UNIVERSIDAD DE EXTREMADURA

TESIS DOCTORAL

Aplicaciones de la Transformación Sinh-arcsinh y Problemas Relacionados

Juan Francisco Rosco Nieves

Departamento de Matemáticas

Conformidad del Director

Fdo: Arthur Pewsey

2012

“We can only conclude from the investigations here considered that the normal curve possesses no special fitness for describing errors or deviations such as arise either in observing practice or in nature.”

Karl Pearson, 1900.

Acknowledgements

I would like to express my deep and sincere gratitude to my supervisor, Professor Arthur Pewsey. His wide knowledge and his logical and British way of thinking have been of great value for me. His understanding, encouraging and personal guidance have provided the basis for the present thesis.

I wish to express my warm and sincere thanks to Professors Harry Joe (University of British Columbia, Canada), Claudio Agostinelli (Ca' Foscari University, Italy) and Chris Jones (Open University, UK) who gave me the opportunity to carry out joint research with them. They made me feel at home, and the experience gained from working with them was most enlightening.

I owe loving thanks to Carmen. Without her encouragement and understanding it would have been impossible for me to complete this thesis. Special gratitude is due to my parents, Sebastián and Francisca, and my brother José Manuel for their loving support throughout these years of ups and downs. Their influence is surely what has moulded me into who I now am, and it is their model that encourages me to strive to become a better person.

The financial support of the European Social Fund through the Junta de Extremadura is gratefully acknowledged.

Contents

Dedicated to my grandparents whose austere lives remind us that we do not need to own but to share

1 Summary

This thesis is based around the four papers in the appendix that summarise research work conducted between 2008 and 2012 under the supervision of Professor Arthur Pewsey. The first two have already been published in the statistical journals *Test* and *The American Statistician*, respectively. The third paper is currently under revision, and an electronic version of the fourth has been published on the *Statistical Papers* website.

In Section 2 we provide an introduction to the main ideas underpinning the four papers. The motivation and main objectives of the work conducted are described in Section 3, whilst Section 4 provides a joint discussion of the results obtained. Finally, conclusions are presented, and prospects for future research discussed, in Section 5.

The first of the four papers in the appendix, entitled 'Skew-t distributions via the sinh-arcsinh transformation', presents results for a new skew-symmetric family of distributions with Student's t distribution as its symmetric member. The family was obtained by applying a slightly modified version of the sinh-arcsinh transformation of Jones & Pewsey (2009) to the t distribution. The properties of the new skew-t family are developed, and particular attention given to quantile-based measures of kurtosis which are skewness-invariant. Maximum likelihood inference is studied and illustrated in the analysis of a real data set. Multivariate extensions of the proposed skew-t family are also considered. A comparison with other skew-t distributions that have been proposed within the literature is made throughout the article.

The second paper, entitled 'Skewness-invariant measures of kurtosis', further develops the topic of skewness-invariant measures of kurtosis. Two classes of quantile-based measures of kurtosis are identified as being skewness-invariant for certain

families of skew-symmetric distributions obtained via transformation of a symmetric random variable. For this type of distributions we state a condition, in terms of the transformation used, that ensures the invariance of the measures of kurtosis. A transformation satisfying this condition is given as an example, namely, the sinh-arcsinh transformation. Another class of measures of kurtosis, based on densities, is briefly studied for so-called transformations of scale distributions.

The third paper presented in the appendix is entitled 'On Blest's measure of kurtosis adjusted for skewness'. There the topic of moment-based kurtosis measures that are invariant to skewness is investigated. In particular, the measure of kurtosis adjusted for skewness introduced in Blest (2003) is studied and an alternative to it proposed. The performance of both measures is analysed using skew-symmetric families of distributions, and lower bounds for them derived. Results are also presented from an extensive simulation study designed to identify the best performing sample versions of the measures obtained by plugging-in different moment estimators available in popular statistical packages.

In the fourth paper, entitled 'Measures of tail asymmetry for bivariate copulas', we start by identifying a set of desirable properties that a tail asymmetry measure should satisfy. We then propose three families of tail asymmetry measures for bivariate copulas. The first two families, one moment-based and the other quantile-based, are obtained using concepts associated with asymmetry measures for univariate distributions. The third family is derived using an L_∞ distance approach. Bounds for all three measures are obtained together with the copulas that attain them. Two examples involving real data sets illustrate the levels of asymmetry that might be expected in practice.

Resumen

Esta tesis está basada en los cuatro artículos que se adjuntan en el apéndice y que sintetizan el trabajo de investigación llevado a cabo entre los años 2008 y 2012 bajo la tutela del profesor Arthur Pewsey. De los cuatro artículos, los dos primeros han sido publicados en las revistas estadísticas *Test* y *The American Statistician*, respectivamente. El tercero se encuentra actualmente bajo revisión y una versión digital del cuarto artículo ha sido publicada en la página web de la revista estadística *Statistical Papers*.

En la sección 2 se presenta una introducción a las ideas principales que se desarrollan en los cuatro artículos. Las motivaciones y los objetivos principales del trabajo desarrollado se describen en la sección 3 mientras que una discusión conjunta de los resultados obtenidos se presenta en la sección 4. Para finalizar, se presentan las conclusiones y se discuten posibles líneas de investigación a seguir en el futuro en la sección 5.

El primero de los cuatro artículos anexados, titulado 'Skew-t distributions via the sinh-arcsinh transformation', presenta resultados para una nueva familia de distribuciones con miembros tanto simétricos como asimétricos y cuya subclase simétrica está formada por las distribuciones t de Student. En adelante, nos referiremos a una de tales familias como familia skew-t. Si no se especifica la subclase simétrica, nos referiremos a ella como familia skew-symmetric. La nueva familia skew-t se obtuvo aplicando una ligera modificación de la transformación sinh-arcsinh de Jones & Pewsey (2009) a la distribución t. Las propiedades de esta nueva familia skew-t se desarrollan en el artículo, prestando especial atención a medidas de curtosis basadas en cuantiles que son invariantes a la asimetría. También se estudia la estimación por máxima verosimilitud y se ilustra su uso mediante el análisis de un conjunto de datos reales. Para completar el estudio, una extensión multivariante de la nueva familia skew-t es considerada. A lo largo

de todo el artículo se realiza una comparación con otras distribuciones skew-t que pueden encontrarse en la literatura.

En el segundo artículo, de título 'Skewness-invariant measures of kurtosis', se desarrolla más a fondo el tema de las medidas de curtosis invariantes a la asimetría. Se identifican dos clases de medidas de curtosis, basadas en cuantiles, que son invariantes a la asimetría para ciertas familias skew-symmetric obtenidas mediante la transformación de una variable aleatoria simétrica. Para este tipo de distribuciones se establece una condición suficiente, en términos de la transformación usada, cuya verificación asegura la invariancia a la asimetría de las medidas de curtosis. Como ejemplo de una tal transformación se presenta la transformación sinh-arcsinh. Finalmente otra clase de medidas de curtosis, basada en densidades, se estudia brevemente para transformaciones de distribuciones de escala.

El tercer artículo presente en el apéndice tiene por título 'On Blest's measure of kurtosis adjusted for skewness'. En él, el tema de medidas de curtosis que son invariantes a la asimetría se estudia para medidas basadas en momentos. Concretamente, se estudia la medida de curtosis ajustada para la asimetría introducida en Blest (2003) y se propone una medida alternativa. Se analiza el comportamiento de ambas medidas usando familias skew-symmetric y se derivan cotas inferiores para ellas. Además, se presentan los resultados obtenidos en un estudio de simulación diseñado para identificar la mejor versión muestral de las medidas. Estas versiones muestrales fueron obtenidas reemplazando los momentos poblacionales por sus distintas versiones muestrales en la definición de las medidas. Las versiones muestrales de los momentos que se utilizaron se corresponden con las implementadas en los paquetes estadísticos más populares.

En el cuarto artículo, titulado 'Measures of tail asymmetry for bivariate copulas', comenzamos estableciendo un conjunto de propiedades deseables que una medida de asimetría debiera satisfacer. A continuación proponemos tres familias de medidas de asimetría para cópulas bivariantes. Las dos primeras familias, una basada en momentos y otra en cuantiles, se obtienen usando conceptos asociados con medidas de asimetría para distribuciones de probabilidad univariantes. La tercera medida se deriva usando la distancia L_∞. Seguidamente se obtienen cotas para las tres medidas y se identifican cópulas que alcanzan dichas cotas. El grado de asimetría que se puede esperar en la práctica se ilustra con dos ejemplos usando datos reales.

2 Introduction

In this introduction to the thesis we consider the background to the main ideas underpinning the papers contained in the appendix. In Section 2.1 we consider the normal distribution, asymmetric data and approaches to generating distributions capable of modelling their main features. As we will see, one popular means of generating asymmetric distributions is via the transformation of a symmetric random variable. In Section 2.2 we focus on a particular form of transformation, the so-called sinh-arcsinh transformation, which plays a crucial role throughout the thesis. In the paper entitled 'Skew t distributions via the sinh-arcsinh transformation' we study a new family of distributions arising from the t distribution when a special case of the sinh-arcsinh transformation is applied to it, and hence in Section 2.3 we provide some background to Student's t distribution and various asymmetric extensions of it that have recently been proposed in the literature.

As background to the articles entitled 'Skewness-invariant measures of kurtosis' and 'On Blest's measure of kurtosis adjusted for skewness', in Section 2.4 we provide an overview of the classical coefficient of kurtosis and how its various interpretations have spawned numerous articles addressing not only the interpretation of the classical coefficient of kurtosis but also the concept of kurtosis in the presence of asymmetry as well as alternatives to the classical coefficient.

In Section 2.5 we provide a brief introduction to copula theory, a research topic that has stimulated considerable activity in recent years. First we review the ideas underpinning copulas and various particular classes of bivariate copulas. One such class, that of vine copulas, has proved particularly important in the modelling of financial data. In Section 2.5.7, we consider the few asymmetry measures that have been proposed in the literature for use with bivariate copulas. Their paucity was the catalyst for the fourth paper in the appendix entitled 'Measures of tail asymmetry for bivariate copulas'.

An overview of the remainder of the thesis is provided in Section 2.6.

2.1 Normality, asymmetric data and approaches to modelling them

In classical Statistics it is usually assumed that the distribution from which the data were drawn is symmetric or, more specifically, normal. As Lippmann stated in a remark to Poincaré: "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" (see Whittaker & Robinson, 1965, p. 179).

In the 18th Century, Abraham de Moivre was often called upon to make the long and tedious computations involved in calculating probabilities such as that of obtaining 20 tails when tossing a fair coin 100 times. At the time it was well-known that the binomial distribution could be used to resolve this sort of problem. During his investigations, de Moivre noted that as the number of events increased, the shape of the binomial probability mass function approached a smooth curve. He reasoned that obtaining an expression for such a curve would lead to elegant and easy-to-calculate solutions. He derived the equation of the curve and discovered not only the normal distribution but also that the binomial distribution could be well approximated by a normal distribution when the number of Bernoulli trials is large. His results were published in de Moivre (1738).

An historically important application of the normal distribution was in the analysis of measurement errors made in astronomical observations; the errors caused by imperfections in the instruments used and, of course, the observers. Various distributions had been hypothesised for such errors, but it was not until the early 19th Century that it was discovered that these errors tended to follow a normal distribution. Independently, in 1809, Gauss and Robert Adrain derived the formula for the normal distribution. Laplace also made significant contributions, being the first to obtain the normalising constant of the normal distribution. It was also Laplace who, in 1810, proved the fundamental central limit theorem, thus establishing the statistical importance of the normal distribution. Thereafter, its mathematical properties and tractability made it an increasingly popular model.

In 1835, Adolphe Quetelet became the first person to apply it to human characteristics such as height, weight and strength.

However, as time went by and applications of the normal distribution proliferated, it was increasingly found that real data are seldom symmetric, let alone normal. A seminal reference in this regard is Pearson (1900). In response, various approaches were proposed to deal with data that are skew.

Historically, perhaps the most popular way of dealing with asymmetric data has been to transform them in an attempt to produce data for which the classical assumptions are more reasonable. As the best choice of transformation was not necessarily obvious for a particular data set, various standard transformations were established and applied in practice. The logarithmic transformation, for instance, was generally found to be appropriate for data associated with growth. When an even stronger transformation was required, the reciprocal transformation was often employed. For count data the square-root transformation was often found to be suitable. Indeed, it was discovered that data from the Poisson distribution could be transformed to normality by applying such a transformation. The arcsine transformation was found to be especially useful when the data were percentages or proportions. The generally *ad hoc* use of transformations to approximate normality led Tukey (1957), and more famously Box & Cox (1964), to propose the power family of transformations, generally referred to nowadays as the Box-Cox family, as well as likelihood-based methods for identifying the optimal transformation from within it.

Despite the transformation approach being a popular one, there are two major problems associated with its use. First, it is possible that, for a given data set, no adequate power transformation can be identified. Secondly, in situations in which an adequate power transformation can be found, inference will generally be of interest on the scale upon which the data were originally observed, not on the transformed scale. This involves the use of back-transformation of the results for the transformed data, which generally introduces bias.

The alternative approach to analysing asymmetric data, which is the one explored in this thesis, is to model the data on the scale on which they were observed using flexible models capable of describing their main features. In particular, we will generally concentrate on models for unimodal data with parameters controlling the location, scale, skewness and kurtosis of the distributions contained within them.

Non-normal probability distributions were, of course, known long before the introduction of the normal distribution. As mentioned above, de Moivre discovered the normal curve whilst trying to approximate the binomial distribution. In 1763, Bayes identified the beta distribution as the posterior density for the probability of success in a Bernoulli trial. Thirty-five years prior to the publication of the work on the normal distribution by Gauss, Laplace (1774) studied the problem of aggregating several observations and discovered the double exponential distribution. When developing his theory of probability, Poisson (1837) introduced what we know today as the Poisson distribution. De Forest (1882) derived the gamma distribution when approximating binomial coefficients using differential equations.

The development of most of the fundamental ideas underpinning the generation of asymmetric models took place in a highly fruitful period around the end of the 19th Century and the beginning of the 20th. Edgeworth (1886) is generally cited as the first reference addressing the problem of fitting asymmetric distributions to asymmetric frequency data. Shortly after its publication, the modelling of non-normal frequency distributions attracted the attention of Karl Pearson. His interest in the topic was triggered by Walter Weldon, a zoologist who, together with his wife, collected data on 23 characteristics of 1000 female crabs whilst holidaying in Malta and the Bay of Naples. Weldon found that one of the characteristics, the frontal breadth of the carapace, did not follow a normal distribution, and asked Pearson for assistance. A potential solution to the modelling of such asymmetric data was provided in Pearson (1893), where finite mixtures of normal densities were advocated. The research conducted by Edgeworth and Pearson on asymmetric distributions led to considerable competition between them. Details of the their correspondence on the theme can be found in Stigler (1978), a biographical paper about Edgeworth.

A further important contribution of Karl Pearson to the modelling of skew data was Pearson (1895) in which he proposed various types of distributions obtained as solutions to a particular differential equation. In Pearson (1901, 1916) he extended the range of solutions so as to obtain what we know today as the Pearson family of distributions. In addition to the normal distribution, the Pearson family includes the: gamma, inverse-gamma, beta, inverted beta and Student's t distributions, as well as a generalisation of the beta distribution and a skew-symmetric extension of Student's t distribution.

Also towards the end of the 19th Century, Fechner (1897) had proposed an approach to constructing a skew distribution based on combining two differentially scaled halves of normal distributions. The result is referred to nowadays as the two-piece normal distribution. The two-piece construction was revisited by Gibbons & Mylroie (1973), and more recently by Fernández & Steel (1998) and Mudholkar & Hutson (2000).

