Flexible modelling in statistics: past, present and future

Christophe Ley[1]

Abstract: In times where more and more data become available and where the data exhibit rather complex structures (significant departure from symmetry, heavy or light tails), flexible modelling has become an essential task for statisticians as well as researchers and practitioners from domains such as economics, finance or environmental sciences. This is reflected by the wealth of existing proposals for flexible distributions; well-known examples are Azzalini's skew-normal, Tukey's *g*-and-*h*, mixture and two-piece distributions, to cite but these. My aim in the present paper is to provide an introduction to this research field, intended to be useful both for novices and professionals of the domain. After a description of the research stream itself, I will narrate the gripping history of flexible modelling, starring emblematic heroes from the past such as Edgeworth and Pearson, then depict three of the most used flexible families of distributions, and finally provide an outlook on future flexible modelling research by posing challenging open questions.

Résumé : Dans des temps où de plus en plus de données deviennent accessibles et où ces données sont de plus en plus complexes (asymétrie évidente, queues lourdes ou légères), la modélisation flexible est devenue une tâche essentielle pour les statisticiens ainsi que pour les chercheurs et praticiens de domaines tels que l'économie, la finance ou les sciences environnementales. Ceci est reflété par la richesse de propositions existantes pour des distributions flexibles ; des exemples connus sont la skew-normale d'Azzalini, la *g*-et-*h* de Tukey, des distributions de mixture ainsi que des distributions deux-morceaux, pour ne citer que celles-là. Mon but dans cet article est de donner une introduction à ce domaine de recherche, destinée à être utile à la fois pour des novices et des professionnels du domaine. Après une brève description du courant de recherche lui-même, je vais raconter l'histoire passionnante de la modélisation flexible, mettant en vedette des héros emblématiques comme Edgeworth et Pearson, puis je vais décrire trois familles de distributions flexibles qui figurent parmi les plus utilisées, et finalement donner un aperçu sur le futur de la modélisation flexible en posant des questions ouvertes stimulantes.

---

[1] ECARES and Département de Mathématique, Université libre de Bruxelles
Boulevard du Triomphe, CP210, 1050 Brussels, Belgium
E-mail: chrisley@ulb.ac.be

1. Introduction.

> "Everybody believes in the exponential law of errors : the experimenters, because they think it can be proved by mathematics; and the mathematicians, because they believe it has been established by observation." [1]
>
> Lippmann to Poincaré, in Poincaré (1896), p. 149

The "experimental law of errors" of course refers to the normal density

$$\frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

with location parameter $\mu \in \mathbb{R}$ and scale parameter $\sigma > 0$, and the above statement provocatively underlines the often assumed ubiquity of this distribution that is believed to represent the "normal" state of data. Indeed, its popularity has numerous sources: the Central Limit Theorem, the strong link to the empirical mean via maximum likelihood characterizations (result due to Gauss, 1809, see Duerinckx et al., 2014 for a recent detailed account), the maximum entropy characterization (see Cover and Thomas, 2006), the numerous well-studied and highly tractable stochastic properties, the bell shape of its curve, the simplicity inherent to the assumption of normality, etc. However, more and more empirical evidence has been provided over the years that the assumption of normality is often only a very poor representation of reality, and many data sets cannot be satisfactorily fitted by the normal distribution.

One nowadays very important domain, where the assumption of normality fails to hold, is finance. Due to the occurrence of extreme events, financial return data have a large amount of probability mass in their tails. Moreover, negative events are usually more extreme than positive events, entailing some form of asymmetry (or skewness) in the data. Both these effects, heavy tails and skewness, which are considered as stylized facts in the finance literature, cannot be captured by the normal distribution. I will now describe three further examples of data sets whose behavior goes beyond normality:

Example 1. *Stochastic Frontier Analysis (SFA) is concerned with the specification and estimation of a frontier production function, e.g., for firms. Economic modelling for SFA has been initiated simultaneously by Aigner et al. (1977) and Meeusen and van den Broeck (1977) and can be formulated as follows:*

$$Y = f(\mathbf{x};\boldsymbol{\beta}) + \varepsilon, \tag{1.1}$$

where Y is the observed scalar output, the production frontier f depends on the input $\mathbf{x} \in \mathbb{R}^k$ and some parameter $\boldsymbol{\beta} \in \mathbb{R}^p$ to be estimated, and ε is the error term. This term itself can be expressed as

$$\varepsilon = V - U \tag{1.2}$$

where V is a random shock, assumed to be symmetric, and U, independent of V, is the random non-negative technical (in-)efficiency component inherent to each firm. Now, the structure (1.2) clearly shows that the composed error term ε cannot follow a normal distribution, since it is the sum of a symmetric term (V) and a negative term ($-U$), leading to skewness in the error term.

---

[1] Original formulation in French: "Tout le monde y croit cependant, me disait un jour M. Lippmann, car les expérimentateurs s'imaginent que c'est un théorème de mathématiques, et les mathématiciens, que c'est un fait expérimental."
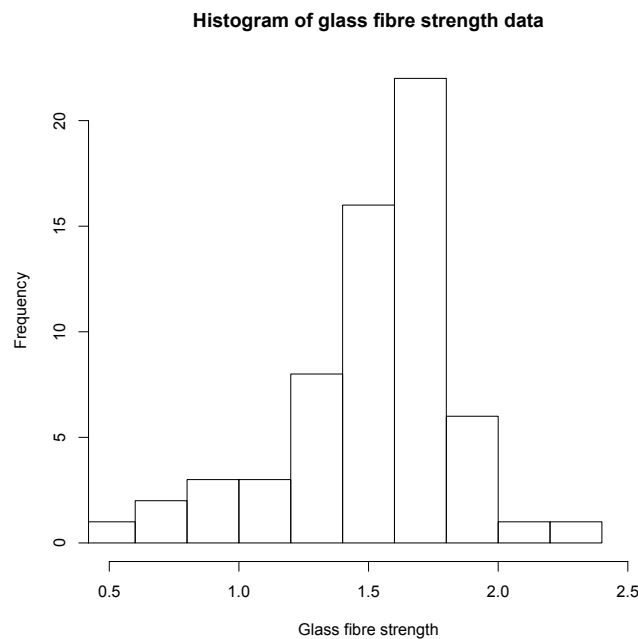
Histogram of glass fibre strength data

FIGURE 1. *Histogram of glass fibre strength data (Example 2).*

Example 2. *Smith and Naylor (1987) presented the breaking strengths of 63 glass fibres of length 1.5 cm, obtained by workers at the U.K. National Physical Laboratory. Determining the behavior of this breaking strength, especially in the tails, is of utmost importance, and it can readily be observed from Figure 1 that the normal distribution does not present an adequate fit. This explains why this data set is very popular among researchers working on new non-normal distributions.*

Example 3. *In order to get further insight into relationships in body dimensions, Heinz et al. (2003) have studied data from the so-called HP study (named after the first two authors of that paper). Diverse body measurements, such as weight and girth, have been taken on 247 men and 260 women primarily in their twenties and early thirties, all exercising several hours per week. A natural question of interest is which distribution fits well the distinct body measurements.*

For one particular measurement, namely waist girth, Heinz et al. (2003) remark "Note that in a well-nourished group the lower limit of waist girth will not fall more than a few centimeters below what can be expected from body build, but the upper limit of waist girth is determined by fatness in addition to body build". Consequently, these data are by nature skewed, as their right tail is heavier than the left tail, which excludes the normal as a possible fit for these data.

These examples further underline the clear need for distributions exhibiting features that the normal does not possess. In other words, the need for *flexible distributions*. Now, one may think that this is an easy task—heavy tails can be obtained by replacing the normal with the Student *t* distribution, and skewness can be attained by simply mixing two normal distributions. This is

still too much of a simplified vision, as these two parametric models also have their limitations, especially when one has to combine skewness and excess kurtosis in a specific way, as is the case with financial data. As a reaction to such limitations, one may want to consider models with a very large number of parameters in order to ensure as many different shapes as possible. Such a solution in turn suffers from a risk of over-fitting and complexity in calculations.

Thus the ideal solution is a compromise between both requirements: wide range of possible shapes and tractability. This is precisely the objective of the research stream called *flexible modelling*, and the topic of the present paper. The importance of flexible models is further stressed by their successful combination with classical statistical theories and tools such as time series analysis (Hansen, 1994), space-state models (Naveau et al., 2005), random fields (Allard and Naveau, 2007), regression models (Azzalini and Genton, 2008, where several other inferential tools are discussed, too), Bayesian statistics (Rubio and Steel, 2014) or hypothesis testing (see Section 4), to cite but a few.

