Raton, FL: Chapman and Hall/CRC

Atkinson A. 2020. The Box-Cox transformation: review and extensions. *Stat. Sci.* In press

Azzalini A. 2017. `sn`: the skew-normal and related distributions such as the skew-*t*. *R Package*, version 1.6-2. **https://CRAN.R-project.org/package=sn**

Azzalini A, Capitanio A. 2003. Distributions generated by perturbation of symmetry with emphasis on a multivariate skew *t*-distribution. *J. R. Stat. Soc. Ser. B* 65:367–89

Azzalini A, Dalla Valle A. 1996. The multivariate skew-normal distribution. *Biometrika* 83:715–26

Azzalini A, Genton MG. 2008. Robust likelihood methods based on the skew-*t* and related distributions. *Int. Stat. Rev.* 76:106–29

Babić S, Ley C, Veredas D. 2019. Comparison and classification of flexible distributions for multivariate skew and heavy-tailed data. *Symmetry* 11(10):1216

Balakrishnan N, Lai CD. 2009. *Continuous Bivariate Distributions*. New York: Springer

Batschelet E. 1981. *Circular Statistics in Biology*. London: Academic

Bedford T, Cooke RM. 2002. Vines: a new graphical model for dependent random variables. *Ann. Stat.* 30:1031–68

Bermúdez L, Karlis D, Santolino M. 2017. A finite mixture of multiple discrete distributions for modelling heaped count data. *Comput. Stat. Data Anal.* 112:14–23

Böhning D. 1999. *Computer-Assisted Analysis of Mixtures and Applications: Meta-Analysis, Disease Mapping, and Others*. Boca Raton, FL: Chapman and Hall/CRC

Box GE, Cox DR. 1964. An analysis of transformations. *J. R. Stat. Soc. Ser. B* 26:211–43

Breiman L. 2001. Statistical modeling: the two cultures. *Stat. Sci.* 6:199–231

Charemza W, Vela CD, Makarova S. 2013. *Too many skew-normal distributions? The practitioners perspective*. Discuss. Pap. Econ. 13/07, Univ. Leicester, UK

Clayton DG. 1978. A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika* 65:141–51

Craens D, Ley C. 2018. Invited opinion paper: Analysis of biological and biomedical data with circular statistics. *Biostat. Biom. Open Access J.* 5:555671

Dominicy Y, Sinner C. 2017. Distributions and composite models for size-type data. In *Advances in Statistical Methodologies and Their Applications to Real Problems*, ed. T Hokimoto, pp. 159–84. Rijeka, Croatia: InTech

Dryden IL. 2005. Statistical analysis on high-dimensional spheres and shape spaces. *Ann. Stat.* 33:1643–65

Embrechts P, Hofert M. 2013. Statistical inference for copulas in high dimensions: a simulation study. *ASTIN Bull. J. IAA* 43:81–95

Everitt B, Hand D. 1981. *Finite Mixture Distributions*. New York: Chapman and Hall

Fang HB, Fang KT, Kotz S. 2002. The meta-elliptical distributions with given marginals. *J. Multivar. Anal.* 82:1–16

Field C, Genton MG. 2006. The multivariate g-and-h distribution. *Technometrics* 48:104–11

Forbes F, Wraith D. 2014. A new family of multivariate heavy-tailed distributions with variable marginal amounts of tailweight: application to robust clustering. *Stat. Comput.* 24:971–84

Frank MJ. 1979. On the simultaneous associativity of $F(x, y)$ and $x + y - F(x, y)$. *Aequ. Math.* 19:194–226

Fréchet M. 1951. Sur les tableaux de corrélation dont les marges sont données. *Ann. Univ. Lyon 3e Sér. Sci. Sect. A* 14:53–77

Gatto R, Jammalamadaka SR. 2007. The generalized von Mises distribution. *Stat. Methodol.* 4:341–53

Gauss CF. 1809. *Theoria motus corporum coelestium in sectionibus conicis solem ambientium*. Hamburg, Ger.: Perthes and Besser

Gelfand AE, Banerjee S. 2017. Bayesian modeling and analysis of geostatistical data. *Annu. Rev. Stat. Appl.* 4:245–66

Genest C, Rivest LP. 1993. Statistical inference procedures for bivariate Archimedean copulas. *J. Am. Stat. Assoc.* 88:1034–43