The use of transformation (or, originally "translation") to obtain new distributions from existing ones makes its first appearance in Edgeworth (1898). There, Edgeworth considered transformations which can be represented by polynomials. Subsequently, Kapteyn & van Uven (1916), Wicksell (1917, 1923) and Rietz (1922) extended the approach using different types of transformation, although the resulting distributions displayed only a limited variety of shapes. In the mid 20th Century the approach gained renewed interest with the publication of Johnson (1949) in which the log-normal, a slight modification of it, and the arcsinh transformations were applied to a normal random variable to obtain random variables from the S_L (log-normal), S_B (bounded) and S_U (unbounded) families, respectively. In Tadikamalla & Johnson (1982), the same three transformations were applied to a logistic random variable, obtaining the corresponding L_L, L_B and L_U families. Tukey (1977) applied the transformation approach to a normal random variable to obtain the flexible g-and-h distribution, defined through its quantile function. Both g and h are parameters, the former controlling asymmetry and the latter tailweight. Towards the end of the last century, Rieck & Nedelman (1991) obtained sinh-normal distributions by applying the sinh, rather than the arcsinh, transformation to a normal random variable. Recently, Jones & Pewsey (2009) combined the use of the sinh and arcsinh functions in their sinh-arcsinh transformation, and studied the sinh-arcsinhed normal (SAS-normal) distribution obtained by applying it to a normal random variable. Heavy-tailed symmetric members of the SAS-normal family behave like Johnson S_U distributions, whilst their lighter-tailed counterparts behave like sinh-normal distributions. The sinh-arcsinh transformation plays a crucial role throughout the remainder of the thesis. It is considered in detail in the next subsection, and is applied to a Student's t random variable to obtain the sinh-arcsinhed t (SAS-t) distribution considered in the first of the papers in the appendix.

A contribution from the beginning of the 20th Century, largely ignored until relatively recently, was that of de Helguero (1908). In his paper, de Helguero considered modelling asymmetric data not only using a mixture of two normal distributions but also via a distribution obtained by applying a selection mechanism to a normal population. The model he derived is a form of skew-normal distribution intimately related to what, following the publication of Azzalini (1985), is generally referred to Azzalini's skew-normal distribution. It should however be noted, as Azzalini (2005) points out, that this particular skew-normal distribution had previously appeared in various guises; for example, in Birnbaum (1950), Nelson (1964), Roberts (1966), Aigner et al. (1977) and Anděl et al. (1984). More generally, Lemma 1 of Azzalini (1985) considers the following construction. Let f denote a density which is symmetric about 0, and G an absolutely continuous distribution function whose derivative is symmetric about 0. Then,

$$2G(\alpha x)f(x), \qquad x \in (-\infty, +\infty),$$

where $-\infty < \alpha < \infty$ is a shape parameter, is a skew-symmetric density function. In particular, Azzalini (1985) studied the skew-normal family of distributions obtained when G and f are the distribution and density functions of the standard normal distribution, respectively. This surprisingly simple perturbation construction has stimulated a vast literature. At the time of writing it has received no less than 400 citations in the Web of Knowledge and 840 in Google Scholar. Many of the most significant contributions to this line of research are referenced in Genton (2004) and Azzalini (2005).

So, in summary, we have highlighted five major means of obtaining models for skew data: (i) finite mixtures; (ii) the solution to a differential equation; (iii) transformation; (iv) piecing together two different halves of symmetric densities; (v) selection or Azzalini-type perturbation.

2.2 The sinh-arcsinh transformation and its properties

The sinh-arcsinh transformation, $S_{\varepsilon,\delta}(x) = \sinh(\delta \sinh^{-1}(x) - \varepsilon)$, where $-\infty < \varepsilon < +\infty$ and $0 < \delta < +\infty$, was introduced in Jones & Pewsey (2009) as a means of generating new skew-symmetric families of distributions from symmetric ones. Given a random variable, X, that is symmetric about 0, a new random variable, $Y_{\varepsilon,\delta}$, is obtained through

$$X = S_{\varepsilon,\delta}(Y_{\varepsilon,\delta}) = \sinh(\delta \sinh^{-1}(Y_{\varepsilon,\delta}) - \varepsilon). \tag{2.1}$$

The following three equivalent representations of $S_{\varepsilon,\delta}$ prove useful:

$$\begin{aligned} S_{\varepsilon,\delta}(x) &= \frac{1}{2}\left(e^{-\varepsilon}\exp(\delta\sinh^{-1}(x)) - e^{\varepsilon}\exp(-\delta\sinh^{-1}(x))\right) \\ &= \frac{1}{2}\left(e^{-\varepsilon}(\sqrt{x^2+1}+x)^{\delta} - e^{\varepsilon}(\sqrt{x^2+1}+x)^{-\delta}\right) && (2.2) \\ &= \frac{1}{2}\left(e^{-\varepsilon}(\sqrt{x^2+1}+x)^{\delta} - e^{\varepsilon}(\sqrt{x^2+1}-x)^{\delta}\right). && (2.3) \end{aligned}$$

In particular, Equations (2.2) and (2.3) are very useful for computational purposes. Inverting Equation (2.1) leads to

$$Y_{\varepsilon,\delta} = S^{-1}_{\varepsilon,\delta}(X) = \sinh(\delta^{-1}(\sinh^{-1}(X) + \varepsilon)) = S_{-\varepsilon/\delta,1/\delta}(X).$$

Thus random variate generation is straightforward provided that X can be simulated.

Given F_X, f_X and Q_X, the distribution, density and quantile functions, respectively, of X, it follows that the corresponding functions of $Y_{\varepsilon,\delta}$ are given by

$$F_{\varepsilon,\delta}(y) = F_X(S_{\varepsilon,\delta}(y)), \tag{2.4}$$

$$f_{\varepsilon,\delta}(y) = \frac{\delta C_{\varepsilon,\delta}(y)}{\sqrt{1+y^2}} f_X(S_{\varepsilon,\delta}(y)), \tag{2.5}$$

and

$$Q_{\varepsilon,\delta}(u) = S_{-\varepsilon/\delta,1/\delta}(Q_X(u)), \quad 0 < u < 1, \tag{2.6}$$

where $C_{\varepsilon,\delta}(x) = \cosh(\delta\sinh^{-1}(x) - \varepsilon) = \sqrt{1 + S^2_{\varepsilon,\delta}(x)}$. It is easily shown that $f_{-\varepsilon,\delta}(x) = f_{\varepsilon,\delta}(-x)$. Also, as X is symmetric about 0, the median of $Y_{\varepsilon,\delta}$ is given by $\sinh(\varepsilon/\delta)$.

A particularly appealing property of the sinh-arcsinh transformation is that its parameters have clear interpretations. For fixed δ, ε acts as a skewness parameter in the sense of the skewness ordering of van Zwet (1964). That ordering is defined as follows. Given the distribution functions, F_1 and F_2, of two random variables, we say that F_2 is more positively skew than F_1 if $F_2^{-1}(F_1)$ is convex in its domain. For members of a sinh-arcsinh family, $F_{\varepsilon_2,\delta}$ is more positively skew than $F_{\varepsilon_1,\delta}$ whenever $\varepsilon_2 > \varepsilon_1$. In a similar fashion, when $\varepsilon = 0$, δ acts as a kurtosis parameter in the sense of van Zwet's (1964) ordering.

2.3 Student's t distribution and asymmetric extensions

The t distribution is generally considered to have been introduced in Student (1908), despite it having been derived as a posterior distribution by Helmert (1875) and Lüroth (1876) as well as being a special case of the Pearson type IV distribution introduced in Pearson (1895). 'Student' was, of course, the pseudonym of William Sealy Gosset, Gosset being a "student" of Pearson. In his highly influential paper, Gosset obtained the distribution of the statistic

$$T^* = \frac{\sqrt{n}(\bar{X} - \mu)}{\left(\sum_{i=1}^{n}(X_i - \bar{X})^2\right)^{1/2}} = \frac{T}{\sqrt{n-1}},$$

where $X_1, \ldots, X_n$ is a sequence of independent and identically distributed random variables, with sample mean $\bar{X}$, sampled from a normal distribution with mean μ, and T denotes the usual "t-statistic"

$$T = \frac{\sqrt{n}(\bar{X} - \mu)}{\left[(n-1)^{-1}\sum_{i=1}^{n}(X_i - \bar{X})^2\right]^{1/2}},$$

which follows what we refer to nowadays as Student's t distribution with $n-1$ degrees of freedom. The contemporary use of T rather than T^* and the terminology used for its distribution can be traced to Fisher (1925). The vast literature related to the t distribution is summarised in Johnson et al. (1994a, chap. 8).

The t distribution has become a popular model for financial data. Although *a priori* there is no reason why financial data should behave in any particular fashion, empirical studies have identified common features amongst them. These common features are known as stylised facts. Two of them are that the underlying distribution of asset returns generally (i) has heavier than normal tails and (ii) tends to be negatively skew. Due to the asymmetry inherent in their distribution, there has been considerable interest in extending the t distribution to a family of distributions with symmetric as well as asymmetric members. The Pearson type IV distribution (see Johnson et al., 1994a, chap. 12) and the noncentral t distribution (see Johnson et al., 1994b, chap. 31) are classical four-parameter extensions which include the t distribution and asymmetric versions of it. Amongst more recent proposals, here we highlight the: two-piece t distribution (Hansen, 1994; Fernández & Steel, 1998); skew-t distribution based on Azzalini-type perturbation (Branco & Dey, 2001; Azzalini & Capitanio, 2003; Genton, 2004; Ma & Genton, 2004); skew-t of Jones & Faddy (2003), one construction of which involves the transformation of a beta random variable.

The two-piece t distribution is a skewed version of the t distribution obtained by joining together two differently scaled halves of a Student's t distribution. More precisely, let t_ν denote the density function of a t distribution with ν degrees of freedom and $\gamma \in (0, +\infty)$ a constant, then the density function of the two-piece t distribution can be expressed as

$$f_{TP}(x) = \frac{2}{\gamma + (1/\gamma)} \left[t_\nu(x/\gamma) I(x \geq 0) + t_\nu(\gamma x) I(x < 0) \right].$$

The parameter γ controls the skewness. Values of $\gamma < 1$ ($\gamma > 1$) lead to negatively (positively) skewed distributions. Also, the density with a parameter value of $1/\gamma$ is the mirror image about 0 of the density with parameter value γ. Note that in the paper entitled 'Skew-t distributions via the sinh-arcsinh transformation', a different parameterisation of this density is referred to.

The Azzalini-type skew-t distribution is derived using a generalised version of the approach introduced in Azzalini (1985). Thus its density is a perturbation of the Student t density. More specifically, denoting the distribution function of the t distribution with ν degrees of freedom by T_ν, its density is given by

$$f_A(x) = 2 t_\nu(x) T_{\nu+1}\left(\alpha x \sqrt{\frac{\nu + 1}{\nu + x^2}} \right),$$

where, as in Section 2.1, $\alpha \in \mathbb{R}$ is the shape parameter. The density is symmetric if $\alpha = 0$; otherwise it is positively (negatively) skewed if $\alpha > 0$ ($\alpha < 0$).

The skew-t distribution of Jones & Faddy (2003) can be obtained in at least three different ways; one of which involves transforming a beta random variable. Let X

denote a beta random variable on $(0,1)$ with parameters $a > 0$ and $b > 0$, then the random variable

$$Y = \frac{\sqrt{(a+b)}(2X-1)}{2\sqrt{X(1-X)}}$$

is distributed according to Jones & Faddy's skew-t distribution. Alternatively, a random variable from the distribution can be obtained by transforming two independent χ^2 random variables with $2a$ and $2b$ degrees of freedom, respectively. A third approach is given in Jones (2001) and involves factorising Student's t distribution in two parts and raising them to different powers. All three constructions lead to the density

$$f_{a,b}(x) = C_{a,b}^{-1}\left(1+\frac{x}{\sqrt{a+b+x^2}}\right)^{a+1/2}\left(1-\frac{x}{\sqrt{a+b+x^2}}\right)^{b+1/2},$$

where $C_{a,b} = 2^{a+b-1}B(a,b)(a+b)^{1/2}$ and $B(\cdot,\cdot)$ denotes the beta function. The parameters a and b control the asymmetry. The density is symmetric if $a = b$, whereas if $a > b$ ($a < b$) it is skew to the right (left).

2.4 Measures of kurtosis

It has been over one hundred years since Thiele (1889) introduced, and Pearson (1905) popularised, the fourth standardised moment about the mean,

$$\alpha_4 = \frac{\mathrm{E}\left[(X-\mu)^4\right]}{\sigma^4},$$

where X denotes a random variable with finite fourth moment, mean μ and variance σ^2, as a measure of "kurtosis". Nevertheless, it is still unclear as to what exactly it measures and what its relationship with the shape of a distribution is.

Because α_4 involves fourth-order moments, and not all probability distributions have finite fourth moments, it need not necessarily exist. In order to circumvent this drawback, numerous alternative measures representing what their advocates understood "kurtosis" to be have been proposed in the literature. Invariably such measures are invariant to location-scale transformations. Many of the alternative measures are quantile-based, including those involving L-moments. As some authors found a single number summary of kurtosis too restrictive, functional measures have also been proposed. Most of the alternative measures are designed to reflect some interpretation of kurtosis in the absence of asymmetry because precisely what kurtosis should represent in the presence of asymmetry was generally considered to be even more puzzling. Nevertheless, various measures have also been proposed for use with skew distributions, including those of Moors (1988), Balanda & MacGillivray (1990), Hosking (1990), Blest (2003) and Critchley & Jones (2008).

Moors (1986) asserts that kurtosis can be interpreted as a measure of dispersion of that part of a distribution within one standard deviation of the mean. Using Q to denote the quantile function, in Moors (1988) he defines the kurtosis measure

$$M = \frac{Q(7/8) - Q(5/8) + Q(3/8) - Q(1/8)}{Q(3/4) - Q(1/4)}.$$

This measure exists and is unique so long as the six values of the quantile function involved in its calculation exist and are unique. M is a more robust measure of kurtosis than α_4 since it is less sensitive to extreme tails. In the paper entitled 'Skewness-invariant measures of kurtosis' we prove that, for distributions generated using certain kinds of transformation of a symmetric random variable, such as the sinh-arcsinh transformation for instance, Moors's measure does not depend on the asymmetry parameter; i.e. it is skewness-invariant.

A functional, quantile-based, summary of kurtosis was proposed by Balanda & MacGillivray (1990). They consider that, like location, scale and skewness, kurtosis is essentially a comparative concept and concentrate on developing orderings in terms of what they refer to as the "spread function", defined for a distribution with continuous density as

$$Q\left(1/2 + u\right) - Q\left(1/2 - u\right), \quad 0 < u < 1/2.$$

The same authors venture to provide an interpretation of the concept of kurtosis, defining it vaguely as "the location and scale-free movement of probability mass from the shoulders of a distribution into its centre and tails" (Balanda & MacGillivray, 1988).

Hosking (1990) does not introduce new ideas about the concept of kurtosis but defines measures of location, scale, skewness, kurtosis, etc. in terms of L-moments rather than classical moments. L-moments are expectations of certain linear combinations of order statistics which exist if, and only if, the mean of the distribution exists. This is clearly a weaker restriction than that underpinning α_4. L-kurtosis is defined as

$$\tau_4 = \frac{\int_0^1 P_3^*(u)Q(u)\mathrm{d}u}{\int_0^1 P_1^*(u)Q(u)\mathrm{d}u},$$

where $P_1^*(x) = 2x - 1$ and $P_3^*(x) = 20x^3 - 30x^2 + 12x - 1$ are the first and third shifted Legendre polynomials, respectively. Hosking notes that, like the classical measure, the new measure does not have a unique interpretation and is best thought of as a measure similar to the classical one but giving less weight to the extreme tails of the distribution. For distributions generated using the sinh-arcsinh transformation this measure is also skewness-invariant.