The paper is organized as follows. In Section 2, I will shortly narrate the beginnings of flexible modelling, and then describe three of the nowadays most popular flexible families in Section 3. After this review of past and present, I will finish the paper with an outlook on the future of flexible modelling in Section 4 by posing (and discussing) six open questions and problems in this field.[2]

2. A brief historic account: flexible modelling before the 1980s

The systematic quest for non-normal distributions has been initiated at the end of the 19th century, when data sets exhibiting marked departures from normality have been collected, *inter alia* by the famous Belgian scientist Adolphe Quetelet. Due to the predominance and well-understood properties of the "normal theory", the first reflex at that time was to transform the data so as to make the resulting transformed data follow a normal distribution, and then apply the usual procedures. For instance, Galton used in 1879 the logarithmic transformation, which resulted in the nowadays well-known log-normal distribution. It was Francis Ysidro Edgeworth who, in 1898, formally developed this concept, which he coined "Method of Translation" instead of transformation. While Edgeworth's transformations were restricted to polynomial functions, the subsequent proposals emerging over the course of the 20th century were more diverse, the most notable contributions being the Johnson (1949) families and Tukey's *g*-and-*h* distributions from 1977, both of which will be described in the next section. Still nowadays, transformations are mainly applied to the normal distribution, for the reasons mentioned above, albeit the Student *t* and logistic have been shown to be good candidates, too.

The paper Edgeworth (1898) was not Edgeworth's first publication on the subject. Already in the 1880's was he interested in fitting non-normal data, and Edgeworth (1886) is commonly considered as the first systematic attempt to fit asymmetric distributions to asymmetric frequency

---

[2] In the same year 2014 where the present paper has been written, an excellent discussion paper, Jones (2015b), has reviewed diverse flexible models, and I had the privilege to figure among its discussants. I have tried my best to avoid overlaps with that paper, whose focus is on the univariate setting and on a structural comparison of four families of distributions; for instance, the paper Jones (2015b) contains no historical account nor an outlook on the future like the one from my Section 4. The two papers are meant to be complementary in their desire to present an increasingly important research domain.

data. A short time after Edgeworth's publication, another heavyweight of statistics, namely Karl Pearson, entered the scene of flexible modelling. His interest was raised by the zoologist Walter Weldon who collected data on 23 characteristics of 1000 female crabs during his holidays in Malta and the Bay of Naples and who noticed that one of these characteristics did not possess normal-like behavior. Pearson's first attempts resulted in a mixture of two normal distributions (hereby laying the foundations of mixture distributions) as well as in a "generalized form of the normal curve of an asymmetrical character" (this was actually a precursor of the Gamma distribution), both published in a letter to *Nature* (Pearson, 1893). This publication triggered a sort of competition between Edgeworth and Pearson. For instance, Edgeworth drew Pearson's attention to the fact that the Gamma-like curve had previously been derived by Erastus De Forest in De Forest (1882, 1883), a fact that Pearson acknowledged in a further letter to *Nature* (Pearson, 1895b).

Pearson's next attempt on skew frequency curves was even more groundbreaking: in Pearson (1895a), he defined several probability distributions as solutions to a particular differential equation. This seminal idea formed the basis of the thoroughly studied and still in modern days used *Pearson family of distributions* (see Johnson et al., 1994 for a detailed description of that family), which he further developed in Pearson (1901) and Pearson (1916). Among the 1895 distributions figures the so-called Pearson Type IV distribution, which contains as special case the Student t distribution; Pearson thus predates William Sealy Gosset's 1908 paper written under the pseudonym "Student". Edgeworth's reaction to Pearson's works was the already mentioned Edgeworth (1898) paper; the latter was further backed up by Kapteyn (1903), the second father of the transformation approach. On top of the Pearson-Edgeworth rivalry, a heated exchange started between Pearson and Kapteyn, each promoting his own family of skew curves and criticizing the other. The Pearson-Kapteyn dispute is reported in Stamhuis and Seneta (2009), while the competition between Edgeworth and Pearson and their constant correspondence has been mostly reprinted in Stigler (1978); see also the excellent Stigler (1986) for further historical details about that fruitful period.

Exactly at that same time appeared the *Kollectivmasslehre* by Carl Gustav Fechner (Fechner, 1897), the founder of psychophysics, the science studying the link between psychological sensations and physical stimulus. Fechner died in 1887, and it was Gottlob Friedrich Lipps who took on the heavy task of completing his manuscript, which explains the posthumous publication in 1897. In that work, Fechner proposed a skew curve by binding together two halves of normal curves, each having a different scaling. Fechner thus was the father of what we now call *two-piece distributions*. However, again Pearson vividly opposed himself to Fechner's idea, both from a statistical as well as historical viewpoint. Indeed, he erroneously believed that de Vries (1894) had made the same proposal. On statistical grounds, he argued that, compared to his family of curves, Fechner's was not general enough. This fierce opposition by Pearson [3] implied that Fechner's work was overlooked for a long time and eventually fell into oblivion. As a consequence, the

---

[3] Quite remarkably, Pearson attacked Fechner's curves in a paper from 1905 where he actually reacted to the work Ranke and Greiner (1904), two anthropologists who claimed that, for their domain, only the normal distribution mattered, and hence disqualified both Pearson's and Fechner's work. Pearson published his reaction in his own journal Biometrika, hereby criticizing Fechner, although both were on the same side w.r.t. Ranke and Greiner. Another remarkable fact: this paper, Pearson (1905), is most well-known as Pearson there introduced the terminology mesokurtic, leptokurtic and platykurtic.

wheel has been re-invented under the names "joined half-Gaussian", "three-parameter two-piece normal" and "binormal" distribution; see Sections 3 and 5 of Wallis (2014) for details about these rediscoveries (including a paper by Edgeworth) and many more details about two-piece distributions. It is to be noted that Fechner's contribution has been brought back to the statistical community mainly through Hald (1998).

At the beginning of the 20th century, a young Italian scholar, Fernando de Helguero, adopted a completely different viewpoint on *abnormal curves* as he called them. In two fundamental papers, de Helguero (1909a,b), he set up his own program to deal with non-normal data. He therein criticized Edgeworth's and Pearson's works by noting that, although clearly more valuable than the normal since they are broader, their proposals suffer from a major drawback: their constructions do not allow to understand which mechanism might have actually generated the data. To quote de Helguero on their non-normal curves: "are defective in my view because they only limit themselves to tell us that the infinitesimal elementary causes of variation are interdependent". His own view is that non-normal behavior only occurs subject to external perturbations, and abnormality is the upshot of some selection mechanism. The ensuing abnormal curve he derived is the pre-cursor of the celebrated *skew-normal distribution* which, re-appearing under several guises in e.g. Birnbaum (1950); O'Hagan and Leonard (1976); Aigner et al. (1977), has come to fame thanks to the seminal paper Azzalini (1985). See the review paper Azzalini (2005) for a more detailed account on these rediscoveries of the skew-normal, and Azzalini and Regoli (2012b) for a historic analysis of de Helguero's work.

The Azzalini (1985) paper had a phenomenal impact on flexible modelling (note, however, that it took until the late 1990s until this success story really started), as it is the starting point for a systematic treatment thereof with growing activity till today. This is why I decide to stop the historical account here [4]; for interested readers, besides the review papers I already cited above, I recommend the Kotz and Vicari (2005) survey paper, (the short) Section 3 of Pewsey (2015) as well as the introductions of the recent PhD theses of Juan Francisco Rosco from the University of Extremadura (Rosco (2012)) and of Francisco Javier Rubio from the University of Warwick (Rubio (2013)).

3. Description of the main families of flexible distributions

If we credit a distribution with flexibility as soon as its shape significantly diverges from that of the normal distribution, then of course the task of this section is near-impossible given the infinity of non-normal distributions. The historical developments of the previous section teach us, in my opinion, an important lesson to which I shall stick throughout the rest of the paper: flexible models ought to be distributions that, besides the usual location and scale/scatter parameters, possess either a skewness or a kurtosis parameter, or, optimally, both. Now this still leaves us with a plethora of distributions, ranging from the Pearson family to the hyperbolic and the Tukey lambda distribution by passing across the α-stable and generalized extreme value families.

I here "restrict" my attention to flexible modelling understood as modifying a given base density f, symmetric about the origin [5]. Most constructions nowadays follow this seemingly

---

[4] There would obviously be several further developments to relate such as, for example, the history of copulas; this would however shift away the focus of the present paper.

[5] In higher dimensions, symmetry of **X** with values in $\mathbb{R}^k$ shall mean, according to the situation, spherical symmetry

natural pattern. The key advantages of modifying a symmetric density f is that in doing so it is possible to retain some of the properties of f, which are often well known, and to pave the way for goodness-of-fit tests of the density f itself (see also OQC6 in Section 4). For example, if in the univariate case two parameters are added, f is mostly normal; in case a single skewness parameter is added, f is mostly the Student t density. Given its importance in the sequel, I provide the latter density in the k-variate case:

$$t_{\boldsymbol{\mu},\boldsymbol{\Sigma},\nu}(\mathbf{x};k) := \frac{\Gamma\left(\frac{\nu+k}{2}\right)}{(\pi\nu)^{k/2}\Gamma\left(\frac{\nu}{2}\right)}|\boldsymbol{\Sigma}|^{-1/2}\left(1+\|\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{x}-\boldsymbol{\mu})\|^2/\nu\right)^{-\frac{\nu+k}{2}}$$

with location parameter $\boldsymbol{\mu}\in\mathbb{R}^k$, scatter parameter $\boldsymbol{\Sigma}\in S_k$, the class of symmetric and positive definite $k\times k$ matrices, and tail-weight parameter $\nu\in\mathbb{R}_0^+$, and where the Gamma function is given by $\Gamma(z)=\int_0^\infty \exp(-t)t^{z-1}\,dt$. Throughout this section, $\boldsymbol{\mu}$ and $\sigma/\boldsymbol{\Sigma}$ are location and scale/scatter parameters, respectively.