Genton MG. 2004. *Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality*. Boca Raton, FL: Chapman and Hall/CRC

Genton MG, Thompson KR. 2004. Skew-elliptical time series with application to flooding risk. In *Time Series Analysis and Applications to Geophysical Systems*, ed. D Brillinger, EA Robinson, FP Schoenberg, pp. 169–85. New York: Springer

Ghalanos A, Theussl S. 2015. `Rsolnp`: general non-linear optimization using augmented Lagrange multiplier method. *R Package*, version 1.16. **https://CRAN.R-project.org/package=Rsolnp**

Gugliani G, Sarkar A, Ley C, Mandal S. 2018. New methods to assess wind resources in terms of wind speed, load, power and direction. *Renew. Energy* 129:168–82

Gumbel EJ. 1960. Bivariate exponential distributions. *J. Am. Stat. Assoc.* 55:698–707

Hallin M, Ley C. 2012. Skew-symmetric distributions and Fisher information—a tale of two densities. *Bernoulli* 18:747–63

Hamelryck T, Mardia KV, Ferkinghoff-Borg J, eds. 2012. *Bayesian Methods in Structural Bioinformatics*. Berlin: Springer

Hofert M, Kojadinovic I, Maechler M, Yan J. 2017. `copula`: multivariate dependence with copulas. *R Package*, version 1.0-0. **https://CRAN.R-project.org/package=copula**

Hunt E. 2011. *Human Intelligence*. Cambridge, UK: Cambridge Univ. Press

Jammalamadaka SR, SenGupta A. 2001. *Topics in Circular Statistics*. Singapore: World Sci.
Jangamshetti SH, Rau GV. 2001. Normalized power curves as a tool for identification of optimum wind turbine generator parameters. *IEEE Trans. Energy Convers.* 16:283–88
Jia L, Li K, Yu J, Guo X, Zhao T. 2020. Prediction and analysis of Coronavirus Disease 2019. arXiv:2003.05447 [q-bio.PE]
Joe H. 1996. Families of m-variate distributions with given margins and $m(m-1)/2$ bivariate dependence parameters. *Lect. Notes Monogr. Ser.* 28:120–41
Joe H. 1997. *Multivariate Models and Multivariate Dependence Concepts*. Berlin: Springer
Johnson RA, Wehrly TE. 1978. Some angular-linear distributions and related regression models. *J. Am. Stat. Assoc.* 73:602–6
Jones MC. 2015. On families of distributions with shape parameters (with discussion). *Int. Stat. Rev.* 83:175–92
Jones MC, Faddy MJ. 2003. A skew extension of the t-distribution, with applications. *J. R. Stat. Soc. Ser. B* 65:159–74
Jones MC, Pewsey A. 2005. A family of symmetric distributions on the circle. *J. Am. Stat. Assoc.* 100:1422–28
Jones MC, Pewsey A. 2009. Sinh-arcsinh distributions. *Biometrika* 96:761–80
Jones MC, Pewsey A, Kato S. 2015. On a class of circulas: copulas for circular distributions. *Ann. Inst. Stat. Math.* 67:843–62
Kato S, Jones MC. 2010. A family of distributions on the circle with links to, and applications arising from, Möbius transformation. *J. Am. Stat. Assoc.* 105:249–62
Kato S, Jones MC. 2015. A tractable and interpretable four-parameter family of unimodal distributions on the circle. *Biometrika* 102:181–90
Kato S, Pewsey A. 2015. A Möbius transformation-induced distribution on the torus. *Biometrika* 102:359–70
Kato S, Shimizu K. 2008. Dependent models for observations which include angular ones. *J. Stat. Plan. Infer.* 138:3538–49
Kelker D. 1970. Distribution theory of spherical distributions and a location-scale parameter generalization. *Sankhyā Indian J. Stat. A* 32:419–30
Kleiber C, Kotz S. 2003. *Statistical Size Distributions in Economics and Actuarial Sciences*. New York: Wiley
Kotz S, Balakrishnan N, Johnson NL. 2004. *Continuous Multivariate Distributions: Models and Applications*, Vol. 1. New York: Wiley
Kowalczyk H. 2013. *Inflation fan charts and different dimensions of uncertainty. What if macroeconomic uncertainty is high?* Work. Pap. 157, Natl. Bank Pol., Warsaw, Pol.
Kurowicka D, Joe H. 2010. *Dependence Modeling: Vine Copula Handbook*. Singapore: World Sci.
Lagona F. 2019. Correlated cylindrical data. In *Applied Directional Statistics: Modern Methods and Case Studies*, ed. C Ley, T Verdebout, pp. 45–60. Boca Raton, FL: Chapman and Hall/CRC
Lagona F, Picone M, Maruotti A. 2015. A Hidden Markov model for the analysis of cylindrical time series. *Environmetrics* 26:535–44
Lambert P, Vandenhende F. 2002. A copula-based model for multivariate non-normal longitudinal data: analysis of a dose titration safety study on a new antidepressant. *Stat. Med.* 21:3197–217
Lee C, Famoye F, Alzaatreh A. 2013. Methods for generating families of univariate continuous distributions in the recent decades. *WIREs Comput. Stat.* 5:219–38
Ley C. 2015. Flexible modelling in statistics: past, present and future. *J. Soc. Fr. Stat.* 156:76–96
Ley C, Paindaveine D. 2010a. Multivariate skewing mechanisms: a unified perspective based on the transformation approach. *Stat. Probab. Lett.* 80:1685–94
Ley C, Paindaveine D. 2010b. On the singularity of multivariate skew-symmetric models. *J. Multivar. Anal.* 101:1434–44
Ley C, Van de Wiele T, Van Eetvelde H. 2019. Ranking soccer teams on the basis of their current strength: a comparison of maximum likelihood approaches. *Stat. Model.* 19:55–77
Ley C, Verdebout T. 2017. *Modern Directional Statistics*. Boca Raton, FL: Chapman and Hall/CRC
Lombardi MJ, Veredas D. 2009. Indirect estimation of elliptical stable distributions. *Comput. Stat. Data Anal.* 53:2309–24
Mardia KV. 1975. Statistics of directional data (with discussion). *J. R. Stat. Soc. Ser. B* 37:349–93
Mardia KV, Hughes G, Taylor CC, Singh H. 2008. A multivariate von Mises distribution with applications to bioinformatics. *Can. J. Stat.* 36:99–109