Ignoring the limitations of classical moments, and thus flying in the face of many of the developments described above, Blest (2003) proposes a new moment-based measure of kurtosis designed with the aim of removing the influence of skewness on α_4. His new measure is obtained by replacing the mean in the definition of α_4 by an alternative measure of central location; denoted by ξ and referred to as the "meson", mesos meaning "middle" in Greek. The meson is defined as that point about which the fourth moment of a distribution is minimum. Equivalently, it is that point about which the third moment is zero. Thus, under the same conditions as those for α_4, Blest's coefficient of kurtosis is defined as

$$\alpha_4^* = \frac{\mathrm{E}\left[(X-\xi)^4\right]}{\sigma^4}.$$

In fact, the effect of skewness on this kurtosis measure is not completely eliminated. In the paper entitled 'On Blest's measure of kurtosis adjusted for skewness', we provide examples which illustrate this fact. We also study the performance of a modified version of Blest's coefficient in which its denominator is replaced by $\mathrm{E}[(X-\xi)^2]^2$. Whilst still not skewness-invariant, this modified version is generally less affected by asymmetry than Blest's measure.

Like Balanda & MacGillivray (1990), Critchley & Jones (2008) perceive kurtosis as something that should be summarised functionally. However, their approach to describing it is very different. They first define functional measures of skewness for continuous univariate unimodal distributions, which they refer to as asymmetry functions. Then, as a means of describing kurtosis, they introduce left and right gradient asymmetry functions which are asymmetry functions of simple functions of the left and right parts of the derivative of the density. They also provide scalar measures of kurtosis obtained by integrating the gradient asymmetry functions.

2.5 Copulas

The first appearance in the statistical literature of a copula, although then not actually referred to as such, is often traced to Fréchet (1951). He studied the following problem. Given two univariate distribution functions, F_1 and F_2, can anything be said about the class of bivariate distributions functions with marginal distribution functions F_1 and F_2? It is obvious that the class is non-empty since, if the random variables are independent, the distribution function $F_1(x)F_2(y)$ is in the class. Of the numerous publications on this problem, the most profound results were presented in Sklar (1959). Sklar was also the person who formalised the concept of, and introduced the name for, copulas. Moreover, in Sklar (1973), he sketched the proof of the important theorem, to be considered in Section 2.5.3, which bears his name. During this period, most of the fundamental breakthroughs in copula theory were obtained in the course of the development of probabilistic metric spaces and the main source of basic information on them was Schweizer & Sklar (1983). Further important developments within the field were stimulated by a highly successful series of six major conferences which took place between 1990 and 2007. Also, in the late nineties, two seminal books were published and became the standard references in the field: Joe (1997) and Nelsen (1999). The latter was subsequently enhanced with new results and republished as Nelsen (2006).

Nevertheless, it was the discovery of copulas by researchers in applied fields that led to the increasing interest in them and the rapid development of copula theory. The research activity associated with the application of copulas in finance was particularly frenetic (see, for instance, Cherubini et al., 2004). However, copulas were also seen as useful tools in fields, like hydrology (see Salvadori et al., 2007), in which there was a need for more flexible multivariate models.

In the remainder of this section we introduce the notation and terminology that we will employ henceforth before providing the formal definition of a d-dimensional copula and proceeding to a consideration of bivariate copulas. Various fundamental results are also stated and three different classes of copulas introduced before the topic of measures of asymmetry for bivariate copulas is finally addressed.

2.5.1 Notation and terminology

We denote the extended real line by $\overline{\mathbb{R}}$ and given d, a positive integer, let $\overline{\mathbb{R}}^d$ denote the extended d-dimensional space $\overline{\mathbb{R}} \times \ldots \times \overline{\mathbb{R}}$. Vector notation will be used, as in $\mathbf{a} = (a_1, \ldots, a_d)$, and we will write $\mathbf{a} \leq \mathbf{b}$ if $a_k \leq b_k$ for all $k \in \{1, \ldots, d\}$. Given $\mathbf{a} \leq \mathbf{b}$, we will use $[\mathbf{a}, \mathbf{b}]$ to denote the d-box, or hyperrectangle or orthotope, $[a_1, b_1] \times \ldots \times [a_d, b_d]$. The vertices of a d-box are the points $\mathbf{c} = (c_1, \ldots, c_d)$ where c_k is equal to either a_k or b_k. The unit d-cube is the Cartesian product of d unit closed intervals, that is, $\mathbf{I}^d = \mathbf{I} \times \ldots \times \mathbf{I}$. A 2-box is a rectangle $[a_1, b_1] \times [a_2, b_2]$ and the unit 2-cube is the unit square $\mathbf{I}^2$ in $\overline{\mathbb{R}}^2$. A d-place real function H is a function of the form

$$\mathrm{Dom}H \subseteq \overline{\mathbb{R}}^d \quad \longrightarrow \quad \mathrm{Ran}H \subseteq \mathbb{R},$$

where Dom and Ran denote the domain and range, respectively.

Let $S_1, \ldots S_d$ denote non-empty subsets of $\overline{\mathbb{R}}$, and H a d-place real function such that $\mathrm{Dom}H = S_1 \times \ldots \times S_d$. The H-volume of a d-box $[\mathbf{a}, \mathbf{b}]$ with all its vertices in $\mathrm{Dom}H$ is given by

$$V_H([\mathbf{a}, \mathbf{b}]) = \sum \mathrm{sign}(\mathbf{c}) H(\mathbf{c}),$$

where the sum is taken over all vertices $\mathbf{c}$ of $[\mathbf{a}, \mathbf{b}]$ and

$$\mathrm{sign}(\mathbf{c}) = \begin{cases} 1, & \text{if } c_k = a_k \text{ for an even number of } k\text{'s}, \\ -1, & \text{if } c_k = a_k \text{ for an odd number of } k\text{'s}. \end{cases}$$

As an example, when $d = 2$, the H-volume of a rectangle $[a_1, b_1] \times [a_2, b_2]$ satisfying $(a_1, a_2), (a_1, b_2), (b_1, a_2), (b_1, b_2) \in S_1 \times S_2$ is given by

$$V_H([a_1, b_1] \times [a_2, b_2]) = H(b_1, b_2) - H(b_1, a_2) - H(a_1, b_2) + H(a_1, a_2).$$

We say that a d-place real function H is d-increasing if $V_H([\mathbf{a}, \mathbf{b}]) \geq 0$ for all d-boxes $[\mathbf{a}, \mathbf{b}]$ whose vertices lie in $\mathrm{Dom}H$.

Let H be a d-place real function with domain $\text{Dom}H = S_1 \times \ldots \times S_d$ where each S_k has a least element a_k, that is, $a_k \leq s$ for every $s \in S_k$. We say that H is grounded if $H(\mathbf{t}) = 0$ for all $\mathbf{t} \in \text{Dom}H$ such that $t_k = a_k$ for at least one $k \in \{1, \ldots, d\}$. If each S_k is non-empty and has greatest element b_k, then we say that H has margins and the 1-dimensional margins of H are the functions H_k given by

$$\begin{aligned} \text{Dom}H_k = S_k \subseteq \overline{\mathbb{R}} &\longrightarrow \mathbb{R} \\ x &\mapsto H(b_1, \ldots, b_{k-1}, x, b_{k+1}, \ldots, b_d). \end{aligned}$$

Higher dimensional margins are defined by fixing fewer places in H.

2.5.2 Definition of a d-copula

We now have the notation and terminology required to proceed with the definition of a copula.

Definition 1. *A d-dimensional copula is a function C with the following properties:*

1. $\text{Dom}C$ *is the unit d-cube* $\mathbf{I}^d$,
2. *C is grounded and d-increasing,*
3. *C has 1-dimensional margins C_k, $1 \leq k \leq d$, such that $C_k(u) = u$ for all $u \in S_k$.*

Equivalently, a d-copula is a function C from $\mathbf{I}^d$ to $\mathbf{I}$ satisfying

a. *for every* $\mathbf{u} \in \mathbf{I}^d$,

$$\textit{if at least one coordinate of } \mathbf{u} \textit{ is } 0, \ \textit{then } C(\mathbf{u}) = 0,$$

and

$$\textit{if all coordinates of } \mathbf{u} \textit{ are 1 except } u_k, \ \textit{then } C(\mathbf{u}) = u_k;$$

b. *for every* $\mathbf{a}, \mathbf{b} \in \mathbf{I}^d$ *such that* $\mathbf{a} \leq \mathbf{b}$,

$$V_C([\mathbf{a}, \mathbf{b}]) \geq 0.$$

2.5.3 Bivariate copulas

Since in the last paper of the appendix, entitled 'Measures of tail asymmetry for bivariate copulas', only bivariate copulas are considered, in the remainder of this section we concentrate on that class of copulas. We start with the definition of a bivariate copula.

Definition 2. *A bivariate copula, C, is a function from $\mathbf{I}^2$ to $\mathbf{I}$ with the following properties*

a. for every $u, v \in \mathbf{I}$,

$$C(u, 0) = 0 = C(0, v)$$

and

$$C(u, 1) = u, \text{ and } C(1, v) = v;$$

b. for every $u_1, u_2, v_1, v_2 \in \mathbf{I}$ such that $u_1 \leq u_2$ and $v_1 \leq v_2$,

$$C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0.$$

The first examples of bivariate copulas that we will consider are the independence copula, Π, and the Fréchet-Hoeffding lower, W, and upper, M, bounds. These three copulas are defined as

$$\Pi(u, v) = uv, \quad W(u, v) = \max\{u + v - 1, 0\}, \quad M(u, v) = \min\{u, v\}.$$

Plots of these functions are shown in Figure 2.1. Later we will justify the name given to Π. With regard to the copulas W and M, for every copula C and every $(u, v) \in \mathbf{I}^2$,

$$W(u, v) \leq C(u, v) \leq M(u, v).$$

This inequality is known as the Fréchet-Hoeffding bounds inequality.

One of the most important results in copula theory is Sklar's theorem, mentioned previously. This theorem explains the relationship between multivariate distributions and their univariate marginal distributions, and is the foundation for most of the applications of copulas in Statistics.

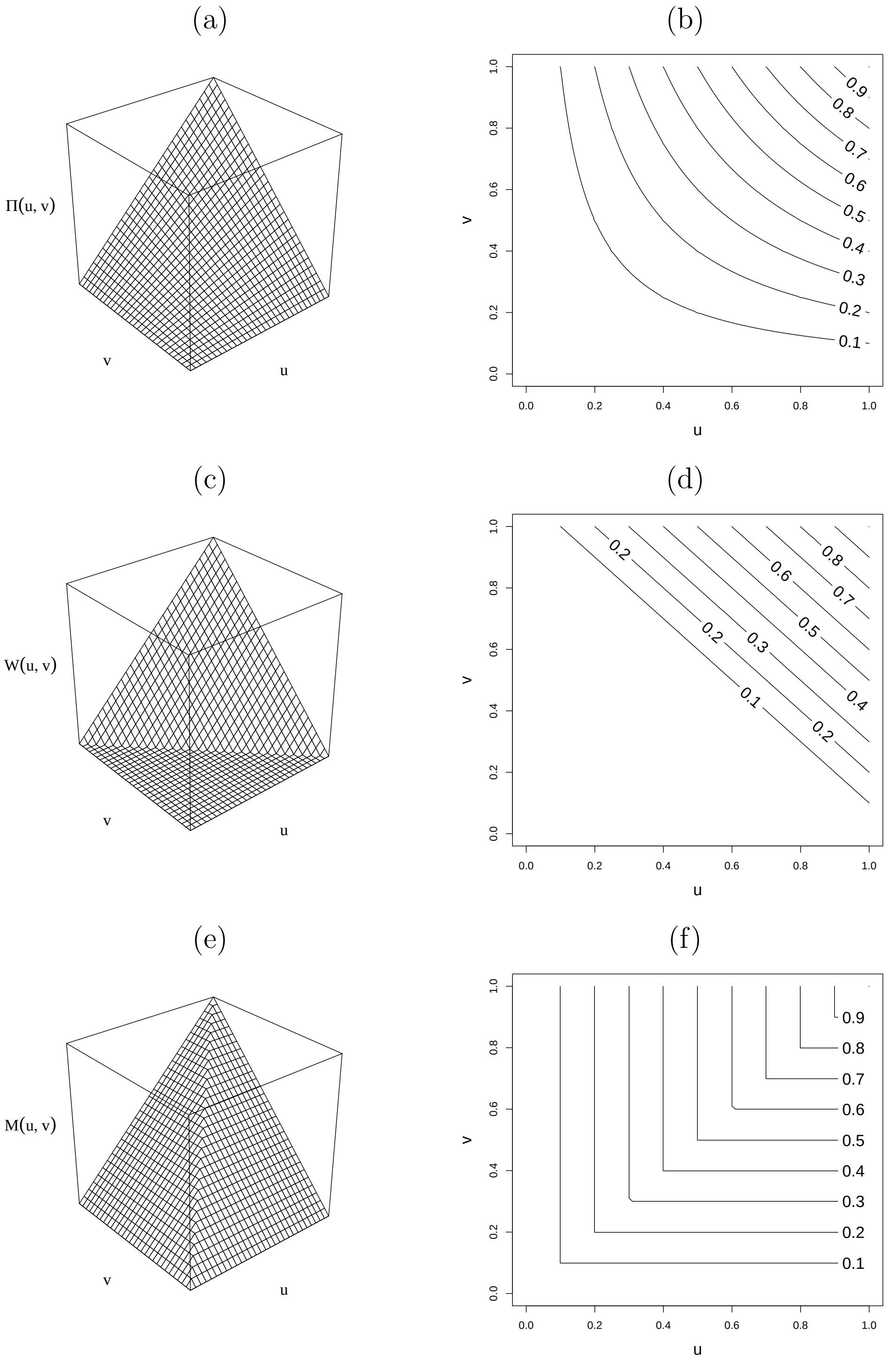


Figure 2.1: 3-dimensional plots (panels a, c and e) and contour plots (panels b, d and f) of the copulas Π, W and M.

Sklar's theorem. *Let F denote a joint distribution function with marginal distribution functions F_1 and F_2. Then there exists a copula C such that, for all $x, y \in \mathbb{R}$,*

$$F(x,y) = C(F_1(x), F_2(y)). \tag{2.7}$$

If F_1 and F_2 are continuous, then C is unique; otherwise, C is uniquely determined on $\mathrm{Ran}F_1 \times \mathrm{Ran}F_2$*. Conversely, if C is a copula and F_1 and F_2 are distribution functions, then the function F defined by (2.7) is a joint distribution function with margins F_1 and F_2.*

Equation (2.7) provides an expression for a joint distribution function in terms of a copula and two given univariate distribution functions. It can be inverted to obtain an expression for a copula in terms of a joint distribution function and the inverses of its two marginal distribution functions. However, a marginal distribution function does not always have an inverse, and thus we require the concept of quasi-inverse.

Definition 3. *Given a univariate distribution function F, a quasi-inverse of F is any function $F^{(-1)}$ with domain $\mathbf{I}$ such that:*

1. *if u is in* $\mathrm{Ran}F$*, then $F^{(-1)}(u)$ is any number $x \in \mathbb{R}$ such that $F(x) = u$, in other words, for all $u \in \mathrm{Ran}F$, $F(F^{(-1)}(u)) = u$;*

2. *if u is not in* $\mathrm{Ran}F$*, then*

$$F^{(-1)}(u) = \inf\{x : F(x) \geq u\}.$$

We are now in a position to state a corollary of Sklar's theorem which can be used to construct copulas from joint distribution functions.