In what follows, I will depict three of the most used families of flexible distributions, namely *Azzalini-type distributions*, *transformation-approach distributions* and *two-piece or scale-transformed distributions*; the order of presentation is in line with Jones (2015b). Evidently, this enumeration could easily be extended by other very popular flexible modelling families and tools such as finite mixture models (McLachlan and Peel, 2000), variance-mean mixtures (Barndorff-Nielsen et al., 1982), copulas (Nelsen, 2006), power transformations including the Box-Cox transformation (Box and Cox, 1964), order-statistics-based distributions (Jones, 2004), the very general probability integral transformations of Ferreira and Steel (2006) or the classical Pearson system of distributions. However, such a broad description would be far beyond the scope of the paper and dilute its main focus of providing a concise idea of the flexible modelling research. Also, I here do not consider data on supports other than $\mathbb{R}^k$, although much could be said about data on finite or semi-finite intervals (e.g., the logarithmic and power transformations) or on directional data (see Section 4). For further general information on flexible distributions, see my encyclopedia contribution Ley (2012), whose focus is solely on skew distributions, the contribution Lee et al. (2013) and the excellent discussion paper Jones (2015b), where four flexible families are compared in terms of their stochastic and statistical properties. To avoid any redundancy, I will not proceed to such a comparison and confine myself to a pure description, which is intended to differ as much as possible from the aforementioned references.

3.1. Family 1: Azzalini-type distributions

The construction underlying this first family of distributions is purely of a skewing nature, although tails can partly be affected, see Section 3.1 of Ferreira and Steel, 2006. This nature of Family 1 is underpinned by the terminology "modulating symmetry" advocated by Adelchi Azzalini in recent years. The modulation can well be seen from (3.3) below: the symmetric part is multiplied by a *skewing function*. This pure-skewing construction explains why the skew-t distribution defined in (3.4) is a popular choice in flexible modelling. Speaking of popularity, Family 1 has encountered an enormous success over the past decades, with an incredible number of papers dedicated to

---

($\mathbf{X}\overset{d}{=}\mathbf{OX}$ for all orthogonal $k\times k$ matrices $\mathbf{O}$), elliptical symmetry ($\mathbf{X}\overset{d}{=}\mathbf{AY}$ where $\mathbf{Y}$ is spherically symmetric and $\mathbf{A}$ is a full-rank $k\times k$ matrix) and central symmetry ($\mathbf{X}\overset{d}{=}-\mathbf{X}$).

it. Reasons for this are good stochastic properties, nice generating mechanisms (see the part on de Helguero of the previous section), good fitting properties and some unfortunate inferential problems that researchers have tried to explain and to solve/circumvent. This cocktail of attributes has helped to forge the fame of what one may call Azzalini-type distributions. It is therefore also not surprising that there is nearly no ambiguity when speaking about *the* skew-normal distribution: it is Azzalini's from 1985 (his systematic treatment and study of the law makes him the father of the skew-normal, despite previous similar proposals as mentioned in Section 2).

At the core of this first family figures evidently the (scalar) skew-normal density

$$2\sigma^{-1}\phi\left(\frac{x-\mu}{\sigma}\right)\Phi\left(\delta\left(\frac{x-\mu}{\sigma}\right)\right)$$

with $\delta \in \mathbb{R}$ the skewness parameter and where ϕ and Φ respectively denote the standard normal density and cumulative distribution function (cdf). The role of δ as skewness parameter becomes most apparent by noting that only at $\delta = 0$ the skew-normal density is symmetric, namely one retrieves the normal density. The multivariate skew-normal has been proposed some ten years later in Azzalini and Dalla Valle (1996) and further studied in Azzalini and Capitanio (1999). Subsequent generalizations include skew-elliptical (skewing an elliptically symmetric density), generalized skew-elliptical and, most generally, *skew-symmetric* distributions. The latter, proposed by Azzalini and Capitanio (2003) and Wang et al. (2004), admit densities of the form

$$2|\boldsymbol{\Sigma}|^{-1/2}f\left(\boldsymbol{\Sigma}^{-1/2}(\mathbf{x}-\boldsymbol{\mu})\right)\Pi\left(\boldsymbol{\Sigma}^{-1/2}(\mathbf{x}-\boldsymbol{\mu}),\boldsymbol{\delta}\right) \tag{3.3}$$

where f is centrally symmetric and $\Pi : \mathbb{R}^k \times \mathbb{R}^k \to [0,1]$ is a skewing function satisfying $\Pi(\mathbf{y},\boldsymbol{\delta}) + \Pi(-\mathbf{y},\boldsymbol{\delta}) = 1$ and $\Pi(\mathbf{y},\mathbf{0}) = 1/2$ for all $\mathbf{y},\boldsymbol{\delta} \in \mathbb{R}^k$ (for diverse choices of Π and resulting skew-f distributions, see Hallin and Ley, 2014). For a very recent account on these distributions, I refer to the book Azzalini and Capitanio (2014).

Besides these general proposals, parametric skew-f distributions have as well emerged, such as the skew-Cauchy, skew-exponential power, skew-logistic and, most prominently, the skew-t distribution (under various forms). I here consider the multivariate skew-t as defined in Azzalini and Capitanio (2003), with density

$$2t_{\boldsymbol{\mu},\boldsymbol{\Sigma},\nu}(\mathbf{x};k)T_{0,1,\nu+k}\left(\boldsymbol{\delta}'\boldsymbol{\sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\left(\frac{\nu+k}{||\boldsymbol{\Sigma}^{-1/2}(\mathbf{x}-\boldsymbol{\mu})||^2+\nu}\right)^{1/2};1\right), \tag{3.4}$$

where $\boldsymbol{\sigma}$ is a $k \times k$ diagonal matrix with diagonal entries $\boldsymbol{\sigma}_{ii} = \boldsymbol{\Sigma}_{ii}^{1/2}$, $i = 1,\ldots,k$, and $T_{\mu,\sigma,\eta}(\cdot;1)$ is the cdf of the univariate Student density $t_{\mu,\sigma,\eta}(\cdot;1)$, $\eta > 0$. This skew-t distribution is often used for modelling purposes, as it incorporates both a skewness and a tail-weight parameter; see Azzalini and Genton (2008) for an overview of statistical procedures involving the skew-t. For the sake of illustration, I present in Figure 2 various one-dimensional skew-normal and skew-t densities.

The aforementioned good properties of skew-symmetric distributions can be found in the references I have given above; see also Azzalini and Regoli (2012a). In particular, the second part of the monograph Genton (2004) contains numerous real data situations where skew-symmetric
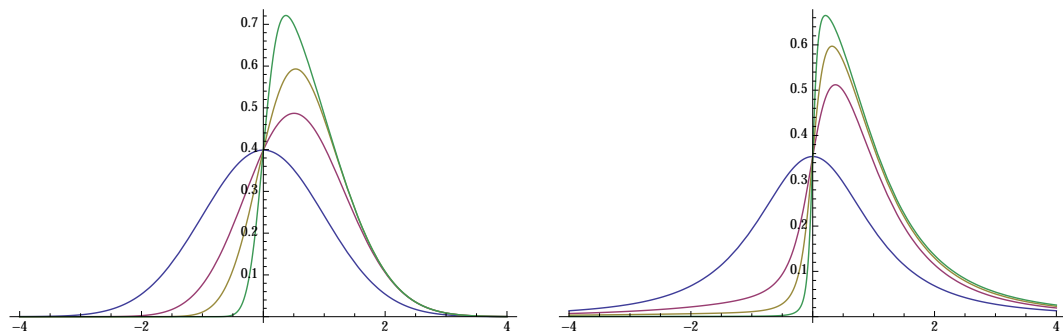
FIGURE 2. *Plots of the original* Azzalini *(1985) skew-normal density* $2\phi(x)\Phi(\delta x)$ *(left) and of the* Azzalini and Capitanio *(2003) skew-t density* (3.4) *with* $\mu = 0, \sigma = 1, \nu = 2$ *(right), for* $\delta = 0$ *(blue), 1 (red), 2 (yellow), and 5 (green).*

distributions are put to use. Regarding physical generating mechanisms or stochastic representations, I refer the reader to Azzalini (2005) or to the discussion paper Arnold and Beaver (2002) where especially hidden truncation and selective sampling are treated. I conclude this section by briefly discussing the peculiar inferential issues related to Family 1.