Mardia KV, Jupp PE. 2000. *Directional Statistics*. New York: Wiley

Mardia KV, Ley C. 2018. Directional distributions. In *Wiley StatsRef: Statistics Reference Online*, ed. N Balakrishnan, T Colton, B Everitt, W Piegorsch, F Ruggeri, J Teugels, pp. 1–13. New York: Wiley

Mardia KV, Patrangenaru V. 2005. Directions and projective shapes. *Ann. Stat.* 33:1666–99

Mardia KV, Sutton TW. 1975. On the modes of a mixture of two von Mises distributions. *Biometrika* 62:699–701

Mardia KV, Sutton TW. 1978. A model for cylindrical variables with applications. *J. R. Stat. Soc. Ser. B* 40:229–33

Mardia KV, Taylor CC, Subramaniam GK. 2007. Protein bioinformatics and mixtures of bivariate von Mises distributions for angular data. *Biometrics* 63:505–12

McLachlan GJ, Basford KE. 1988. *Mixture Models: Inference and Applications to Clustering*. New York: M. Dekker

McLachlan GJ, Lee SX, Rathnayake SI. 2019. Finite mixture models. *Annu. Rev. Stat. Appl.* 6:355–78

McNeil AJ, Frey R, Embrechts P. 2005. *Quantitative Risk Management: Concepts, Techniques and Tools*, Vol. 3. Princeton, NJ: Princeton Univ. Press

Mengersen KL, Robert C, Titterington M. 2011. *Mixtures: Estimation and Applications*. New York: Wiley

Mooney JA, Helms PJ, Jolliffe IT. 2003. Fitting mixtures of von Mises distributions: a case study involving sudden infant death syndrome. *Comp. Stat. Data Anal.* 41:505–13

Müller D, Czado C. 2019. Dependence modelling in ultra high dimensions with vine copulas and the Graphical Lasso. *Comput. Stat. Data Anal.* 137:211–32

Nagler T, Bumann C, Czado C. 2019. Model selection in sparse high-dimensional vine copula models with an application to portfolio risk. *J. Multivar. Anal.* 172:180–92

Nagler T, Schepsmeier U, Stoeber J, Brechmann EC, Graeler B, Erhardt T. 2020. VineCopula: statistical inference of vine copulas. *R Package*, version 2.4.1. **https://CRAN.R-project.org/package=VineCopula**

Nelsen RB. 2003. Properties and applications of copulas: A brief survey. In *Proceedings of the First Brazilian Conference on Statistical Modelling in Insurance and Finance*, ed. J Dhaene, N Kolev, P Morettin, pp. 10–28. Sao Paulo: Inst. Math. Stat. Univ. Sao Paulo

Nelsen RB. 2006. *An Introduction to Copulas*. New York: Springer. 2nd ed.