Corollary. *Let F be a joint distribution function with continuous margins F_1 and F_2, and let $F_1^{(-1)}$ and $F_2^{(-1)}$ be their respective quasi-inverses. If C is the copula satisfying Equation (2.7), then for every $(u,v) \in \mathbf{I}^2$,*

$$C(u,v) = F(F_1^{(-1)}(u), F_2^{(-1)}(v)).$$

Sklar's theorem can be restated in terms of random variables and their distribution functions as follows.

Sklar's theorem. *Let X and Y be random variables with distribution functions F_1 and F_2, respectively, and joint distribution function F. Then there exists a copula C such that (2.7) holds. If F_1 and F_2 are continuous, then C is unique. Otherwise, C is uniquely determined on* $\mathrm{Ran}F_1 \times \mathrm{Ran}F_2$.

The copula C in the last theorem is referred to as the copula of X and Y.

Previously we defined the independence copula as $\Pi(u, v) = uv$. Its name is a consequence of the result that two random variables X and Y are independent if, and only if, their copula is the independence copula.

We continue with the definitions of absolutely continuous and singular copulas.

Definition 4. *Given a copula C, let*

$$C(u, v) = A_C(u, v) + S_C(u, v),$$

with

$$A_C(u, v) = \int_0^u \int_0^v \frac{\partial^2}{\partial s \partial t} C(s, t) \mathrm{d}t \mathrm{d}s$$

and $S_C(u, v) = C(u, v) - A_C(u, v)$. If $C = A_C$ on $\mathbf{I}^2$, then C is absolutely continuous, whereas if $C = S_C$ on $\mathbf{I}^2$, then C is singular. Otherwise, C has an absolutely continuous component A_C and a singular component S_C.

For an absolutely continuous copula C, the function

$$c(u, v) = \frac{\partial^2}{\partial u \partial v} C(u, v)$$

is called the density of C. The support of a copula is the complement of the union of all open subsets of $\mathbf{I}^2$ with C-volume zero. When the support of C is $\mathbf{I}^2$ we say that C has full support. When C is singular, its support has Lebesgue measure zero.

For a copula, C, its associated survival copula, C_R, is defined via the equation

$$C_R(u, v) = u + v - 1 + C(1 - u, 1 - v),$$

where the subscript R denotes reflection. Thus, we have a relationship between the univariate and joint survival functions analogous to the one between univariate and joint distribution functions.

Given two random variables, X and Y, let C denote the copula associated with them and $\bar{F}$, $\bar{F}_1$ and $\bar{F}_2$ their joint and univariate survival functions, respectively. Then

$$\bar{F}(x,y) = C_R(\bar{F}_1(x), \bar{F}_2(y)),$$

and we refer to C_R as the survival copula of X and Y.

2.5.4 Symmetry in bivariate copulas

The concept of symmetry is clear-cut for univariate distributions but is not so for multivariate ones. Focusing on the 2-dimensional case, we consider three different forms of symmetry that have been proposed within the literature.

Given two random variables, X and Y, and a point $(a,b) \in \mathbb{R}^2$, we say that (X,Y) is:

- *marginally symmetric* about (a,b) if X and Y are symmetric about a and b, respectively;
- *radially symmetric* about (a,b) if the joint distribution function of $X-a$ and $Y-b$ is the same as the joint distribution function of $a-X$ and $b-Y$;
- *jointly symmetric* about (a,b) if the four pairs of random variables $(X-a, Y-b)$, $(X-a, b-Y)$, $(a-X, Y-b)$ and $(a-X, b-Y)$ have a common joint distribution.

Another form of symmetry is exchangeability. We say that two random variables X and Y are exchangeable if the random vectors (X,Y) and (Y,X) are identically distributed. This notion of symmetry translates to copula theory as follows. Let X and Y be continuous random variables with copula C. Then X and Y are exchangeable if, and only if, $C(u,v) = C(v,u)$ for all $u, v \in \mathbf{I}$. A copula satisfying this last statement will be referred to as being symmetric.

2.5.5 Families of bivariate copulas

This subsection considers two popular families of copulas which provide a wide variety of copulas that can be very useful when building stochastic models with properties such as heavy tails, asymmetries, etc.

The first family is that of elliptical copulas, associated with elliptical distributions. A random vector $\mathbf{X} = (X, Y)$ is said to have an elliptical distribution with mean vector $\mu \in \mathbb{R}^2$, covariance matrix $\Sigma = (\sigma_{ij})$ and generator $g : [0, +\infty) \to [0, +\infty)$, and one denotes the fact by $\mathbf{X} \sim \mathcal{E}(\mu, \Sigma, g)$, if it can be expressed in the form

$$X = \mu + RAU,$$

where $AAT = \Sigma$ is the Cholesky decomposition of Σ, U is a 2-dimensional random vector uniformly distributed on the unit circle $S_1 = \{(u_1, u_2) \in \mathbb{R}^2 : u_1^2 + u_2^2 = 1\}$, and R is a positive random variable independent of U, with density given, for every $r > 0$, by

$$f_g(r) = 2\pi r g\left(r^2\right).$$

Although $\mathbf{X}$ does not always have a density function, if the density does exist it has the form

$$|\Sigma|^{-1/2} g\left(\frac{1}{2}(\mathbf{x} - \mu)'\Sigma^{-1}(\mathbf{x} - \mu)\right),$$

For instance, when $\mathbf{X}$ has a bivariate normal distribution then $g(r) = (2\pi)^{-1}e^{-r/2}$. Another member of this family is the bivariate Student t distribution with $g(r) = c(1 + r/\nu)^{-(\nu+2)/2}$, where ν denotes the degrees of freedom and c is a normalising constant.

Consider now an elliptical random vector, (X, Y), then the copula of X and Y is an elliptical copula. A closed form for elliptical copulas is generally not available. The bivariate normal copula, for instance, is given by

$$C_\rho(u, v) = \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{s^2 - 2\rho st + t^2}{2(1-\rho^2)}\right) \mathrm{d}s\mathrm{d}t,$$

where Φ denotes the distribution function of the standard normal distribution and $\rho \in (-1, 1)$ denotes the correlation coefficient between X and Y. Likewise, the bivariate Student t copula is obtained as

$$C_{\rho,\nu}(u, v) = \int_{-\infty}^{T_\nu^{-1}(u)} \int_{-\infty}^{T_\nu^{-1}(v)} \frac{1}{2\pi\sqrt{1-\rho^2}} \left(1 + \frac{s^2 - 2\rho st + t^2}{\nu(1-\rho^2)}\right)^{-(\nu+2)/2} \mathrm{d}s\mathrm{d}t,$$

with, as before, T_ν denoting the distribution function of a univariate t distribution with ν degrees of freedom.

Before considering Archimedean copulas, we introduce some extra terminology and notation. Any convex, strictly decreasing and continuous function ψ from $[0, +\infty]$ to $\mathbf{I}$ satisfying $\psi(0) = 1$ will be called an Archimedean generator. Given such a function ψ, its pseudo-inverse, with domain $[0, +\infty)$ and range $\mathbf{I}$, is given by

$$\psi^{[-1]}(t) = \begin{cases} \psi^{-1}(t), & \text{if } 0 \leq t \leq \psi(0), \\ 0, & \text{if } \psi(0) \leq t \leq +\infty. \end{cases}$$

A bivariate copula C is called Archimedean if it admits the representation

$$C(u, v) = \psi^{[-1]}(\psi(u) + \psi(v)),$$

for an Archimedean generator ψ. The importance of Archimedean copulas results from the ease with which they can be constructed, their wide variety, and their appealing properties. Two such properties are that Archimedean copulas are symmetric and associative; that is, $C(C(u, v), w) = C(u, C(v, w))$ for every $u, v, w \in \mathbf{I}$. Two particular families of bivariate Archimedean copulas are the:

- Frank family (Frank, 1978),

  $$-\frac{1}{\theta} \log\left(1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{e^{-\theta} - 1}\right), \ \theta \in \mathbb{R} \setminus \{0\},$$

  with generator $-\log((e^{-\theta t} - 1)/(e^{-\theta} - 1))$;

- BB2 family (Joe, 1997, p. 150–151),

  $$\left[1 + \delta^{-1} \log\left(e^{\delta u^{-\theta}} + e^{\delta v^{-\theta}} - 1\right)\right]^{1/\theta}, \ \theta, \delta > 0,$$

  with generator $e^{\delta(t^{-\theta} - 1)} - 1$.

2.5.6 Vine copulas

One of the main reasons why copulas have received so much interest in the statistical literature is that they can be used to model dependencies and marginal distributions separately. Nevertheless, the elliptical and Archimedean copulas considered so far do not allow for different forms of dependence between variables. It is in this context that vine copulas come to the fore.

In Joe (1996), a probabilistic construction of multivariate distribution functions was given by iteratively mixing conditional distributions. This was the first appearance of a pair-copula construction. In Bedford & Cooke (2001, 2002), that construction was rephrased in terms of the graph theory notion of a regular vine. Aas et al. (2009) used the pair-copula construction to obtain flexible multivariate copulas based on normal and t copulas. Results presented at four recent workshops on vine copulas have been published in Kurowicka & Joe (2010). Before presenting the definition of a vine copula, we need to consider the pair-copula construction and introduce various concepts from graph theory.

Let C be a bivariate copula of an absolutely continuous bivariate random vector (X, Y) with joint distribution function F, marginal distribution functions F_1 and F_2, joint density f, and marginal densities f_1 and f_2. Denoting the density of C by c_{12}, the joint density of (X_1, X_2) can be expressed as

$$f(x_1, x_2) = c_{12}(F_1(x_2), F_2(x_2))f_1(x_1)f_2(x_2),$$

and the conditional density as

$$f(x_2|x_1) = c_{12}(F_1(x_2), F_2(x_2))f_2(x_2).$$

This is the germ of the pair-copula construction. We can represent a joint density $f(x_1, \ldots, x_d)$ as a product of pair-copula densities and marginal densities. For instance, when $d = 3$, one possible decomposition is

$$f(x_1, x_2, x_3) = f_{3|12}(x_3|x_1, x_2)f_{2|1}(x_2|x_1)f_1(x_1).$$

However,

$$\begin{aligned}
f_{2|1}(x_2|x_1) &= c_{12}(F_1(x_1), F_2(x_2))f_2(x_2), \\
f_{13|2}(x_1, x_3|x_2) &= c_{13|2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2))f_{1|2}(x_1|x_2)f_{3|2}(x_3|x_2), \\
f_{3|2}(x_3|x_2) &= c_{23}(F_2(x_2), F_3(x_3))f_3(x_3), \\
f_{3|12}(x_3|x_1, x_2) &= c_{13|2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2))f_{3|2}(x_3|x_2), \\
&= c_{13|2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2))c_{23}(F_2(x_2), F_3(x_3))f_3(x_3).
\end{aligned}$$

Thus,

$$\begin{aligned} f(x_1, x_2, x_3) &= f_1(x_1)f_2(x_2)f_3(x_3) \text{ (marginals)} \\ &\times c_{12}(F_1(x_1), F_2(x_2))c_{23}(F_2(x_2), F_3(x_3)) \text{ (unconditional pairs)} \\ &\times c_{13|2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2)) \text{ (conditional pair).} \end{aligned}$$

The general expression for the pair-copula decomposition in d dimensions is

$$f(x_1, \ldots, x_d) = \prod_{j=1}^{d-1} \prod_{i=1}^{d-j} c_{i,(i+j)|(i+1),\ldots,(i+j-1)} \prod_{k=1}^{d} f_k(x_k), \tag{2.8}$$

with $c_{i,j|i_1\ldots,i_k} = c_{i,j|i_1\ldots,i_k}(F(x_i|x_{i_1}, \ldots, x_{i_k}), F(x_j|x_{i_1}, \ldots, x_{i_k}))$ for $i < j$ and $i_1 < \ldots < i_k$. Note, however, that this decomposition is not unique. The complexity of Equation (2.8) prompted the introduction of the regular vine structure as a means of organising the pair-copula construction.

In graph theory, a graph G is an ordered pair (N, E) consisting of a set N of nodes and a set E of edges. An edge is a 2-element subset of N. A path in a graph is a sequence of vertices such that there is an edge from each of its vertices to the next vertex in the sequence. A path is called a cycle if the first and last vertices in the sequence coincide. Two vertices in a graph are connected if a path exists from one of them to the other, and a graph is said to be connected if every pair of its vertices is connected. A graph is referred to as being a tree if it is connected and has no cycles.

A d-dimensional regular vine is a sequence of $d - 1$ trees where:

- tree 1 has d nodes and $d - 1$ edges;
- tree j has $d + 1 - j$ nodes, which correspond to the edges of tree $j - 1$, and $d - j$ edges,

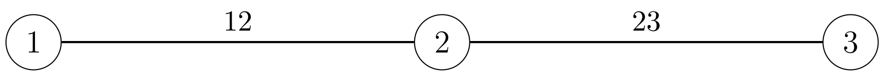
with a proximity condition: if two nodes in tree $j + 1$ are joined by an edge, the corresponding edges in tree j share a node.

A d-dimensional regular vine distribution is defined by a d-dimensional regular vine tree structure in which each node in the first tree corresponds to a marginal density and each edge corresponds to a pair-copula density. The density of a regular vine distribution is defined by the product of the pair-copula densities over the $d(d-1)/2$ edges identified by the regular vine tree structure and the product of the marginal densities.

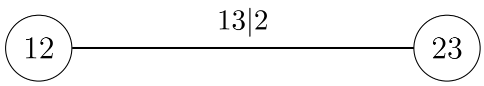
According to the characteristics of the regular vine structure, we can distinguish between C-vines (canonical) and D-vines (drawable). We say that a regular vine structure is canonical if tree j has a unique node that is connected to $d-j$ nodes. A regular vine structure is said to be drawable if no node in any tree is connected to more than two nodes.

As an example of a D-vine structure, which can be used to organise the $d = 3$ dimensional pair-copula construction considered above, is:

- Tree 1:

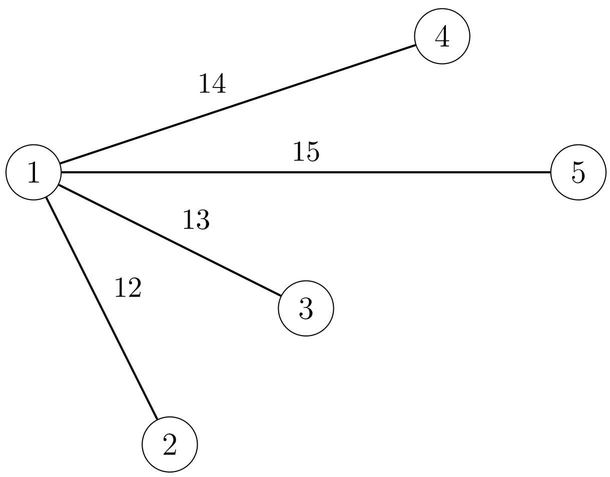


- Tree 2:

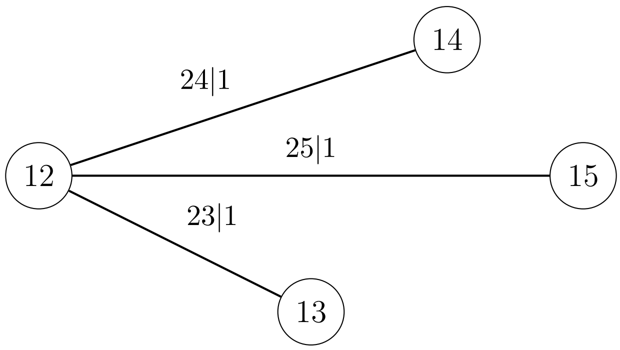


An example of a $d = 5$ dimensional C-vine structure is given below. The pair-copula decomposition for this example is far more complex than the one for $d = 3$.