First, in the vicinity of symmetry, that is, at $\boldsymbol{\delta} = \mathbf{0}$, certain subfamilies of densities (3.3) suffer from a singular Fisher information matrix (w.r.t. $\boldsymbol{\mu}, \Sigma, \boldsymbol{\delta}$) due to a (partial) collinearity between location and skewness scores. The most prominent candidate is the skew-normal itself, which is the main reason for the large number of publications on the topic, some trying to explain the phenomenon, others to determine all skew-symmetric distributions suffering from that singularity [6], and again others to provide reparameterizations avoiding the phenomenon. The question of "who is infected" has been solved in the series of papers Ley and Paindaveine (2010c,b) and Hallin and Ley (2012), while suggested remedies are the centred parameterization of Azzalini (1985), extended to higher dimensions in Arellano-Valle and Azzalini (2008), and the Gram-Schmidt-based reparameterization put forward in Hallin and Ley (2014). In the latter paper, we also classify distributions according to their degrees of singularity (how many orthogonalization and reparameterization steps are necessary to obtain a non-singular matrix), hereby providing intriguing and entertaining insights into the structure of skew-symmetric distributions. Second, for some data sets (even for some simulated data from a skew-normal with finite skewness $\boldsymbol{\delta}$), the maximum likelihood estimate of $\boldsymbol{\delta}$ becomes infinite, that is, lies at the frontier of the parameter space. This anomaly is well described in Section 6.3 of Azzalini and Capitanio (1999), and as remedy Azzalini and Arellano-Valle (2013) suggest penalized maximum likelihood estimation. In the literature there is no consensus on whether this is really an issue or simply a natural artefact (e.g., Jones, 2015b considers the problem as being "overblown").

---

[6] Closely related to the Fisher information singularity is another singularity, namely a stationary point at $\boldsymbol{\delta} = \mathbf{0}$ in the profile log-likelihood function for skewness.

3.2. *Family 2: Transformation-approach distributions*

As mentioned at the beginning of Section 2, the initial incentive behind the transformation approach was the predominance of the normal distribution, and, if data were obviously not normally distributed, then one should strive to find a transformation H such that $H(\text{data}) \approx \mathcal{N}(\mu, \sigma^2)$. The success of the transformation approach has entailed that nowadays the transformations *per se* are interesting, as they lead to very diverse shapes in the resulting density. Indeed, compared to Family 1 and 3, Family 2 is the most flexible, thanks to the freedom of choice in H. In the subsequent lines, I will comment on some of the most used such transformations.

In their most general k-variate form, transformation-approach densities built upon some base density f take on the guise

$$|\mathbf{\Sigma}|^{-1/2} f\left(H_{\boldsymbol{\delta},\boldsymbol{\eta}}\left(\mathbf{\Sigma}^{-1/2}(\mathbf{x}-\boldsymbol{\mu})\right)\right) \left|DH_{\boldsymbol{\delta},\boldsymbol{\eta}}\left(\mathbf{\Sigma}^{-1/2}(\mathbf{x}-\boldsymbol{\mu})\right)\right|, \qquad (3.5)$$

where $\boldsymbol{\delta} \in \mathbb{R}^k$ is a skewness parameter and $\boldsymbol{\eta} \in (\mathbb{R}_0^+)^k$ a tail-weight parameter [7], and where the transformation $H_{\boldsymbol{\delta},\boldsymbol{\eta}} : \mathbb{R}^k \to \mathbb{R}^k$ is a monotone increasing diffeomorphism with $|DH_{\boldsymbol{\delta},\boldsymbol{\eta}}(\cdot)|$ representing the absolute value of the determinant of the Jacobian matrix DH. To avoid any ambiguity, I prefer to stress that H-transformed densities, let us write them f_H, are obtained by transforming $\mathbf{X} \sim f$ into $\mathbf{Y} \sim f_H$ via the (inverse) mapping $H_{\boldsymbol{\delta},\boldsymbol{\eta}}^{-1}$ so that $H_{\boldsymbol{\delta},\boldsymbol{\eta}}(\mathbf{Y}) \sim f$. Densities (3.5) have been studied in detail in Ley and Paindaveine (2010a) with particular focus on the skewness parameter. For instance, one can easily check that (3.5) is symmetric iff $H_{\boldsymbol{\delta},\boldsymbol{\eta}}(\cdot)$ is odd. Changes to tail-weight, in terms of largest finite directional moments, can be explicitly calculated via Theorem 3.2 in Ley and Paindaveine (2010a). In that same paper, we have identified the minimal requirements on $H_{\boldsymbol{\delta},\boldsymbol{\eta}}(\cdot)$ to satisfy the surjectivity property, meaning that every random k-vector $\mathbf{Y}$ can be obtained from any $\mathbf{X}$ via one such transformation. Another popular approach to define multivariate versions of existing univariate transformation-approach distributions consists in transforming each marginal with a one-dimensional transformation $H_{\delta,\eta}$ and introducing some dependence structure by means of a correlation matrix.

The classical Johnson (1949) paper applies the arcsinh transformation to the normal distribution to obtain the Johnson-S_U distribution, Tukey (1977) transforms the normal via $H_{\delta,\eta}^{-1}(x) = \frac{1}{\delta}(\exp(\delta x) - 1)\exp(\eta x^2/2)$ to define his famous g-and-h distributions ($g = \delta$ and $h = \eta$), later extended to higher dimensions in Field and Genton (2006), Rieck and Nedelman (1991) replace Johnson's arcsinh transformation with the sinh transformation, whereas Jones and Pewsey (2009) combine both in their sinh-arcsinh (SAS) transformation $H_{\delta,\eta}(x) = \sinh(\eta \operatorname{arcsinh}(x) + \delta)$ which they also generalize to the multivariate setting. Transformations that affect only tail-weight are, among others, the K-transformation $H_\eta^{-1}(x) = x(1 + x^2)^\eta$ of Haynes et al. (1997) and the E-, J-transformations of Fischer and Klein (2004). For further examples of transformations, see Rosco (2012) and Rubio (2013). One may wonder why, in those examples, I switched between writing out H and H^{-1}. This is due to the fact that Tukey-type transformations (the g-and-h as well as the K-, E-, J-transformations) do not admit an inverse, hence their density cannot be written out properly and the parameters are often estimated by quantile-based methods. Maximum likelihood estimation, the typical procedure for Family 2, has been examined for Tukey-type distributions

---

[7] Note that neither the skewness nor the tail-weight parameter need be of dimension k but could be either larger or smaller.
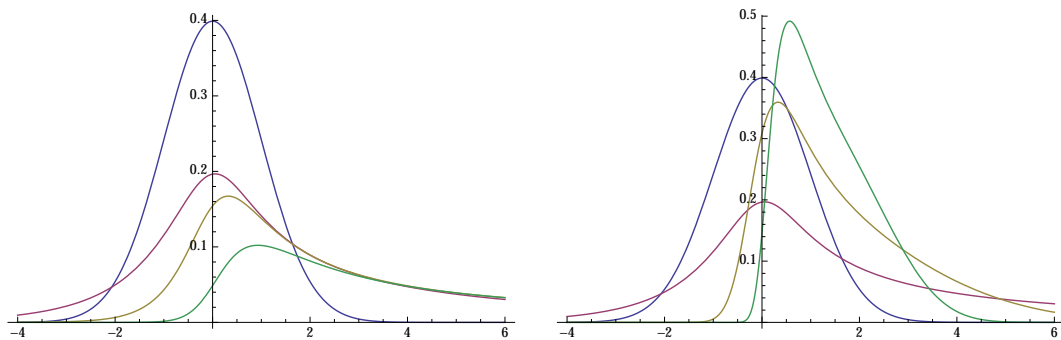
FIGURE 3. *Plots of the SAS-normal of Jones and Pewsey (2009) with $\mu = 0, \sigma = 1$ and, on the left picture, $(\delta, \eta) = (0,1)$ (blue), (-0.5,0.5) (red), (-1,0.5) (yellow) and (-1.5,0.5) (green) and, on the right picture, $(\delta, \eta) = (0,1)$ (blue), (-0.5,0.5) (red), (-1,1) (yellow) and (-1.5,1.5) (green).*

in Rayner and MacGillivray (2002). As an illustration, I provide in Figure 3 diverse shapes of SAS-normal densities.

3.3. Family 3: Two-piece or scale-transformed distributions

The third family of distributions is, like Family 1, a pure skew family of distributions, the tail-weight being again regulated by applying this construction to e.g. the Student t distribution. While Family 1 can be best understood via the expression "modulating symmetry", Family 3 can be described as follows: take a symmetric distribution, change the weight on both sides of the center (re-scaling operation), and glue both parts together! This nice interpretation definitely represents an advantage of these distributions pioneered by Fechner (1897). It is therefore not astonishing that this intuitive model has been applied by the Bank of England and the Sveriges Riksbank in their probabilistic forecast of future inflation to show that the deviation from the central forecast might be asymmetric [8]. In recent years, researchers even went one step further in generality by adding a scaling function to the initial density f, hence extending the image from above by now re-weighting continuously all points on the real line. This is why I opted for the title "Two-piece or scale-transformed distributions"; it is also for the same reason that Chris Jones in Jones (2015b) categorizes them as Family 3 and Family 3A.