Paindaveine D. 2012. Elliptical symmetry. In *Encyclopedia of Environmetrics*, ed. AH El-Shaarawi, W Piegorsch, pp. 802–7. New York: Wiley. 2nd ed.

Pewsey A. 2008. The wrapped stable family of distributions as a flexible model for circular data. *Comp. Stat. Data Anal.* 52:1516–23

Pewsey A, Kato S. 2016. Parametric bootstrap goodness-of-fit testing for Wehrly–Johnson bivariate circular distributions. *Stat. Comput.* 26:1307–17

Pewsey A, Neuhäuser M, Ruxton GD. 2013. *Circular Statistics in R*. Oxford, UK: Oxford Univ. Press

Ramachandran GN. 1963. Stereochemistry of polypeptide chain configurations. *J. Mol. Biol.* 7:95–99

Ranalli M, Lagona F, Picone M, Zambianchi E. 2018. Segmentation of sea current fields by cylindrical hidden Markov models: a composite likelihood approach. *J. R. Stat. Soc. Ser. C* 67:575–98

Scrucca L, Fop M, Murphy TB, Raftery AE. 2016. mclust 5: Clustering, classification and density estimation using Gaussian finite mixture models. *R J.* 8:205–33

Sellers KF, Swift AW, Weems KS. 2017. A flexible distribution class for count data. *J. Stat. Distrib. Appl.* 4:22

Siffer A. 2018. `Rfolding`: the folding test of unimodality. *R Package*, version 1.0. **https://CRAN.R-project.org/package=Rfolding**

Singh H, Hnizdo V, Demchuk E. 2002. Probabilistic model for two dependent circular variables. *Biometrika* 89:719–23

Sklar M. 1959. Fonctions de répartition à *n* dimensions et leurs marges. *Publ. Inst. Stat. Univ. Paris* 8:229–31

Stigler SM. 1986. *The History of Statistics: The Measurement of Uncertainty Before 1900*. Cambridge, MA: Harvard Univ. Press

Tibshirani R, Taylor J, Lockhart R, Tibshirani R, Fithian W, et al. 2015. *Recent advances in post-selection statistical inference*. Breiman Lecture presented at NIPS 2015, the Twenty-Ninth Conference on Neural Information Processing Systems, Dec. 7–12, Montreal, Can.

Tukey JW. 1977. *Modern techniques in data analysis*. Presented at the NSF-Sponsored Regional Research Conference, Southern Massachusetts University, North Dartmouth, MA

Wang J, Boyer J, Genton MG. 2004. A skew-symmetric representation of multivariate distributions. *Stat. Sin.* 14:1259–70

Wehrly TE, Johnson RA. 1980. Bivariate models for dependence of angular observations and a related Markov process. *Biometrika* 66:255–56

Wraith D, Forbes F. 2015. Location and scale mixtures of Gaussians with flexible tail behaviour: properties, inference and application to multivariate clustering. *Comput. Stat. Data Anal.* 90:61–73

Wuertz D, Setz T, Chalabi Y. 2020. `fmultivar`: Rmetrics—analysing and modeling multivariate financial return distributions. *R Package*, version 3042.80.1. **https://CRAN.R-project.org/package=fMultivar**

Ye Y. 1987. *Interior algorithms for linear, quadratic, and linearly constrained non-linear programming*. PhD Thesis, Dep. ESS, Stanford University

Yeo IK, Johnson RA. 2000. A new family of power transformations to improve normality or symmetry. *Biometrika* 87:954–59

Annual Review of Statistics and Its Application

Volume 8, 2021

Contents

Errata

An online log of corrections to *Annual Review of Statistics and Its Application* articles may be found at http://www.annualreviews.org/errata/statistics