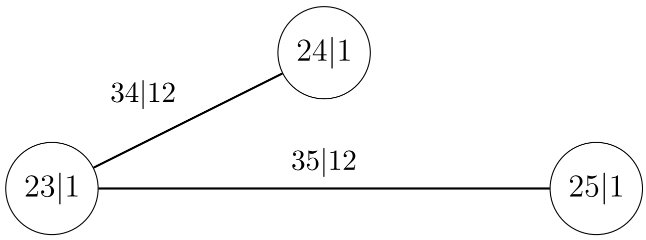
- Tree 1:



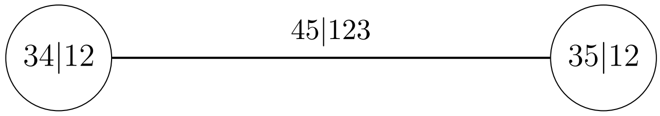
- Tree 2:



- Tree 3:



- Tree 4:



Thus, the pair-copula decomposition for this example is

$$f_{12345} = \left(\prod_{i=1}^{5} f_i\right)\left(\prod_{j=2}^{5} c_{1j}\right)\left(\prod_{k=3}^{5} c_{2k|1}\right)\left(\prod_{l=4}^{5} c_{3l|12}\right) c_{45|123}.$$

The generic steps involved in fitting a pair-copula decomposition to a data set are as follows. An appropriate vine structure is chosen, and then a suitable family of copulas. The parameters of the selected copula family are then estimated. The main problems associated with this general scheme are that, i) there is a huge number of vine structures from which to choose, and ii), the number of pair-copulas we have to select in high dimensions is also vast. Asymmetry measures for bivariate copulas can be helpful when resolving the latter issue.

2.5.7 Asymmetry measures for bivariate copulas

As we explained in Section 2.5.4, there are various ways the concept of symmetry can be extended from the univariate context to the multivariate. As far as copulas are concerned, the most studied notion of symmetry is that of exchangeability.

The first measure of nonexchangeability for bivariate random vectors was proposed in Nelsen (2007). Nelsen related the concept of nonexchangeability to the notion of an asymmetric copula as follows. Given a continuous random vector (X, Y), where X and Y are identically distributed, with joint distribution function F and copula C, the set of values of $|F(x, y)-F(y, x)|$ for $x, y \in \mathbb{R}$ is the same as the set of values of $|C(u, v) - C(v, u)|$ for $u, v \in \mathbf{I}$. Thus, for continuous and identically distributed random variables, nonexchangeability implies asymmetry of the copula. Nelsen's measure of asymmetry is

$$3 \sup_{u,v\in\mathbf{I}} \{|C(u, v) - C(v, u)|\},$$

the multiplier 3 being chosen so as to obtain values between 0 and 1; a value of 0 corresponding to a symmetric copula. A generalised version of this measure was studied in Durante et al. (2010), who used L_p distances to define nonexchangeability measures and state properties that a measure of nonexchangeability should satisfy. Their measures are

$$d_p(C, C^t),\ p \in [1, +\infty],$$

where d_p is an L_p distance and $C^t(u, v) = C(v, u)$ for every $u, v \in \mathbf{I}$.

Dehgani et al. (2012) explored measures of radial asymmetry for bivariate random vectors. The structure of their paper is similar to that of Durante et al. (2010), in that they state various properties that a measure of radial asymmetry should satisfy and then define the measures

$$d_p(C, C_R),\ p \in [1, +\infty].$$

Another recent approach to measure radial asymmetry is that given in Krupskii & Joe (2012). They develop tail-weighted dependence measures and then consider the difference between the tail-weighted dependence measure in the lower and the upper tails as a measure of radial asymmetry.

2.6 Overview of the remainder of the thesis

The remainder of the thesis is structured as follows. In Section 3 the motivation for, and main objectives of, the work summarised in the four papers contained in the appendix are described. Section 4 provides a joint discussion of the results presented in the four papers. Finally, in Section 5, conclusions are drawn and prospects for future research discussed.

3 Motivation and Objectives

In this section we consider the motivation for, and objectives of, the research work summarised in the four papers contained in the appendix.

Skew-t distributions via the sinh-arcsinh transformation

The focus of this paper is the SAS-t family, a skew-symmetric family of distributions, based on Student's t distribution, obtained using an adaptation of the sinh-arcsinh transformation. The primary motivation for proposing this family was to provide a flexible extension of Student's t distribution with members displaying a wide range of shapes. Those shapes are regulated by two parameters controlling the asymmetry and tailweight, respectively, of the distributions within the family.

As explained in Section 2.3, approaches to extending Student's t distribution to obtain flexible families including symmetric as well as asymmetric members have received considerable attention within the literature of late. The motivation underpinning this activity has often been stimulated by the need, especially within the modelling of financial data, for distributions with tails heavier than those of the normal distribution and displaying varying levels of asymmetry. With its tails being at least as heavy as those of the normal distribution, Student's t distribution is then a natural choice for the symmetric base density. The other ingredient in the construction is the sinh-arcsinh transformation, which provides an appealing novel means with which to generate highly flexible skew-symmetric families from a base symmetric distribution. In the paper we show that the use of Student's t distribution together with the adaptation of the sinh-arcsinh transformation results in a highly flexible skew-symmetric family. Once defined, we aimed to establish the family's main properties, obtain results for the estimation of its parameters, apply it in the analysis of skew heavy-tailed data, and draw a comparison between it and two of its most popular existing skew-t competitors.

Skewness-invariant measures of kurtosis

This paper deals with measuring kurtosis in the presence of asymmetry. The classical coefficient of kurtosis, α_4, referred to in Section 2.4, does not have an obvious interpretation and its relationship with the classical coefficient of skewness complicates its meaning in the presence of asymmetry still further. Another inconvenience is its potential non-existence, since it is based on moments. This limitation is a particularly important one for very heavy-tailed distributions. Whilst alternative measures of kurtosis which can be used with asymmetric distributions have been proposed in the literature, little had previously been published regarding measures of kurtosis that are skewness-invariant. Our motivation was to fill this gap in the literature and identify measures of kurtosis that are skewness-invariant for families of distributions obtained via transformation of a base symmetric distribution. Once it was established that such measures do exist, we aimed to identify skewness-inducing transformations that leave certain measures of kurtosis invariant. In addition, we aimed to establish the conditions under which the transformations used would result in families of distributions for which certain measures of kurtosis would be skewness-invariant. A further objective was to identify illustrative families obtained using the transformation approach for which the measures of kurtosis considered are indeed skewness-invariant.

On Blest's measure of kurtosis adjusted for skewness

This paper is closely related to the previous one, and can be considered as an extension of its ideas to the moment-based context. The catalyst for the paper was the coefficient of kurtosis adjusted for asymmetry proposed in Blest (2003). Although the classical moment-based kurtosis measure, α_4, has the drawback of non-existence for distributions without a finite fourth moment, it is still widely used in many applied fields, and thus we, like Blest, perceived the need for skewness-invariant measures of kurtosis based on moments. This was the main motivation for our research work conducted on this topic. As objectives, we aimed first to study Blest's coefficient and establish whether or not it was indeed skewness-invariant. Once we established that Blest's measure does not completely remove the effects of asymmetry, our next goal was to develop an alternative moment-based measure of kurtosis which would better eliminate those effects. We also sought to obtain lower bounds for the two competing measures as functions of the classical coefficient of skewness. Our final objective was to use Monte Carlo simulation to identify the best-performing sample versions of the two measures obtained by plugging-in different moment estimators available in popular statistical packages.

Measures of tail asymmetry for bivariate copulas

This last paper considers tail asymmetry measures for bivariate copulas. The primary motivation for our research into this topic was that such measures can be helpful when seeking appropriate bivariate copulas to use, for instance, in the first level of a vine structure (see Section 2.5). Our first objective was to provide a set of desirable properties that a measure of tail asymmetry should satisfy. Our second aim was to develop tail asymmetry measures capable of discriminating between different families of copulas when a choice between them has to be made. Finally, we sought to use real data sets to illustrate the levels of tail asymmetry that might be expected in practice.

4 Discussion of Results

Skew-t distributions via the sinh-arcsinh transformation

In the first of the four papers in the appendix we used a minor adaptation of the sinh-arcsinh transformation to derive a skew-symmetric family of distributions with Student's t as its symmetric member. In the adaptation used, the degrees of freedom parameter of the t distribution, ν, plays the role of a tailweight parameter and the tailweight parameter of the sinh-arcsinh transformation, δ, is set equal to one. The distributions within the resulting SAS-t family can assume a wide range of shapes with tails at least as heavy as those of the limiting normal distribution obtained as $\nu \to +\infty$. Being based on only a slight modification of the sinh-arcsinh transformation, the family inherits the appealing properties identified in Section 2.2.

A rather unappealing property of the SAS-t family is that distributions contained within it can have densities which are bimodal. However, using numerical methods to explore the derivative of the density, we established that bimodality only results when the value of the tailweight parameter ν is very small (< 0.35). Moreover, the density can be at most bimodal.

When considering the properties of the SAS-t distribution, we obtained a general expression for its nth moment. The formula involves the Gauss hypergeometric function and is therefore relatively complex. Using that formula we identified the attainable region of skewness and kurtosis for SAS-t distributions and established that almost all of the attainable region for the Azzalini-type skew-t distribution proposed in Azzalini & Capitanio (2003) is contained within it.

We also showed that the kurtosis measures of Moors (1988) and Hosking (1990) are skewness invariant for the distributions within the SAS-t family.

Maximum likelihood based inference for the location-scale extension of the SAS-t family involves the use of numerical optimisation methods. During applications involving small-sized samples we found that the likelihood surface needs to be explored quite extensively in order to ensure that the true maximum likelihood solution is identified. In addition, the profile log-likelihood functions for some of the parameters can be rather irregular. Infinite estimates of the tailweight parameter ν were also found to occur relatively frequently when the sample size is small. Such estimates are only slightly problematic from an arithmetical perspective. Their interpretation, nevertheless, is straightforward; they indicate that the best fitting member of the SAS-t family belongs to its skew-normal subclass. The real data example included in the paper illustrates that the SAS-t family can provide a better fit to skew heavy-tailed data than some of its skew-t competitors. The results from an extensive simulation study indicate that there is generally a high negative correlation between the location and asymmetry parameters, ξ and ε, as well as a high positive correlation between the scale and tailweight parameters, η and ν. However, as is shown in Jones & Anaya-Izquierdo (2011), such strong dependencies are not exclusive to the maximum likelihood estimates of the parameters of the SAS-t family.

When exploring multivariate extensions of the SAS-t family we investigated a particular multivariate skew-t, with just a single tailweight parameter, obtained using a t copula.

Skewness-invariant measures of kurtosis

In the second paper presented in the appendix, two classes of quantile-based measures of kurtosis are identified as being invariant to the skewness parameter of families of distributions arising from certain types of skewness inducing transformations. The two classes include various popular measures of kurtosis proposed as alternatives to the classical coefficient of kurtosis. In addition, we provide a sufficient condition for the classes of kurtosis measures to be skewness-invariant for families of distributions obtained via the transformation of a symmetric random variable. An example of a transformation leading to families of distributions for which the measures of kurtosis are skewness-invariant is the sinh-arcsinh transformation introduced in Section 2.2.

We also show that, for distributions obtained by transformation of scale, there exists an appropriately defined measure of kurtosis based on their density which is independent of skewness.

With regard to the classical moment-based approach to measuring kurtosis, the skewness adjusted kurtosis measure proposed in Blest (2003) is identified as being equal to the classical coefficient of kurtosis, α_4, minus an expression involving a sinh-arcsinh transformation of the classical coefficient of skewness.

On Blest's measure of kurtosis adjusted for skewness

In the third paper in the appendix the skewness adjusted measure of kurtosis proposed by Blest (2003) is shown not to be skewness invariant. After reconsidering the components of its construction, we propose an adaptation of Blest's measure and compare the performance of the two measures. We show that our adaptation of Blest's measure generally removes the influence of asymmetry slightly more successfully. Also, we provide lower bounds for both measures and show that the lower bound for our adaptation of Blest's measure is closer to being constant.

We also present the results obtained from a simulation study designed to identify the sample versions of Blest's measure and our adaptation of it that have lowest mean squared error. The sample versions of the two measures were those obtained by plugging in different sample moments computed in popular statistical packages such as R, STATA, MINITAB, SPSS, SAS, etc. The results indicate that the tailweight of the distribution from which the data are drawn has a considerable influence upon which sample versions are identified as performing best.

Measures of tail asymmetry for bivariate copulas

In the fourth of the papers appearing in the appendix we first identify desirable properties that measures of tail asymmetry should exhibit.

We then introduce three families of measures of tail asymmetry: two based on the univariate skewness of a projection, and another based on a distance measure between a copula and its reflected/survival copula. All three families have finite ranges to facilitate their interpretation. However, none of the three proposed measures satisfies all of the previously identified desirable properties. The first two,

ς_1 and ς_2, measure not only the degree of asymmetry but also the direction of the asymmetry. The third, ς_3, only indicates the degree of asymmetry. A disadvantage of the latter is that there are copulas with upper and lower tail order of infinity which attain the maximum value. Thus, some copulas considered extreme according to ς_3 are not considered extreme according to other tail asymmetry concepts.

Our results show that most tail asymmetry tends to occur for intermediate levels of positive or negative dependence. Moreover, the copulas attaining extreme tail asymmetry are found to depend on the particular measure employed. As a consequence, copulas with very different properties can have identical tail asymmetry values of the three measures.

Sample moments and quantiles, and empirical copulas, can be used to readily obtain sample versions of the three families. The asymptotic behaviour of ς_3 is not easily established, however.

Two real data sets are used to illustrate the levels of tail asymmetry that might be expected in practice. Our results suggest that in applications it would be sensible to employ more than one measure of tail asymmetry as each has quite different properties.

Joint discussion

In statistical modelling it is crucial to have flexible models capable of describing those features, such as asymmetry and varying degrees of kurtosis, often exhibited by real data. As an aid in this endeavour, the SAS-t family offers a highly flexible family containing distributions ranging from the symmetric to the highly skew with tails between those of the normal distribution and the very heavy. The simplicity of many of its properties, particularly those based on quantiles, makes the SAS-t family a potentially appealing one from both a mathematical statistics perspective as well as an applications one.

Copulas are a powerful tool with which to generate multivariate distributions from univariate ones. We used a special case of the t copula to extend the SAS-t family to a multivariate one, the bivariate case of which has a joint density which is relatively simple.

The questions as to what precisely kurtosis is, and how it should be measured, have vexed the minds of statisticians for over a century. What kurtosis represents in the presence of asymmetry is a problem of yet a higher order. We have considered measures of kurtosis that are skewness-invariant. Such measures can be used to summarise the kurtosis of distributions, like the SAS-t, obtained via the transformation of a symmetric base random variable, without having to be concerned about the symmetry, or lack thereof, of the distribution.

Although moment-based measures of kurtosis have the drawback of potential non-existence, they are nevertheless still widely used. Hence we perceive our results on Blest's skewness adjusted measure of kurtosis, and our adaptation of it, to be particularly relevant to potential users. Moreover, the values taken by sample measures of skewness and kurtosis can be highly useful when deciding on the sorts of distributions to employ in the modelling of real data. In this context, our results on the performance of the different estimators of Blest's kurtosis measure and our adaptation of it provide insight on their interpretation.

Similarly, asymmetry measures are useful aids when trying to choose between copulas with which to model multivariate data. We have proposed three different families of tail asymmetry measures. Having more than one such family available means that we can use them to obtain different types of information about the copulas we might employ.