Under their most general form, univariate two-piece-f distributions possess densities of the form

$$\frac{a(\delta)}{\sigma} \begin{cases} f\left(s_\ell(\delta)\frac{(x-\mu)}{\sigma}\right) & \text{if } x < \mu \\ f\left(s_r(\delta)\frac{(x-\mu)}{\sigma}\right) & \text{if } x \geq \mu, \end{cases} \tag{3.6}$$

with $\delta \in \mathbb{R}$ the skewness parameter, $s_\ell(\cdot)$ and $s_r(\cdot)$ left- and right-scaling functions, and resulting

---

[8] According to Wallis (2014), this application brought the two-piece normal distribution to "public attention in the late 1990s".
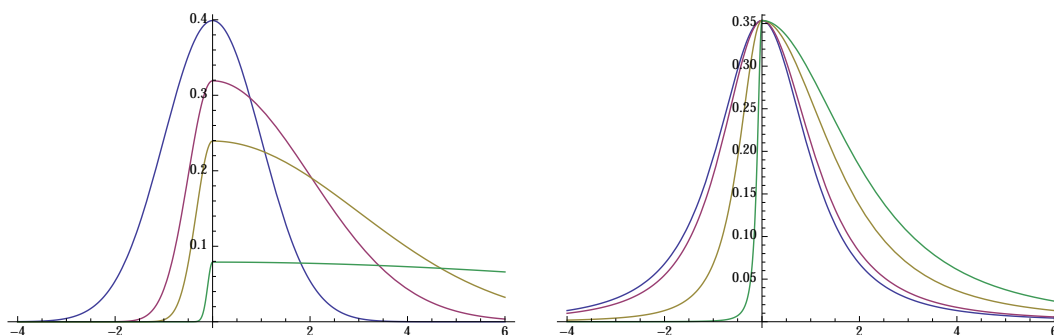
FIGURE 4. *Plots of the ISF-skew-normal (Fernández and Steel, 1998) density with $\mu = 0, \sigma = 1$ (left) and of the epsilon-skew-t (Hansen, 1994) density with $\mu = 0, \sigma = 1, \nu = 2$ (right), for $\delta = 1$ and 0 (blue), 2 and 0.1 (red), 3 and 0.5 (yellow), and 10 and 0.9 (green).*

normalizing constant

$$a(\delta) := \frac{2}{\frac{1}{s_\ell(\delta)} + \frac{1}{s_r(\delta)}}.$$

The probability mass on the left side of μ is $s_r(\delta)/(s_\ell(\delta) + s_r(\delta))$, while on the right side it is $s_\ell(\delta)/(s_\ell(\delta) + s_r(\delta))$. This rescaling on both sides is the idea put forward by Fechner (1897) with f the normal density, and the general construction (3.6) has been examined in Arellano-Valle et al. (2005). The reader is thus referred to that paper for the numerous interesting and good properties of two-piece distributions. One particular choice, suggested in Hansen (1994) in combination with the Student t and in Mudholkar and Hutson (2000) with the normal distribution, corresponds to $s_\ell(\delta) = 1/(1-\delta)$ and $s_r(\delta) = 1/(1+\delta)$ for $\delta \in (-1,1)$; conveniently, $a(\delta) = 1$ in that case. Mudholkar and Hutson (2000) term the resulting two-piece-skew-normal distribution *epsilon-skew-normal* (they use ε as parameter instead of my δ), and investigate its properties. Another famous representative of (3.6) is the *inverse scale factors (ISF)* model introduced by Fernández and Steel (1998). As the well-chosen name indicates, $s_\ell(\delta) = 1/s_r(\delta) = \delta$ for $\delta > 0$, yielding $a(\delta) = 2/(\delta + 1/\delta)$. Multivariate extensions of the latter choice, built along the same lines as the marginal-wise transformations in Family 2, are studied in Bauwens and Laurent (2005) and Ferreira and Steel (2007b). As for the other flexible families, an abundance of papers besides those mentioned above also propose parametric two-piece distributions, such as skew-Laplace or skew-exponential power distributions; see Wallis (2014) for references. Figure 4 contains density plots of ISF-skew-normal and epsilon-skew-*t* distributions (the names should by now be self-explanatory).

For inferential aspects of two-piece distributions, I refer the reader to Jones and Anaya-Izquierdo (2011), where it is shown that two-piece distributions actually enjoy strong parameter orthogonality compared to other flexible four-parameter models, Cassart et al. (2008) for a Le Cam approach and the various cited papers involving Mark Steel for insightful Bayesian aspects.

Scale-transformed distributions, the "younger brothers of two-piece distributions", are actually strongly linked to Family 2, as can be read from their densities

$$2f(H^{-1}(x)), \tag{3.7}$$

where the mapping $H : \mathbb{R} \to \mathbb{R}$ satisfies $H(x) - H(-x) = x$ (for simplicity I have exchanged here the roles of H and H^{-1} compared to Family 2). This strongly resembles (3.5) but without a Jacobian; the condition on H is the cause thereof. Note that the latter condition, when differentiated, yields $H'(x) + H'(-x) = 1$, which is exactly the requirement on skewing functions in Family 1. This again is no coincidence: Families 1-3 are linked, as nicely explained in Jones (2015b). Densities of the form (3.7) have been investigated and brought to the statistical community in Jones (2010) and, mainly, Jones (2014) and Fujisawa and Abe (2015); see those papers for their properties and for further references, especially regarding the classical Cauchy-Schlömilch transformation underpinning the entire construction. Finally, multivariate extensions of (3.7) are analysed in Fujisawa and Abe (2014), following again the independent-component-approach described for Family 2, and Jones (2015a) via marginal transformations from multivariate skew-symmetric distributions of Family 1.

4. The future of flexible modelling: open questions and challenges

The number of existing proposals for flexible distributions has become huge over the past decades and does not cease to increase. Papers proposing new skew and heavy-/light-tailed distributions have become common currency in the statistical literature, which is both good and bad. Good because each article does contribute to the repository of knowledge on statistical modelling, some with very strong impact on the community, and bad because it gets more and more complicated for the practitioner to choose the "best" model among this zoo of new proposals. This is why, besides continuing to improve on existing models, we should also use the current momentum of interest in flexible modelling to lift this research domain onto a new level by structuring the collection of distinct models (critical comparison of existing families both on theoretical and practical grounds) and hence make this research stream more easily accessible and applicable to people outside the community, be it statisticians, researchers from other disciplines or practitioners. Jones (2015b) has successfully started this task, and I hope to further contribute to it by addressing more directly what open problems and research questions should be tackled in the future.

I of course by no means claim to be in a position to say which are the main future problems in flexible modelling! However, from several personal reflections, readings of recent papers and discussions with colleagues from the domain (*inter alia* T. Abe, A. Azzalini, A. Bücher, H. Dette, Y. Dominicy, M. Hallin, M. C. Jones, A. E. Koudou, D. Paindaveine, A. Pewsey, J. Rubio, M. Steel), I have set up the following list of open questions and challenges (OQC's) that are worth being addressed in the future.

OQC1: What desirable properties should a flexible model possess?

The history of flexible modelling reveals that the fathers of this research domain had very divergent views on what is a "good flexible distribution". For instance, de Helguero's viewpoint on the necessity of a generating mechanism allowing to explain data generation, his main criticism to Edgeworth's and Pearson's proposals, has, in turn, been criticized in Pretorius (1930) with the following words: "The superiority of one frequency function over another depends rather on the success with which that function can be applied to graduate data than on the manner in which it originated"; see Section 1.4 of Johnson (1949) for further details.

Given this discrepancy, it is evident that even today we have divergent opinions on what would be good properties that a flexible model should ideally possess. Nevertheless, I undertake the endeavor of writing out what I consider as desirable structural properties (I will not consider fitting issues, see OQC2 for that matter).

– *finite number of well interpretable parameters:* optimally, one skewness and one tail-weight parameter are added to location and scale/scatter (for the sake of clarity: the term "one" does not stand for the dimension of the parameter). Evidently, in mixture models one also has to take into account the mixing parameter. A larger number of parameters entails the risk of a loss in interpretability.

– *well-separated roles for skewness and tail-weight parameters:* Measures of skewness such as those proposed in Arnold and Groeneveld (1995) and Critchley and Jones (2008) should ideally not depend on the tail-weight parameter. Two-piece constructions are particularly viable in this respect. Similarly, kurtosis measures, based on quantiles (e.g., Balanda and MacGillivray, 1988) or on moments (e.g., Ley and Paindaveine, 2010a, Theorem 3.2), ought not be effected by the skewness parameter; here the SAS-transformation constitutes a good choice. For further information on this topic, see Rubio et al. (2014) (where the van Zwet ordering is also discussed) and Jones et al. (2011), respectively.

– *good parameter estimation and tractability:* it is of utmost importance to be able to correctly estimate the diverse parameters of a model and then to produce calculations with it, e.g. to predict the risk of exceeding a certain threshold value.

– *clear stochastic properties:* this point is related to clear roles for each parameter and good tractability. It is desirable for any distribution to display well-identified conditions as to when, for example, a distribution is uni- or bi-modal. In the multivariate case, properties like closure under marginalization seem also desirable.

– *varying shapes:* optimally, a flexible model is able to exhibit as many distinct shapes as possible. For instance, we may wish that a distribution covers (nearly) all values of AG-skewness or of a given kurtosis measure.

– *good inferential properties:* besides good parameter estimation, a flexible model should not be prone to inferential problems prohibiting its use as, for instance, alternative in tests for normality (see OQC6). This explains the numerous attempts to remove the singularities linked to Family 1. Parameter orthogonality is, in this respect, an essential attribute of a flexible model (see Jones and Anaya-Izquierdo, 2011 or Ley and Paindaveine, 2015).

– *data generating mechanism:* although in the discussion Ley and Paindaveine, 2015 I was still reluctant towards the usefulness of this criterion, I have in the preparation of the present paper changed my mind and now also consider it as an advantage to be able to explain what causes might have led to a certain behavior of the data and not to confine myself to just finding the best fitting model. The delicateness of the question is well underlined by the Pretorius citation above.

I recommend the reader to also go through the criteria of Jones (2015b) and the related discussions. Although formulated under a different form, there is a strong overlapping between my and Chris Jones' criteria, indicating that these are indeed important requirements (their validity is further underlined by the opinions of the various discussants of Jones (2015b). Anticipating OQC4, good new flexible distributions should thus fulfill (most of) these desiderata. In line with the present OQC, I also recommend the reading of Ferreira and Steel (2007a).

OQC2: The curse of abundance: which distribution to be used when?

Besides the theoretical comparison of flexible families, another crucial issue consists in providing hints on what family ought to be used with what type of data. Of course, this question is extremely difficult to answer, and I will try to formulate diverse ways of proceeding.

–   The classical procedure would consist in choosing a set of flexible models, say, the skew-normal, skew-t, two-piece-t and SAS-logistic, and then use model selection criteria like likelihood maximization, the Akaike information criterion or the Bayesian information criterion. The best model will readily be obtained, sometimes at the cost of complex numerical maximizations. However, if the initial sub-selection is performed in an arbitrary way, then we may only optimize our fitting among distributions that present a low degree of fitting compared to other models. Moreover...

–   ... the interesting recent paper Charemza et al. (2013) has shown that distinct skew-normal distributions may be indistinguishable in certain circumstances! Their solution in such a case, from the practitioner's perspective, is to select the model on grounds of parameter interpretation, stressing the importance of this requirement in my listing in OQC1.

–   Some problems very naturally guide us towards the flexible model from which the data have (probably) been generated. Coming back to two examples of the Introduction, it is clear by construction that the error term in Example 1 stems from Family 1 (this is an immediate exercise for the interested reader), while the waist girth data from Example 3 can be well-described by members of Family 2 which allows to control left and right tail-weight. Advice of experts from the domains in which the data originate is evidently helpful.

–   In our discussion Ley and Paindaveine, 2015, we have added a further criterion to the list of Jones (2015b) and we term it "Testability: natural or satisfactory goodness-of-fit tests for the considered family can be defined". If satisfied, this criterion (which really means testing the validity of a given family, not testing for the best distribution within a given family!) would of course help to answer OQC2. In our discussion, we have shown that two-piece distributions actually do satisfy the criterion and provided a testing strategy based on the idea of de-scaling both half-distributions. In other families, the situation is more delicate. In Family 1, one can resort to the property that any even function of a skew-symmetric random vector annihilates the effect of the skewing function, leaving only the symmetric part; working with evenly transformed data, the symmetric part can thus be estimated, and the skewing function shall then be detected in a subsequent step within a collection of choices (it appears that, except for the bad matchings identified in Hallin and Ley, 2012, the choice of skewing function does not so much influence the final shape of the distribution, see Umbach, 2007). Finally, in Family 2, I would suggest transforming the data with several choices of $H_{\delta,\eta}$, apply some symmetry or normality test and retain that transformation with the highest p-value (evidently, this is not a formal procedure). Of course, such procedures are again subject to the criticism mentioned at the first point of this OQC and, contrarily to Family 3, I do not provide a Family-membership-test; my ideas are merely first steps into that direction. Another natural option would be to check membership by means of kernel density estimation under a specified-family form versus no specified form.

Even if OQC2 appears very tough, and near-impossible given the diversity of data sets even within the same domain, we should nevertheless aim to provide a list of recipes for which distribu-

tion to use in which circumstance. Such hints would definitely present an important contribution to flexible modelling and its use in other disciplines, especially in finance (what is, in the end, the best skew-heavy-tailed distribution to be combined with GARCH models?), economics and environmental sciences.

OQC3: Improved multivariate modelling via flexible copulas with flexible marginals?

The main appeal of copulas, and the reason for their enormous success story over the past years, lies in the fact that they model separately the dependence structure of multivariate data and the behavior of the individual marginals via the celebrated Sklar theorem of Sklar (1959), which renders them very attractive, especially in financial areas such as risk management and asset pricing (see, e.g., Cherubini et al., 2004). The success of copulas has been further stimulated by the introduction of vine copulas which are tailored for different forms of dependence between the variables (see Kurowicka and Joe, 2010). Thus, it is completely appealing to aim at a stronger combination of both research domains [9]: *flexible copulas* for the dependence structure, and flexible univariate distributions for the margins. This clearly generalizes the dependence structure described for certain members of Family 2 and 3, where instead of a full copula only a correlation matrix is added to independent component models. A neat treatment of OQC3 will most certainly result in promising constructions and should have a strong impact also outside statistics, especially in finance. For further recent information about copulas, I refer to the special volume 154(1) of the Journal de la Société Française de Statistique dedicated to this topic.

OQC4: Holy grail question: does THE flexible distribution exist?

Very naturally, each person has his/her personal favorite flexible model, and divergent viewpoints on the subject have helped shape this research landscape (recall the stormy beginnings of flexible modelling!). Such heated debates, of course under a less aggressive form, would certainly also nowadays push forward the quest for the ultimate flexible model that could be used for "all" purposes and would satisfy "all" requirements one has on a flexible distribution. Can such a universal distribution exist, or is it at least possible to come close to it? And if so, how could this discovery be achieved?

One potential solution might be to efficiently combine distinct flexible proposals. Rubio et al. (2014) combine the symmetric SAS transformation with (i) the two-piece construction and (ii) Azzalini-type skewing functions, while Steel and Rubio (2015) propose merging Ferreira and Steel (2006) with Ley and Paindaveine (2010a); further investigations in this direction, especially under the form of mixture distributions with flexible components, should allow to go beyond the state-of-the-art.

OQC5: Flexible modelling on other supports?

Besides the multivariate setting, a further challenge consists in consolidating respectively developing flexible distributions on supports other than $\mathbb{R}^k$; the word "consolidation" is meant for positive data and data on the unit circle, "developing" for data on the unit hypersphere $\mathscr{S}^{k-1} := \{\mathbf{x} \in \mathbb{R}^k : \mathbf{x}'\mathbf{x} = 1\}, k > 2$. Indeed, somehow paradoxically, while numerous circular

---

[9] The 2013 conference "Non-Gaussian Multivariate Statistical Models and their Applications" precisely followed this goal.

($k = 2$) flexible models have recently been put forward under the impetus of researchers like Abe, Jones, Kato and Pewsey, the same cannot be said when $k > 2$ (see Ley and Verdebout, 2014, one of the rare contributions).

OQC6: Improved tests for normality and symmetry via flexible models?

Testing for normality and symmetry is a classical topic in statistics. Famous normality tests are the Jarque-Bera, Shapiro-Wilk and Lilliefors tests, univariate symmetry is recurrently tested by means of Kolmogorov-Smirnov- and Cramér-von Mises-type tests; interesting alternative symmetry tests are the triples test of Randles et al. (1980) or the runs test of McWilliams (1990). For tests of multivariate symmetry under its different forms, see Serfling (2006).

It is clear that flexible models, besides their role of "improved description of reality", should also serve the purpose of insightful alternatives to the null hypotheses of normality and symmetry. Their specific forms allow to test for varying departures from normality (most naturally under a likelihood ratio form), be it in terms of skewness or kurtosis. Razali and Wah (2011) provide a broad power comparison of several classical normality tests, while Jones and Pewsey (2009) compare the performance of SAS-based normality tests to classical competitors. It would be of great interest to build likelihood ratio tests for normality within each flexible family, and then compare the overall performance of the resulting tests. While it is obvious that the Family-ℓ-test is the best under Family ℓ, it is far from clear if one test is able to exhibit good power under general alternatives. If such a test can be found, it is highly probable that this test can become a competitor for the above-mentioned classical tests.