5 Conclusions and Future Research

Skew-t distributions via the sinh-arcsinh transformation

We have derived and studied an appealing family of skew-symmetric distributions obtained by applying a special case of the sinh-arcsinh transformation to a base Student's t random variable. Many of the properties of the resulting SAS-t family have the same level of complexity as those of the base Student t distribution. Maximum likelihood inference reduces to an optimisation problem, with care needing to be taken when the size of the sample is small. After comparing it to skew-symmetric extensions of the t distribution introduced by Azzalini & Capitanio (2003) and Fernández & Steel (1998), it can be concluded that the two-piece t and the SAS-t distributions have roughly the same number of advantages in terms of tractability. However, the Azzalini-type skew-t distribution has at least two interesting geneses.

We also studied a multivariate extension of the SAS-t family obtained by transforming marginals using a special case of the t copula. A thorough investigation of the properties of potential multivariate extensions is still outstanding. Making use of the general form of the t copula, a multivariate family of distributions in d dimensions with d asymmetry and d degrees of freedom parameters can be obtained. At least for the bivariate case, the number of parameters needed to be estimated would not be prohibitive. Another way of extending the univariate family would be the following. Let $X_1, \ldots, X_d$ be independent and identically distributed random variables with $X_i \sim t_\nu$. Now consider the transformed random variables

$$Y_{\varepsilon_1,\nu} = \sinh\left(\sinh^{-1}(X_1) + \varepsilon_1\right),$$

$$Y_{\varepsilon_i,\nu,\delta_i} = \sinh\left(\delta_i^{-1}\left(\sinh^{-1}(X_i) + \varepsilon_i\right)\right),\ 2 \le i \le d,$$

with $\varepsilon_1, \ldots, \varepsilon_d \in \mathbb{R}$ controlling asymmetries and $\nu, \delta_2, \ldots, \delta_d > 0$ controlling tail-weights. It would be of interest to study the multivariate family of distributions obtained by connecting such random variables using the t copula.

A related line of research that we are presently pursuing is the development of a likelihood ratio test, based on the use of the SAS-normal model of Jones & Pewsey (2009), as an alternative to the four tests considered by Lehmann (2009) for the classic problem of testing the hypothesis of a common underlying population for two independent samples. At present we are in the process of performing an extensive Monte Carlo experiment designed to explore the operating characteristics of the five tests.

A further potential line of future research would be to explore alternatives to the sinh-arcsinh transformation.

Skewness-invariant measures of kurtosis

We have identified measures of kurtosis that are invariant to the skewness parameter of families of distributions derived using certain types of transformations. A sufficient condition is given in terms of the transformation used to obtain a given family.

A potential line of future research would be to investigate the sample versions of the measures of kurtosis. First, choices would have to be made amongst the various quantile estimators available. Dielman et al. (1994) provide a useful survey of quantile estimators, while Hyndman & Fan (1996) study quantile estimators implemented in statistical packages. After the identification of suitable quantile estimators, a Monte Carlo study would need to be performed in order to compare the mean squared errors of the different sample versions. Moreover, a study of the robustness of these sample versions would provide insight into their skewness-invariance in practice.

In our work on skewness-invariant kurtosis measures we focused on quantile-based measures for the univariate context. Another interesting line of future research would be the development of skewness-invariant quantile-based measures of kurtosis for multivariate distributions. First of all the definition of a multivariate quantile with which to work would need to be chosen. A recently proposed definition is given in Hallin et al. (2010). Of course, it would be necessary to establish that the proposed measures were indeed measuring kurtosis. Their skewness-invariance might also be found to depend on the definition of multivariate asymmetry employed.

On Blest's measure of kurtosis adjusted for skewness

We have studied a moment-based measure of kurtosis due to Blest (2003) as well as a modification of it. Although neither measure is skewness-invariant, our adaptation slightly outperforms Blest's in terms of its ability to remove the influence of skewness. Lower bounds for both measures were also derived. From the results of a study into the performance of different sample versions of the two measures, we concluded that the sample version identified as being best depends on the tail behaviour of the population from which the data are drawn.

Since the classical coefficient of kurtosis is widely used, the development of a moment-based measure of kurtosis that might be skewness-invariant, at least for certain kinds of distributions, would be another topic for future research.

Measures of tail asymmetry for bivariate copulas

We have identified a desirable set of properties for tail-asymmetry measures, and proposed three families of measures of tail asymmetry obtained using the univariate skewness of a projection and the distance between a copula and its survival/reflected copula. As the three families have different advantages and disadvantages, in practice it is sensible to make use of more than just one of them. With two examples involving real data we show that tail asymmetry is more pronounced than nonexchangeability.

A potential focus for future research is the development of measures of kurtosis based on $\mathrm{E}\left[|U+V-1|^k\right]$ with $k > 2$ and $(U,V) \sim C$. The minimum occurs for countermonotonic random variables, that is, random variables with copula W.

The maximum perhaps occurs for comonotonic random variables, i.e. whose copula is M. Geometrically, a singular copula with support on the line segments

- $v = u$ for $0 < u < a$,
- $v = 1 + a - u$ for $a < u < 1$,

or

- $v = 1 - a - u$ for $0 < u < 1 - a$,
- $v = u$ for $1 - a < u < 1$,

with $0 < a < 1/2$, would have high kurtosis as defined above.

Bibliography

Aas, K., Czado, C., Frigessi, A. & Bakken, H. (2009). Pair-copula constructions of multiple dependence. *Insur. Math. Econ.* **44**, 182–198.

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

Anděl, J., Netuka, I. & Zvára, K. (1984). On threshold autoregressive processes. *Kybernetika* **20**, 89–106.

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. *Scand. J. Statist.* **32**, 159–188.

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

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

Balanda, K.P. & MacGillivray, H.L. (1990). Kurtosis and spread. *Can. J. Statist.* **18**, 17–30.

Bedford, T. & Cooke, R.M. (2001). Probability density decomposition for conditionally dependent random variables modeled by vines. *Ann. Math. Artif. Intel.* **32**, 245–268.

Bedford, T. & Cooke, R.M. (2002). Vines - a new graphical model for dependent random variables. *Ann. Statist.* **30**, 1031–1068.

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

Blest, D.C. (2003). A measure of kurtosis adjusted for skewness. *Aust. N. Z. J. Statist.* **45**, 175–179.

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

Branco, M.D. & Dey, D.K. (2001). A general class of multivariate skew-elliptical distributions. *J. Multivar. Anal.* **79**, 99–113.

Cherubini, U., Luciano, E. & Vecchiato, W. (2004). *Copula Methods in Finance.* Chichester: Wiley.

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

De Forest, E.L. (1882). On an unsymmetrical probability curve. *Analyst* **9**, 135–142, 161–168.

de Helguero, F. (1908). Sulla rappresentazione analitica delle statistiche abnormali. *Atti del IV Congresso Internazionale dei Matematici* **III**, 288–299.

de Moivre, A. (1738). *The Doctrine of Chances.* London: H. Woodfall, 2nd ed.

Dehgani, A., Dolati, A. & Úbeda-Flores, M. (2012). Measures of radial asymmetry for bivariate random vectors. *Statist. Pap.* DOI:10.1007/s00362-011-0425-y.

Dielman, T., Lowry, C. & Pfaffenberger, R. (1994). A comparison of quantile estimators. *Commun. Statist. Simul.* **23**, 355–371.

Durante, F., Klement, E.P., Sempi, C. & Úbeda-Flores, M. (2010). Measures of non-exchangeability for bivariate random vectors. *Statist. Pap.* **51**, 687–699.

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 formulæ (part I). *J. Roy. Statist. Soc.* **61**, 670–700.

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

Fernández, C. & Steel, M.F.J. (1998). On bayesian modelling of fat tails and skewness. *J. Am. Statist. Assoc.* **93**, 359–371.

Fisher, R.A. (1925). *Statistical Methods for Research Workers*. Edinburgh: Oliver & Boyd.

Frank, M.J. (1978). On the simultaneous associativity of $F(x, y)$ and $x+y-F(x, y)$. *Aeq. Math.* **18**, 266–267.

Fréchet, M. (1951). Sur les tableaux de corrélation dont les marges sont donées. *Ann. Univ. Lyon, Sect. A* **14**, 53–77.

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

Gibbons, J.F. & Mylroie, S. (1973). Estimation of impurity profiles in ion-implanted amorphous targets using joined half-Gaussian distributions. *Appl. Phys. Lett.* **22**, 568–569.

Hallin, M., Paindaveine, D. & Šiman, M. (2010). Multivariate quantiles and multiple-output regression quantiles: from L_1 optimization to halfspace depth. *Ann. Statist.* **38**, 635–669.

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

Helmert, F.R. (1875). Über die Bestimmung des wahrscheinlichen Fehlers aus einer endlichen Anzahl wahrer Beobachtungsfehler. *Z. Math. Phys.* **20**, 300–303.

Hosking, J.R.M. (1990). *L*-moments: analysis and estimation of distributions using linear combinations of order statistics. *J. Roy. Statist. Soc. Ser. B* **52**, 105–124.

Hyndman, R.J. & Fan, Y. (1996). Sample quantiles in statistical packages. *Am. Statist.* **50**, 361–365.

Joe, H. (1996). Families of m-variate distributions with given margins and $m(m-1)/2$ bivariate dependence parameters. In *Distributions with Fixed Marginals and Related Topics*, L. Rüschendorf, B. Schweizer & M. Taylor, eds. Hayward, CA: IMS, pp. 120–141.

Joe, H. (1997). *Multivariate Models and Dependence Concepts*. London: Chapman & Hall.

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

Johnson, N.L., Kotz, S. & Balakrishnan, N. (1994a). *Continuous Univariate Distributions*, vol. 1. New York: Wiley, 2nd ed.

Johnson, N.L., Kotz, S. & Balakrishnan, N. (1994b). *Continuous Univariate Distributions*, vol. 2. New York: Wiley, 2nd ed.

Jones, M.C. (2001). A skew t distribution. In *Probability and Statistical Models with Applications*, C. Charalambides, M. Koutras & N. Balakrishnan, eds. London: Chapman & Hall/CRC, pp. 269–277.

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

Jones, M.C. & Faddy, M.J. (2003). A skew extension of the t-distribution, with applications. *J. Roy. Statist. Soc. Ser. B* **65**, 159–174.

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

Kapteyn, J.C. & van Uven, M.J. (1916). *Skew Frequency Curves in Biology and Statistics.* Gröningen: Gröningen Astronomical Laboratory.

Krupskii, P. & Joe, H. (2012). Tail-weighted measures of dependence and deviation from normal copula. Submitted.

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

Laplace, P.S. (1774). Mémoire sur la probabilité des causes par les événements. In *Mémoires de Mathématique et de Physique, Presentés à l'Académie Royale des Sciences, par divers Savans & lûs dans ses Assemblées*, vol. 6. pp. 621–656.

Lehmann, E.L. (2009). Parametric versus nonparametrics: two alternative methodologies. *J. Nonparametr. Statist.* **21**, 397–405.

Lüroth, J. (1876). Vergleichung von zwei Werten des wahrscheinlichen Fehlers. *Astron. Nachr.* **87**, 209–220.

Ma, Y. & Genton, M.G. (2004). Flexible class of skew-symmetric distributions. *Scand. J. Statist.* **31**, 459–468.

Moors, J.J.A. (1986). The meaning of kurtosis: Darlington reexamined. *Am. Statist.* **40**, 283–284.

Moors, J.J.A. (1988). A quantile alternative for kurtosis. *J. Roy. Statist. Soc. Ser. D* **37**, 25–32.

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

Nelsen, R.B. (1999). *An Introduction to Copulas.* New York: Springer.

Nelsen, R.B. (2006). *An Introduction to Copulas.* New York: Springer, 2nd ed.

Nelsen, R.B. (2007). Extremes of nonexchangeability. *Statist. Pap.* **48**, 329–336.

Nelson, L.S. (1964). The sum of values from a normal and truncated normal distribution. *Technometrics* **6**, 469–471.

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

Pearson, K. (1895). Contributions to the mathematical theory of evolution, II. Skew variation in homogeneous material. *Philos. T. Roy. Soc. Lond. A* **186**, 343–414.

Pearson, K. (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philos. Mag.* **50**, 157–175.

Pearson, K. (1901). Contributions to the mathematical theory of evolution, X. Supplement to a memoir on skew variation. *Philos. T. Roy. Soc. Lond. A* **197**, 443–459.

Pearson, K. (1905). Skew variation, a rejoinder. *Biometrika* **4**, 169–212.

Pearson, K. (1916). Contributions to the mathematical theory of evolution, XIX. Second supplement to a memoir on skew variation. *Philos. T. Roy. Soc. Lond. A* **216**, 429–457.

Poisson, S.D. (1837). *Recherches sur la Probabilité des Jugements en Matière Criminelle et en Matière Civile.* Paris: Bachelier.

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

Rietz, H.L. (1922). Frequency distributions obtained by certain transformations of normally distributed variates. *Ann. Math.* **23**, 292–300.

Roberts, C. (1966). A correlation model useful in the study of twins. *J. Am. Statist. Assoc.* **61**, 1184–1190.

Salvadori, G., De Michele, C., Kottegoda, N.T. & Rosso, R. (2007). *Extremes in Nature. An Approach Using Copulas.* Dordrecht: Springer.

Schweizer, B. & Sklar, A. (1983). *Probabilistic Metric Spaces.* New York: North-Holland.

Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris* **8**, 229–231.

Sklar, A. (1973). Random variables, joint distribution functions, and copulas. *Kybernetika* **9**, 449–460.

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

Student (1908). The probable error of a mean. *Biometrika* **6**, 1–25.

Tadikamalla, P.R. & Johnson, N.L. (1982). Systems of frequency curves generated by transformations of logistic variables. *Biometrika* **69**, 461–465.

Thiele, T.N. (1889). *Forelæsninger over Almindelig Lagttagelseslære: Sandsynlighedsregning og Mindste Kvadraters Methode.* Copenhagen: Reitzel.

Tukey, J.W. (1957). On the comparative anatomy of transformations. *Ann. Math. Statist.* **28**, 602–632.

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

van Zwet, W.R. (1964). Convex transformations of random variables. In *Mathematical Center Tracts 7.* Amsterdam: Mathematisch Centrum.

Whittaker, E.T. & Robinson, G. (1965). *The Calculus of Observations: A Treatise on Numerical Mathematics*. London: Blackie & Son, 4th ed.

Wicksell, S.D. (1917). On the genetic theory of frequency. *Arkiv. Mate. Astron. Fysik.* **12**, no. 20.

Wicksell, S.D. (1923). Contributions to the analytical theory of sampling. *Arkiv. Mate. Astron. Fysik.* **17**, no. 19.

Appendix: Papers

Test (2011) 20:630–652
DOI 10.1007/s11749-010-0222-2

ORIGINAL PAPER

Skew t distributions via the sinh-arcsinh transformation

J.F. Rosco · M.C. Jones · Arthur Pewsey

Received: 23 July 2010 / Accepted: 10 November 2010 / Published online: 26 November 2010

Abstract A version of the sinh-arcsinh transformation is used to generate a skew extension of Student's t distribution which provides an alternative to previously proposed skew t distributions. The basic properties of the resulting sinh-arcsinhed t family of distributions are presented, many of them effectively having the same level of complexity as their Student t counterparts. Quantile-based measures, which come to the fore due to the non-existence of moments, are readily available. The parameters of the distribution have clear interpretations. Limiting distributions as shape parameters tend to their extreme values are especially appealing. The family's simplest sub-class is closely related to a sub-class of the L_U family. Likelihood based inference is considered and applied in the analysis of heavy-tailed and skew data on fibre glass strengths. Comparisons are made throughout with two of the most popular existing competitors to this distribution: it scores very well relative to them on a number of tractability grounds.

Keywords Azzalini-type distributions · Heavy tails · Quantile-based measures · Student's t distribution · Skewness · Two-piece distributions

Mathematics Subject Classification (2000) 60E05 · 62F12 · 62F25

J.F. Rosco · A. Pewsey (✉)
Department of Mathematics, Escuela Politécnica, University of Extremadura, Avenida de la Universidad s/n, 10003 Cáceres, Spain
e-mail: apewsey@unex.es

J.F. Rosco
e-mail: juanfranrn@unex.es

M.C. Jones
Department of Mathematics & Statistics, The Open University, Walton Hall, Milton Keynes MK7 6AA, UK
e-mail: m.c.jones@open.ac.uk

1 Introduction

Student's t distribution is a fundamentally important distribution in Statistics, its discovery by W.S. Gosset and formalisation by R.A. Fisher representing one of the milestones in the development of the field. As well as its primary role in classical normal sampling theory, it now figures prominently as an empirical model for heavy-tailed data, particularly in finance (see Rachev et al. 2005). Variates simulated from it are frequently used in simulation experiments designed to assess robustness. Johnson et al. (1994b, Chap. 28) provide a summary of the extensive literature associated with the t distribution.

There has been much recent interest in the modelling of asymmetry together with heavy tails. Let us refer to any four-parameter family of distributions which contains the t distribution as a three-parameter (location, scale and degrees of freedom) symmetric subfamily as a skew t distribution. Early examples which can be seen in this light are the Pearson Type IV distribution (Johnson et al. 1994a, Chap. 12) and the noncentral t distribution (Johnson et al. 1994b, Chap. 31). Renewed interest has resulted in further, generally more tractable, skew t proposals being made in the literature. Amongst these figure: the two-piece t distributions of Hansen (1994) and Fernández and Steel (1998); the skew t distributions based on Azzalini's (1985) perturbations of the t distribution (Branco and Dey 2001; Azzalini and Capitanio 2003; Genton 2004; Ma and Genton 2004); the skew t distribution of Jones and Faddy (2003), obtained by transforming a beta random variable; and the skew t distribution arising from mean-variance mixing the normal distribution (Aas and Haff 2006). Probably the most successful of these are the Azzalini-type skew t distribution (in the form arising from scale mixing Azzalini's skew-normal distribution; see also Azzalini and Genton 2008) and the two-piece t distribution. The density of the Azzalini-type skew t distribution is

$$f_A(x) = 2 f_\nu(x) F_{\nu+1}\left(\alpha x \sqrt{\frac{\nu+1}{\nu+x^2}}\right), \qquad x, \alpha \in \mathbb{R}; \tag{1}$$

and that of the two-piece t distribution is

$$f_{\mathrm{TP}}(x) = f_\nu\left(\frac{x}{1+\gamma}\right) I(x<0) + f_\nu\left(\frac{x}{1-\gamma}\right) I(x \geq 0), \tag{2}$$

where I denotes the indicator function and $-1 < \gamma < 1$. Here and throughout, f_ν and F_ν denote the density and distribution functions, respectively, of the t distribution on ν degrees of freedom, denoted t_ν. These two distributions are compared briefly in the rejoinder to Jones (2008); an inferential advantage of two-piece distributions has been given by Jones and Anaya-Izquierdo (2011).

Also recently, Jones and Pewsey (2009) proposed the sinh-arcsinh transformation as a general means of generating classes of distributions containing symmetric, as well as asymmetric, cases with varying tailweights. In their formulation, a base random variable symmetric about 0, Z, is related to its skew-symmetric analogue, $X_{\varepsilon,\delta}$, via the (readily invertible) sinh-arcsinh transformation

$$Z = \sinh\left\{\delta \sinh^{-1}(X_{\varepsilon,\delta}) - \varepsilon\right\}, \tag{3}$$

where $\varepsilon \in \mathbb{R}$ is a skewness parameter and $\delta > 0$ controls tailweight. The distributions for $X_{\varepsilon,\delta}$ which result are skewed to the left (right) if $\varepsilon < 0$ ($\varepsilon > 0$). Their tailweights are lighter (heavier) than those of Z if $\delta > 1$ ($\delta < 1$). Obviously, $X_{0,1} = Z$.

In this paper, we propose an alternative skew t distribution generated by a restricted version of the sinh-arcsinh transformation that accommodates the fact that the t_ν distribution already has a parameter controlling tailweight, namely, its degrees of freedom $\nu > 0$. Letting T_ν denote a random variable from the t_ν distribution, we replace Z by T_ν and set $\delta = 1$ in (3) so as to define what we will refer to as a sinh-arcsinhed t random variable, $T_{\varepsilon,\nu}$, through the transformation

$$T_\nu = S_\varepsilon(T_{\varepsilon,\nu}) = \sinh\bigl(\sinh^{-1}(T_{\varepsilon,\nu}) - \varepsilon\bigr). \tag{4}$$

This is a skew t distribution because the symmetric t distributions correspond to $\varepsilon = 0$ and $\varepsilon \neq 0$ introduces and controls skewness as above. $T_{\varepsilon,\nu}$ has tails that, at one mathematical level, correspond to those of T_ν and are therefore at least as heavy as those of the limiting sinh-arcsinhed normal distribution with $\delta = 1$, obtained as $\nu \to \infty$. Note that $S_\varepsilon^{-1} = S_{-\varepsilon}$.

Numerous basic properties of the sinh-arcsinhed t family are studied in the seven subsections of Sect. 2. Likelihood based methods of inference for its parameters are described in Sect. 3 and applied, in Sect. 4, in the analysis of data on glass fibre strengths. Multivariate extensions are briefly considered in Sect. 5. The paper ends with Sect. 6 dedicated to conclusions. Comparisons with the Azzalini-type and two-piece skew t distributions will be made briefly wherever appropriate throughout the paper. The sinh-arcsinhed t distribution proves to be tractable and our comparisons will allow us to claim that it is as advantageous in many ways as the best of its competitors, and correspondingly preferable to others.

2 Properties of the family

2.1 Density, distribution and quantile functions, and simulation

Transformation of Student's t density via inversion of (4) results in $T_{\varepsilon,\nu}$ having density

$$f_{\varepsilon,\nu}(x) = \frac{C_\varepsilon(x)}{\sqrt{1+x^2}} f_\nu\bigl(S_\varepsilon(x)\bigr) = K_\nu \frac{C_\varepsilon(x)}{\sqrt{1+x^2}(1+\nu^{-1}S_\varepsilon^2(x))^{(\nu+1)/2}}, \quad x \in \mathbb{R}, \tag{5}$$

where $C_\varepsilon(x) = \cosh(\sinh^{-1}(x) - \varepsilon) = \sqrt{1 + S_\varepsilon^2(x)}$ and $K_\nu = \Gamma((\nu+1)/2)/(\sqrt{\nu\pi}\,\Gamma(\nu/2))$ is the normalising constant of the t_ν density, f_ν. The level of complexity of the density of $T_{\varepsilon,\nu}$ is therefore effectively that of t_ν. We note that $S_\varepsilon(x)$ can be represented in a variety of further ways, the most simple and compact of which is

$$S_\varepsilon(x) = x\cosh\varepsilon - \sqrt{x^2+1}\,\sinh\varepsilon. \tag{6}$$

Correspondingly, $C_\varepsilon(x) = \sqrt{x^2+1}\cosh\varepsilon - x\sinh\varepsilon$.

Changing the sign of ε leads to $f_{-\varepsilon,\nu}(x) = f_{\varepsilon,\nu}(-x)$. Examples of the shapes that can be assumed by (5) for non-negative ε are presented in Fig. 1. The densities in

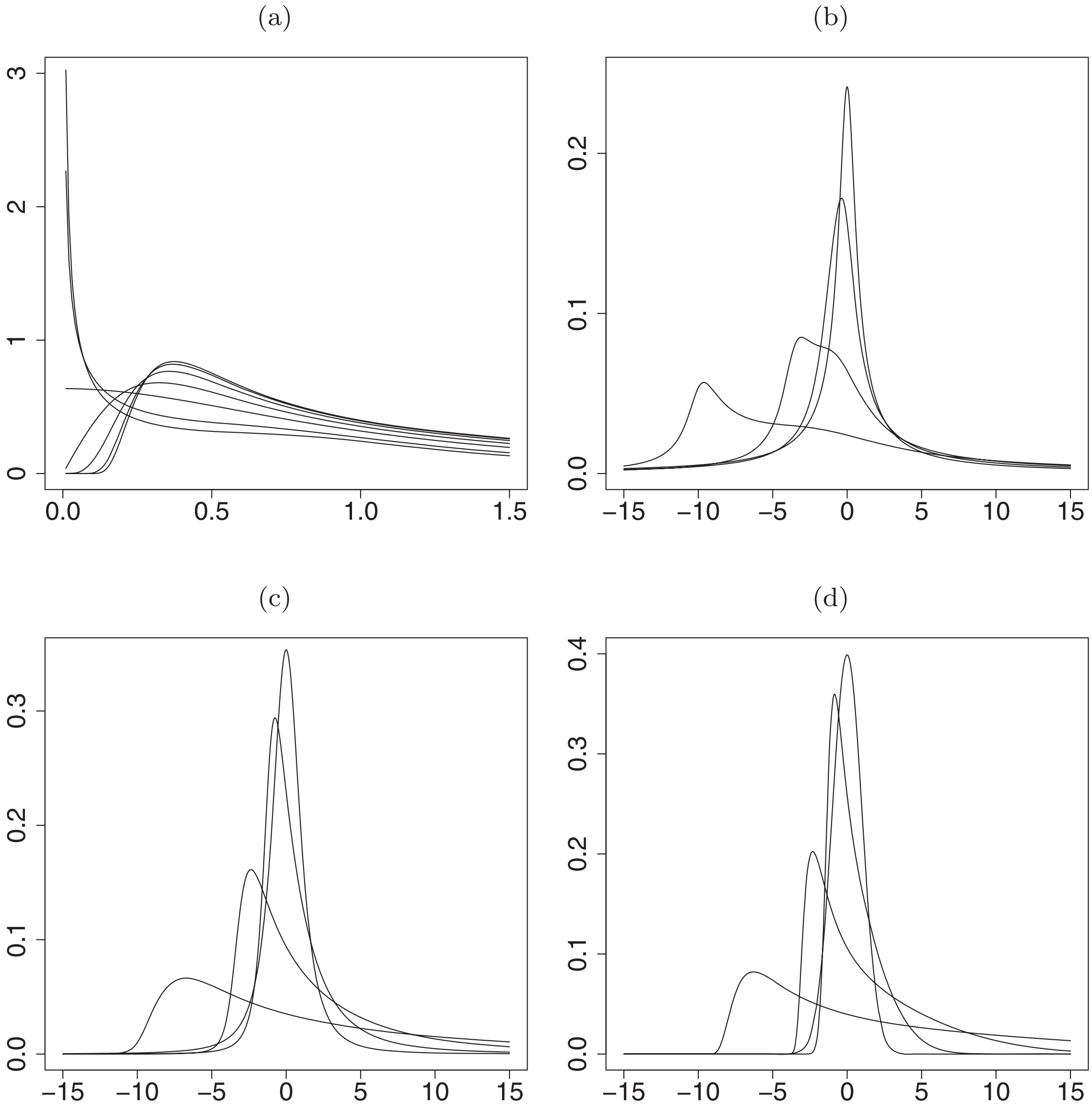


Fig. 1 Examples of density (5) with: (**a**) $\varepsilon = \infty$ and, from bottom to top at $x = 0.4$, $\nu = 0.35, 0.5, 1, 2, 5, 20, \infty$; (**b**) $\nu = 0.35$; (**c**) $\nu = 2$; (**d**) $\nu = \infty$. In (**a**), the densities have been rescaled to match formula (11); in (**b**)–(**d**) the four plotted densities have been shifted to have medians which are zero and correspond, from top to bottom at $x = 0$, to $\varepsilon = 0, 1, 2, 3$

panel (a) are those obtained as $\varepsilon \to \infty$. The other three panels depict symmetric as well as asymmetric densities for three different values of ν, the relevance of which will become apparent as we progress.

The distribution function of $T_{\varepsilon,\nu}$ is given in terms of the distribution function F_ν of t_ν by

$$F_{\varepsilon,\nu}(x) = F_\nu\big(S_\varepsilon(x)\big). \tag{7}$$

Hence the computation of values of $F_{\varepsilon,\nu}(x)$ is simple so long as routines are available to compute values of $F_\nu(x)$. $F_{\varepsilon,\nu}(x)$ can most easily be explicitly expressed as

$$F_{\varepsilon,\nu}(x) = F_B\left(\frac{1}{2}\left(1 + \frac{S_\varepsilon(x)}{\sqrt{\nu + S_\varepsilon^2(x)}}\right); \frac{\nu}{2}, \frac{\nu}{2}\right),$$

where F_B is the beta distribution function (or incomplete beta function ratio) given by $F_B(y; a, b) = B^{-1}(a, b) \int_0^y u^{a-1}(1-u)^{b-1}\,du$.

Inversion of (7) results in the quantile function of $T_{\varepsilon,\nu}$ being

$$Q_{\varepsilon,\nu}(u) = S_{-\varepsilon}\big(Q_\nu(u)\big), \quad 0 < u < 1, \tag{8}$$

where $Q_\nu(u)$ denotes the quantile function of T_ν. So both $F_{\varepsilon,\nu}$ and $Q_{\varepsilon,\nu}$ have essentially the same order of complexity as their counterparts, F_ν and Q_ν, of T_ν. This is also true of the two-piece t distribution but not of the Azzalini-type skew t distribution because the t distribution function appears in the skew t *density* function with consequent complications for its distribution function. A particularly elegant result for the new distribution is that the median of $T_{\varepsilon,\nu}$ is given by $\sinh(\varepsilon)$.

Similarly,

$$T_{\varepsilon,\nu} = S_{-\varepsilon}(T_\nu) = \sinh\big(\sinh^{-1}(T_\nu) + \varepsilon\big) = T_\nu \cosh\varepsilon + \sqrt{T_\nu^2 + 1}\,\sinh\varepsilon, \tag{9}$$

and thus the simulation of $T_{\varepsilon,\nu}$ random variates is essentially at the same level of complexity as that of simulation from the t_ν distribution. To date, the most efficient and simple algorithm available for simulating T_ν variates is Bailey's (1994) adaptation of the Box–Müller method. Both two-piece and Azzalini-type skew t random variates are obtainable from t random variates via simple formulae too, and ours holds only a *slight* edge: each Azzalini-type random variate is a simple function of the bivariate t random variate readily provided by the Box–Müller-type method, while the two-piece distribution requires a further uniform random variate in addition to the single t variate required here.

2.2 Tailweight and limiting distributions

As $|x| \to \infty$, $C_\varepsilon(x) \approx \exp\{-\mathrm{sgn}(x)\varepsilon\}|x|$ and $S_\varepsilon(x) \approx \mathrm{sgn}(x)C_\varepsilon(x)$ where $\mathrm{sgn}(x)$ denotes the sign of x. Using these results we obtain

$$f_{\varepsilon,\nu}(|x|) \approx \frac{\exp\{\mathrm{sgn}(x)\nu\varepsilon\}}{|x|^{\nu+1}}. \tag{10}$$

The dependence of the tails on x is the same as that of the t, Azzalini skew t and two-piece t distributions. Also, the sinh-arcsinhed t distribution shares with the latter two skew t distributions the property that the ratio of the constants in the limiting left and right tailweights depends on ϵ (as well as ν).

As $\nu \to \infty$, (5) tends to the density of a normal-based sinh-arcsinh distribution with $\delta = 1$ (see Jones and Pewsey 2009, Sect. 2.3), various examples of which are portrayed in panel (d) of Fig. 1. These are true "skew-normal" densities in the sense that both their tails are normal-like (a property in common with the two-piece t distribution but not Azzalini's, 1985, skew-normal distribution.)