The same holds true when testing for symmetry. Tests that behave optimally against a given family have been studied, e.g., for two-piece distributions in Cassart et al. (2008) and for Ferreira-Steel type distributions in Ley and Paindaveine (2009). Extending this effort to other flexible families and comparing the ensuing testing procedures in terms of their power seems a promising research topic.

Having arrived at this point, each reader will agree and disagree with me on the relevance of certain OQC's (and certainly have further OQC's in mind); some may be willing to take up a given challenge, others may (hopefully) benefit from certain ideas for their own research. The subdivision into underlined OQC's, which are sometimes provocatively formulated, follows a very precise goal: stimulate reflections on how we may/shall address flexible modelling in the future and open up discussions on this thrilling and ageless research direction!

Acknowledgements

References

Aigner, D. J., Lovell, C. A. K., and Schmidt, P. (1977). Formulation and estimation of stochastic frontier production function model. *J. Economet.*, 12:21–37.

Allard, D. and Naveau, P. (2007). A new spatial skew-normal random field model. *Comm. Statist. Theor. Meth.*, 39:1821–1834.

Arellano-Valle, R. B. and Azzalini, A. (2008). The centred parametrization for the multivariate skew-normal distribution. *J. Multivariate Anal.*, 99:1362–1382. Corrigendum: vol. 100 (2009), p. 816.

Arellano-Valle, R. B., Gómez, H. W., and Quintana, F. A. (2005). Statistical inference for a general class of asymmetric distributions. *J. Statist. Plann. Infer.*, 128:427–443.

Arnold, B. C. and Beaver, R. J. (2002). Skewed multivariate models related to hidden truncation and/or selective reporting (with discussion). *Test*, 11:7–54.

Arnold, B. C. and Groeneveld, R. A. (1995). Measuring skewness with respect to the mode. *Amer. Statist.*, 49:34–38.

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scand. J. Statist.*, 12:171–178.

Azzalini, A. (2005). The skew-normal distribution and related multivariate families (with discussion). *Scand. J. Statist.*, 32:159–188.

Azzalini, A. and Arellano-Valle, R. B. (2013). Maximum penalized likelihood estimation for skew-normal and skew-t distributions. *J. Statist. Plann. Infer.*, 143:419–433.

Azzalini, A. and Capitanio, A. (1999). Statistical applications of the multivariate skew-normal distributions. *J. Roy. Stat. Soc. Ser. B*, 61:579–602.

Azzalini, A. and Capitanio, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew-t distribution. *J. Roy. Stat. Soc. Ser. B*, 65(367-389).

Azzalini, A. and Capitanio, A. (2014). *The Skew-Normal and Related Families*. Cambridge: IMS Monographs. Cambridge University Press.

Azzalini, A. and Dalla Valle, A. (1996). The multivariate skew-normal distribution. *Biometrika*, 83:715–726.

Azzalini, A. and Genton, M. G. (2008). Robust likelihood methods based on the skew-t and related distributions. *Internat. Statist. Rev.*, 76:106–129.

Azzalini, A. and Regoli, G. (2012a). Some properties of skew-symmetric distributions. *Ann. Inst. Statist. Math.*, 64:857–879.

Azzalini, A. and Regoli, G. (2012b). The work of fernando de helguero on non-normality arising from selection. *Chil. J. Statist.*, 3:113–128.

Balanda, K. P. and MacGillivray, H. L. (1988). Kurtosis: a critical review. *Amer. Statist.*, 42:111–119.

Barndorff-Nielsen, O., Kent, J., and Sorensen, M. (1982). Normal variance-mean mixtures and *z* distributions. *Internat. Statist. Rev.*, 50:145–159.

Bauwens, L. and Laurent, S. (2005). A new class of multivariate skew densities, with application to generalized autoregressive conditional heteroscedasticity models. *J. Bus. Econom. Statist.*, 23:346–354.

Birnbaum, Z. W. (1950). Effect of linear truncation on a multinormal population. *Ann. Math. Statist.*, 21:272–279.

Box, G. E. P. and Cox, D. R. (1964). An analysis of transformations. *J. Roy. Stat. Soc. Ser. B*, 26:211–252.

Cassart, D., Hallin, M., and Paindaveine, D. (2008). Optimal detection of Fechner-asymmetry. *J. Statist. Plann. Infer.*, 138:2499–2525.

Charemza, W., Vela, C. D., and Makarova, S. (2013). Too many skew-normal distributions? The practitioners perspective. Discussion papers in economics 13/07, University of Leicester.

Cherubini, U., Luciano, E., and Vecchiato, W. (2004). *Copula Methods in Finance*. New York: Wiley.

Cover, T. M. and Thomas, J. A. (2006). *Elements of Information Theory*. Wiley-interscience.

Critchley, F. and Jones, M. C. (2008). Asymmetry and gradient asymmetry functions: density-based skewness and kurtosis. *Scand. J. Statist.*, 35:415–437.

De Forest, E. (1882). On an asymmetrical probability curve. *Analyst*, 9:135–143.

De Forest, E. (1883). On an asymmetrical probability curve. *Analyst*, 10:67–74.

de Helguero, F. (1909a). Sulla rappresentazione analitica delle curve abnormali. In *Atti del IV Congresso Internazionale dei Matematici (Roma, 6-11 Aprile 1908), volume III, sezione III-B*, pages 288–299. Roma: R. Accademia dei Lincei.

de Helguero, F. (1909b). Sulla rappresentazione analitica delle curve statistiche. *Giorn. Econ.*, 38:241–265.

de Vries, H. (1894). Ueber halbe Galton-Curven als Zeichen discontinuirlicher Variation. *Ber. Deutsch. Bot. Ges.*, 12:197–207.

Duerinckx, M., Ley, C., and Swan, Y. (2014). Maximum likelihood characterization of distributions. *Bernoulli*, 20:775–802.

Edgeworth, F. Y. (1886). The law of error and the elimination of chance. *Philos. Mag.*, 21:308–324.

Edgeworth, F. Y. (1898). On the representation of statistics by mathematical formulae. *J. R. Statist. Soc.*, 61:670–700.

Fechner, G. T. (1897). *Kollektivmasslehre*. Leipzig: Engelmann.

Fernández, C. and Steel, M. F. J. (1998). On Bayesian modeling of fat tails and skewness. *J. Amer. Statist. Assoc.*, 93:359–371.

Ferreira, J. T. A. S. and Steel, M. F. J. (2006). A constructive representation of univariate skewed distributions. *J. Amer. Statist. Assoc.*, 101:823–829.

Ferreira, J. T. A. S. and Steel, M. F. J. (2007a). Model comparison of coordinate-free multivariate skewed distributions with an application to stochastic frontiers. *J. Economet.*, 137:641–673.

Ferreira, J. T. A. S. and Steel, M. F. J. (2007b). A new class of skewed multivariate distributions with applications to regression analysis. *Statist. Sinica*, 17:505–529.

Field, C. and Genton, M. G. (2006). The multivariate *g*-and-*h* distribution. *Technometrics*, 48:104–111.

Fischer, M. and Klein, I. (2004). Kurtosis modelling by means of the *j*-transformation. *Allgemeines Statistisches Archiv*, 88:35–50.

Fujisawa, H. and Abe, T. (2014). A family of multivariate skew distributions with monotonicity of skewness. Manuscript.

Fujisawa, H. and Abe, T. (2015). A family of skew-unimodal distributions with mode-invariance through transformation of scale. *Statistical Methodology*.

Gauss, C. F. (1809). *Theoria motus corporum coelestium in sectionibus conicis solem ambientium*. Cambridge Library Collection. Cambridge University Press.

Genton, M. G. (2004). *Skew-elliptical Distributions and Their Applications: A Journey Beyond Normality*. Edited volume, Boca Raton, FL: Chapman and Hall/CRC.

Hald, A. (1998). *A History of Mathematical Statistics from 1750 to 1930*. Wiley, New York.

Hallin, M. and Ley, C. (2012). Skew-symmetric distributions and Fisher information - a tale of two densities. *Bernoulli*, 18:747–763.

Hallin, M. and Ley, C. (2014). Skew-symmetric distributions and Fisher information: the double sin of the skew-normal. *Bernoulli*, 20:1432–1453.

Hansen, B. E. (1994). Autoregressive conditional density estimation. *Internat. Econ. Rev.*, 35:705–730.

Haynes, M. A., MacGillivray, H. L., and Mergersen, K. L. (1997). Robustness of ranking and selection rules using generalized *g* and *k* distributions. *J. Statist. Plann. Infer.*, 65:45–66.

Heinz, G., Peterson, L. J., Johnson, R. W., and Kerk, G. J. (2003). Exploring relationships in body dimensions. *Journal of Statistics Education (online only)*, 11:www.amstat.org/publications/jse/v11n2/datasets.heinz.html.

Johnson, N. L. (1949). Systems of frequency curves generated by methods of translation. *Biometrika*, 36:149–176.

Johnson, N. L., Kotz, S., and Balakrishnan, N. (1994). *Continuous Univariate Distributions*. New York: John Wiley and Sons.

Jones, M. C. (2004). Families of distributions arising from distributions of order statistics. *Test*, 13:1–43.

Jones, M. C. (2010). Distributions generated by transformation of scale using an extended Schlömilch transformation. *Sankhya A*, 72:359–375.

Jones, M. C. (2014). Generating distributions by transformation of scale. *Statist. Sinica*, 24:749–772.

Jones, M. C. (2015a). On bivariate transformation of scale distributions. *Comm. Statist. Theor. Meth.*

Jones, M. C. (2015b). On families of distributions with shape parameters (with discussion). *Internat. Statist. Rev.*

Jones, M. C. and Anaya-Izquierdo, K. (2011). On parameter orthogonality in symmetric and skew models. *J. Statist. Plann. Infer.*, 141:758–770.

Jones, M. C. and Pewsey, A. (2009). Sinh-arcsinh distributions. *Biometrika*, 96:761–780.

Jones, M. C., Rosco, J. F., and Pewsey, A. (2011). Skewness-invariant measures of kurtosis. *Amer. Statist.*, 65:89–95.

Kapteyn, J. C. (1903). *Skew Frequency Curves in Biology and Statistics*. Groningen: Noordhoff.

Kotz, S. and Vicari, D. (2005). Survey of developments in the theory of continuous skewed distributions. *METRON*, 63:225–261.

Kurowicka, D. and Joe, H. (2010). *Dependence Modeling: Vine Copula Handbook*. World Scientific: Singapore.

Lee, C., Famoye, F., and Alzaatreh, A. (2013). Methods for generating families of univariate continuous distributions in the recent decades. *WIREs Comput. Statist.*, 5:219–238.

Ley, C. (2012). Skew distributions. In *Encyclopedia of Environmetrics Second Edition, A.-H. El-Shaarawi and W.*

Piegorsch (eds), pages 1944–1949. John Wiley & Sons Ltd, Chichester, UK.

Ley, C. and Paindaveine, D. (2009). Le Cam optimal tests for symmetry against Ferreira and Steel's general skewed distributions. *J. Nonparam. Statist.*, 21:943–967.

Ley, C. and Paindaveine, D. (2010a). Multivariate skewing mechanisms: a unified perspective based on the transformation approach. *Stat. Probab. Lett.*, 80:1685–1694.

Ley, C. and Paindaveine, D. (2010b). On Fisher information matrices and profile log-likelihood functions in generalized skew-elliptical models. *METRON*, 68, Special Issue on "Skew-symmetric and flexible distributions":235–250.

Ley, C. and Paindaveine, D. (2010c). On the singularity of multivariate skew-symmetric models. *J. Multivariate Anal.*, 101:1434–1444.

Ley, C. and Paindaveine, D. (2015). Discussion of "On families of distributions with shape parameters" by M. C. Jones. *Internat. Statist. Rev.* To appear.

Ley, C. and Verdebout, T. (2014). Skew-rotsymmetric distributions on unit spheres and related efficient inferential procedures. ECARES Working Paper 2014-46.

McLachlan, G. and Peel, D. (2000). *Finite Mixture Models*. Wiley Series in Probability and Statistics.

McWilliams, T. P. (1990). A distribution-free test for symmetry based on a runs statistic. *J. Amer. Statist. Assoc.*, 85:1130–1133.

Meeusen, W. and van den Broeck, J. (1977). Efficiency estimation from Cobb-Douglas production functions with composed error. *Internat. Econ. Rev.*, 18:435–444.

Mudholkar, G. S. and Hutson, A. D. (2000). The epsilon-skew-normal distribution for analyzing near-normal data. *J. Statist. Plann. Infer.*, 83:291–309.

Naveau, P., Genton, M. G., and Shen, X. (2005). A skewed Kalman filter. *J. Multivariate Anal.*, 94:382–400.

Nelsen, R. B. (2006). *An Introduction to Copulas, Second Edition*. Springer: New York.

O'Hagan, A. and Leonard, T. (1976). Bayes estimation subject to uncertainty about parameter constraints. *Biometrika*, 63:201–203.

Pearson, K. (1893). Asymmetrical frequency curves. *Nature*, 48:615–616.

Pearson, K. (1895a). Contributions to the mathematical theory of evolution. ii. Skew variation in homogeneous material. *Phil. Trans. R. Soc. Lond. A*, 186:343–414.

Pearson, K. (1895b). On skew probability curves. *Nature*, 52:317.

Pearson, K. (1901). Mathematical contributions to the theory of evolution. X. Supplement to a memoir on skew variation. *Phil. Trans. R. Soc. Lond. A*, 197:443–459.

Pearson, K. (1905). Das Fehlergesetz und seine Verallgemeinerungen durch Fechner und Pearson. A rejoinder. *Biometrika*, 4:169–212.

Pearson, K. (1916). Mathematical contributions to the theory of evolution, XIX: Second supplement to a memoir on skew variation. *Phil. Trans. R. Soc. Lond. A*, 216:429–457.

Pewsey, A. (2015). Invited discussion of "On families of distributions with shape parameters" by M. C. Jones. *Internat. Statist. Rev.*

Poincaré, H. (1896). *Calcul des Probabilités*. Carré-Naud, Paris.

Pretorius, S. J. (1930). Skew bivariate frequency surfaces, examined in the light of numerical illustrations. *Biometrika*, 22:109–223.

Randles, R. H., Fligner, M. A., Policello, G. E., and Wolfe, D. A. (1980). An asymptotically distribution-free test for symmetry versus asymmetry. *J. Amer. Statist. Assoc.*, 75:168–172.

Ranke, K. E. and Greiner, A. (1904). Das Fehlergesetz und seine Verallgemeinerungen durch Fechner und Pearson in ihrer Tragweite fur die Anthropologie. *Archiv für Anthropologie*, 2:295–332.

Rayner, G. D. and MacGillivray, H. L. (2002). Numerical maximum likelihood estimation for the *g*-and-*k* and generalized *g*-and-*h* distributions. *Stat. Comput.*, 12:57–75.

Razali, N. M. and Wah, Y. B. (2011). Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. *J. Statist. Model. Analytics*, 2:21–33.

Rieck, J. R. and Nedelman, J. R. (1991). A log-linear model for the Birnbaum-Saunders distribution. *Technometrics*, 33:51–60.

Rosco, J. F. (2012). *Aplicaciones de la Transformacion Sinh-arcsinh y Problemas Relacionados*. PhD thesis, Universidad de Extremadura.

Rubio, F. J. (2013). *Modelling of kurtosis and skewness : Bayesian inference and distribution theory*. PhD thesis, University of Warwick.

Rubio, F. J., Ogundimu, E. O., and Hutton, J. L. (2014). Robust modelling using two-piece sinh-arcsinh distributions.

arXiv:1307.6021.

Rubio, F. J. and Steel, M. F. J. (2014). Inference in two-piece location-scale models with Jeffreys priors (with discussion). *Bayesian Analysis*, 9:1–22.

Serfling, R. J. (2006). Multivariate symmetry and asymmetry. In *Encyclopedia of Statistical Sciences, Second Edition (S. Kotz, N. Balakrishnan, C.B. Read and B. Vidakovic, eds.), Vol. 8*, pages 5338–5345. Wiley.

Sklar, A. (1959). Fonctions de répartition à *n* dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Université de Paris*, 8:229–231.

Smith, R. L. and Naylor, J. C. (1987). A comparison of maximum likelihood and Bayesian estimators for the three-parameter Weibull distribution. *J. Roy. Stat. Soc. Ser. C*, 36:358–369.

Stamhuis, I. H. and Seneta, E. (2009). Pearson's statistics in the Netherlands and the astronomer Kapteyn. *Internat. Statist. Rev.*, 77:96–117.

Steel, M. F. J. and Rubio, F. J. (2015). Discussion of "On families of distributions with shape parameters" by M. C. Jones. *Internat. Statist. Rev.*

Stigler, S. M. (1978). Francis Ysidro Edgeworth, statistician (with discussion). *J. Roy. Stat. Soc. Ser. A*, 141:287–322.

Stigler, S. M. (1986). *The history of statistics: The measurement of uncertainty before 1900*. Cambridge and London: The Belknap Press of Harvard University Press.

Tukey, J. W. (1977). Modern techniques in data analysis. In *NSF-sponsored Regional Research Conference*. Southern Massachusetts University, North Darthmout, MA.

Umbach, D. (2007). The effect of the skewing distribution on skew-symmetric families. *Soochow Journal of Mathematics*, 33:657–668.

Wallis, K. F. (2014). The two-piece normal, binormal, or double Gaussian distribution: its origin and rediscoveries. *Statist. Sci.*, 29:106–112.

Wang, J., Boyer, J., and Genton, M. G. (2004). A skew-symmetric representation of multivariate distribution. *Statist. Sinica*, 14:1259–1270.