As $f_{-\varepsilon,\nu}(x) = f_{\varepsilon,\nu}(-x)$, we focus on results for positive ε when considering limiting skewness distributions. A suitable standardisation of location and scale proves to be to consider the distribution of $Y = 2e^{-\varepsilon}T_{\varepsilon,\nu} - 1$. An easy calculation then shows that as $\varepsilon \to \infty$, $Y \to T_\nu + \sqrt{1+T_\nu^2}$, with density

$$f_{\infty,\nu}(y) = 2^\nu \nu^{(\nu+1)/2} K_\nu \frac{y^{\nu-1}(1+y^2)}{\{1+2(2\nu-1)y^2+y^4\}^{(\nu+1)/2}}, \quad y>0. \tag{11}$$

These are densities of the inverse of the identity-minus-reciprocal transformation $\sinh(\log(Y)) = (Y-(1/Y))/2$ discussed by Jones (2007). An immediate consequence, verifiable from density (11), is that the tails retain the behaviour of the t tails in the sense that $f_{\infty,\nu}(y) \approx y^{-(\alpha+1)}$ as $y \to \infty$ and adapt to $f_{\infty,\nu}(y) \approx y^{-(\alpha+1)}$ as $y \to 0$. Density (11) is plotted in panel (a) of Fig. 1 for a range of ν-values. There, the density for $\nu = \infty$ is that obtained as both ν and ε tend to ∞ and is the density of $Y = Z + \sqrt{1+Z^2}$ where Z is standard normal. These limiting densities display a wider range of shapes than do the limiting skewness distributions of Azzalini-type skew t and two-piece t densities which are both the half-t density. (Only when $\nu = 1$ does density (11) reduce to a half-t, actually half-Cauchy, density.)

2.3 Modality

The first derivative of (5) with respect to x is proportional to

$$-\frac{\{\sqrt{1+x^2}S_\varepsilon(x)(1+\nu S_\varepsilon^2(x)) + x\sqrt{1+S_\varepsilon^2(x)}(\nu+S_\varepsilon^2(x))\}}{(1+x^2)^{3/2}(\nu+S_\varepsilon^2(x))^{(\nu+3)/2}}.$$

The denominator of this quantity is positive and the numerator is zero for any x satisfying

$$-\frac{x}{\sqrt{1+x^2}} = \frac{S_\varepsilon(x)(1+\nu S_\varepsilon^2(x))}{\sqrt{1+S_\varepsilon^2(x)}(\nu+S_\varepsilon^2(x))}. \tag{12}$$

Clearly, the left side of (12) is a decreasing function in x (with horizontal asymptotes at 1 and -1). The right side of (12) is an increasing function of $y = S_\varepsilon(x)$ and hence of x if

$$y^4\left(2\nu^2+\nu-2\right) + y^2\left(3\nu^2-1\right) + \nu \geq 0 \quad \text{for all } y$$

and this is readily seen to be the case provided $\nu \geq (\sqrt{17}-1)/4 \simeq 0.78$ (with horizontal asymptotes at $-\nu$ and ν). In such cases, there is a unique point where equality is reached in (12) i.e. the density is unimodal.

Numerical investigation of the behaviour of expression (12) has led us to conclude that $f_{\varepsilon,\nu}$ is, in fact, always unimodal if $\nu \geq 0.35$. Of course, $f_{0,\nu}$, the density of the t_ν distribution, is always unimodal. However, for $\nu < 0.35$ there are values of ε for which the density is bimodal. As ν decreases, the range of ε-values corresponding to bimodal densities increases and, for positive ε, the right-hand mode increases exponentially. For $f_{\infty,\nu}$ of (11), bimodality emerges at around $\nu = 0.23$.

It must be admitted that we perceive the bimodality of (5) for $\nu < 0.35$ to be an unappealing property, primarily because we generally favour the use of finite mixtures as a means of modelling multimodality. We therefore propose that, in applications involving obviously skew data, ν should be restricted to being greater than

0.35. However, this is a constraint of barely any practical importance; in a closely related context, Jones and Faddy (2003 p. 163) state that "such distributions have *extremely* heavy tails and the whole business of directly modelling data containing many extreme outliers is not to be recommended". On the other hand, two-piece t distributions are clearly always unimodal with mode at 0; Azzalini-type skew t distributions appear to be unimodal for all ν too although this is harder to prove and, like (5), no explicit formula is available for the mode.

2.4 Moments

Property (10) implies that, as for t distributions, the rth moment of $T_{\varepsilon,\nu}$ exists provided $r < \nu$. In order to obtain the moments of $T_{\varepsilon,\nu}$ we make use of expression (9); it is found that

$$E\left[T_{\varepsilon,\nu}^{r}\right] = \sum_{m=0}^{\lfloor r/2 \rfloor} \binom{r}{2m} \cosh^{2m}(\varepsilon)\sinh^{r-2m}(\varepsilon) E\left\{T_\nu^{2m}\left(1+T_\nu^2\right)^{(r/2)-m}\right\},$$

where $\lfloor\cdot\rfloor$ denotes the floor function. Now,

$$\begin{aligned}
E\left[T_\nu^{2m}\left(1+T_\nu^2\right)^{(r/2)-m}\right] &= 2\nu^{(\nu+1)/2}K_\nu \int_0^\infty \frac{t^{2m}(1+t^2)^{(r/2)-m}}{(\nu+t^2)^{(\nu+1)/2}}\,\mathrm{d}t \\
&= \nu^{(\nu+1)/2}K_\nu \int_0^\infty \frac{w^{m-(1/2)}(1+w)^{(r/2)-m}}{(\nu+w)^{(\nu+1)/2}}\,\mathrm{d}w \\
&= \nu^{(\nu+1)/2}K_\nu B\big((\nu-r)/2, m+(1/2)\big) \\
&\quad \times {}_2F_1\big((\nu+1)/2, (\nu-r)/2; m+(\nu+1-r)/2; 1-\nu\big) \\
&= \frac{\nu^{\nu/2}\Gamma((\nu+1)/2)\Gamma((\nu-r)/2)\Gamma(m+(1/2))}{\sqrt{\pi}\,\Gamma(\nu/2)\Gamma(m+(\nu+1-r)/2)} \\
&\quad \times {}_2F_1\big((\nu+1)/2, (\nu-r)/2; m+(\nu+1-r)/2; 1-\nu\big),
\end{aligned}$$

the equalities arising from a slight rearrangement of the definition, the substitution $w=t^2$, (3.197.9) of Gradshteyn and Ryzhik (1994), and expansion of K_ν and the beta function, respectively. The formula from Gradshteyn and Ryzhik is

$$\int_0^\infty x^{\lambda-1}(1+x)^{-\mu+\nu}(x+\beta)^{-\nu}\,\mathrm{d}x = B(\mu-\lambda,\lambda)\,{}_2F_1(\nu,\mu-\lambda;\mu;1-\beta),$$

for $\mu > \lambda > 0$, and we set $\lambda = m+(1/2)$, $\mu = m+(\nu+1-r)/2$, $\nu = (\nu+1)/2$ and $\beta = \nu$. Denoting the ith moment of a T_ν random variable by N_i ($i < \nu$ and even), i.e.

$$N_i = \left\{\sqrt{\pi}\,\Gamma\left(\frac{\nu}{2}\right)\right\}^{-1} \nu^{i/2}\Gamma\left(\frac{i+1}{2}\right)\Gamma\left(\frac{\nu-i}{2}\right),$$

we note that $E[T_\nu^{2m}(1+T_\nu^2)^{(r/2)-m}]$ reduces to N_r when $m = r/2$ since the hypergeometric function then reduces to $\nu^{(r-\nu)/2}$ by (9.121.1) of Gradshteyn and Ryzhik (1994).

The first four moments of $T_{\varepsilon,\nu}$ simplify to:

$$\begin{aligned} E[T_{\varepsilon,\nu}] &= \sinh(\varepsilon)\frac{\nu^{\nu/2}\Gamma((\nu+1)/2)\Gamma((\nu-1)/2)}{\Gamma^2(\nu/2)} \\ &\quad \times {}_2F_1((\nu+1)/2, (\nu-1)/2; \nu/2; 1-\nu); \qquad (13) \\ E[T_{\varepsilon,\nu}^2] &= \sinh^2(\varepsilon) + \{1 + 2\sinh^2(\varepsilon)\}N_2; \\ E[T_{\varepsilon,\nu}^3] &= \sinh(\varepsilon)\frac{\nu^{\nu/2}\Gamma((\nu+1)/2)\Gamma((\nu-3)/2)}{2\Gamma^2(\nu/2)} \\ &\quad \times \left[3\cosh^2(\varepsilon){}_2F_1\big((\nu+1)/2, (\nu-3)/2; \nu/2; 1-\nu\big)\right. \\ &\quad \left. + (\nu-2)\sinh^2(\varepsilon){}_2F_1\big((\nu+1)/2, (\nu-3)/2; (\nu-2)/2; 1-\nu\big)\right]; \\ E[T_{\varepsilon,\nu}^4] &= \sinh^4(\varepsilon) + 2\sinh^2(\varepsilon)\big(3 + 4\sinh^2(\varepsilon)\big)N_2 \\ &\quad + \left[1 + 8\sinh^2(\varepsilon)\big(1 + \sinh^2(\varepsilon)\big)\right]N_4. \end{aligned}$$

The presence of the hypergeometric functions makes these formulae a little more complex than those for the Azzalini-type skew t and two-piece t distributions.

2.5 Skewness measures

As with all sinh-arcsinh distributions, the parameter ϵ in the sinh-arcsinhed t distribution is a bona fide skewness parameter in the classical convex ordering sense of van Zwet (1964); see Jones and Pewsey (2009, Sect. 2.2). (The same is true of the parameter intended to control skewness in the two-piece distributions, Klein and Fischer 2006, Sect. 5, but not in the Azzalini-type skew t distribution.) Many well-known summary measures of skewness are therefore monotone functions of ϵ—and take their maximum values for distribution (11)—including the two we discuss next (MacGillivray 1986).

First, we consider the classical moment based measure of skewness $\gamma_1 = \mu_3/\mu_2^{3/2}$, where $\mu_k = E[(X-\mu)^k]$. Panel (a) of Fig. 2 is a contour plot of $\gamma_1/(1+\gamma_1)$ for density (5) with $\varepsilon \geq 0$. Second, as γ_1 is unavailable when $\nu \leq 3$ (i.e. $\lambda \leq 0.75$), we consider the quantile based Bowley coefficient as an alternative measure of skewness. Panel (b) of Fig. 2 is a contour plot of

$$B_{\varepsilon,\nu} = \frac{Q_{\varepsilon,\nu}(3/4) - 2Q_{\varepsilon,\nu}(1/2) + Q_{\varepsilon,\nu}(1/4)}{Q_{\varepsilon,\nu}(3/4) - Q_{\varepsilon,\nu}(1/4)} = \frac{\tanh(\varepsilon)(\sqrt{1+Q_\nu^2(3/4)}-1)}{Q_\nu(3/4)},$$

again as a function of ρ and λ. (The denominator in the definition of Bowley's coefficient re-scales the numerator so that the coefficient lies in $[-1, 1]$.) Thus, essentially, $B_{\varepsilon,\nu}$ is at the same level of complexity as Q_ν, the quantile function of T_ν (similarly

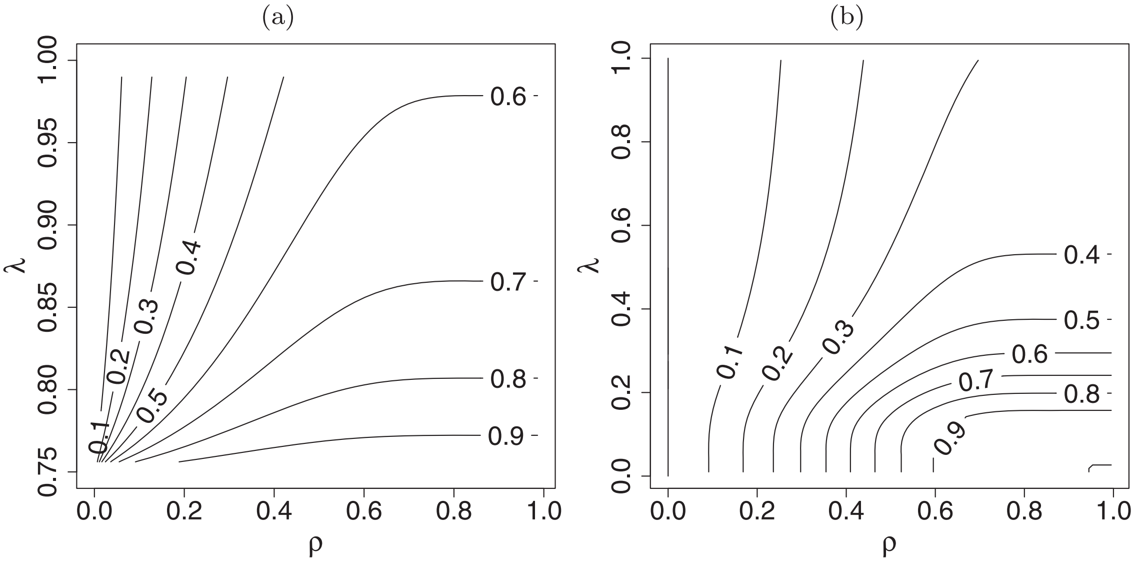


Fig. 2 Contour plots, for density (5) with $\varepsilon \geq 0$, of (**a**) $\gamma_1/(1+\gamma_1)$ and (**b**) Bowley's skewness coefficient as functions of $\rho = \varepsilon/(1+\varepsilon)$ and $\lambda = \nu/(1+\nu)$. In (**a**), γ_1 is undefined for $\nu \leq 3$. In (**b**), the vertical contour line corresponds to a value of zero for Bowley's coefficient

to the two-piece t distribution but unlike the Azzalini-type skew t distribution). As $\varepsilon \to \pm\infty$, $B_{\varepsilon,\nu} \to \pm(\sqrt{1+Q_\nu^2(3/4)}-1)/Q_\nu(3/4)$. Comparing the two panels of Fig. 2, it can be seen that, for a given value of ν, neither skewness measure increases appreciably as ε increases beyond about 2.3 ($\rho > 0.7$).

In a slightly different vein, we consider the behaviour of the density-based asymmetry function proposed by O'Hagan (1994, Sect. 2.6) and Critchley and Jones (2008) for unimodal distributions; see also Avérous et al. (1996) and Boshnakov (2007). Figure 3 shows several examples of their asymmetry function

$$\gamma^*(p) = \frac{x_R(p) - 2m + x_L(p)}{x_R(p) - x_L(p)}, \quad 0 < p < 1,$$

where m denotes the mode of the density and $x_R(p)$ and $x_L(p)$ both satisfy $f(x) = pf(m)$. We note the peculiar behaviour of the asymmetry function for $\varepsilon > 1$ and ν in the neighbourhood of the lower bound $\nu = 0.35$ ensuring unimodality. For larger values of ν, the asymmetry functions behave like examples shown in Critchley and Jones (2008, Sect. 3), increasing monotonically with increasing ε and decreasing p. For such ν-values, the marginal increase in its values as ε increases from 2 to 3 is consistent with the behaviour of the other two skewness measures considered previously.

2.6 Measures of kurtosis

We first consider the behaviour of the classical moment based measure of (excess) kurtosis $\gamma_2 = (\mu_4/\mu_2^4) - 3$. Figure 4 presents a contour plot for density (5) with $\varepsilon \geq 0$ of $\gamma_2/(1+\gamma_2)$ as a function of $\rho = \varepsilon/(1+\varepsilon)$ and $\lambda = \nu/(1+\nu)$. This only exists for $\nu > 4$ ($\lambda > 0.8$). As expected, γ_2 decreases with increasing ν. It also increases with

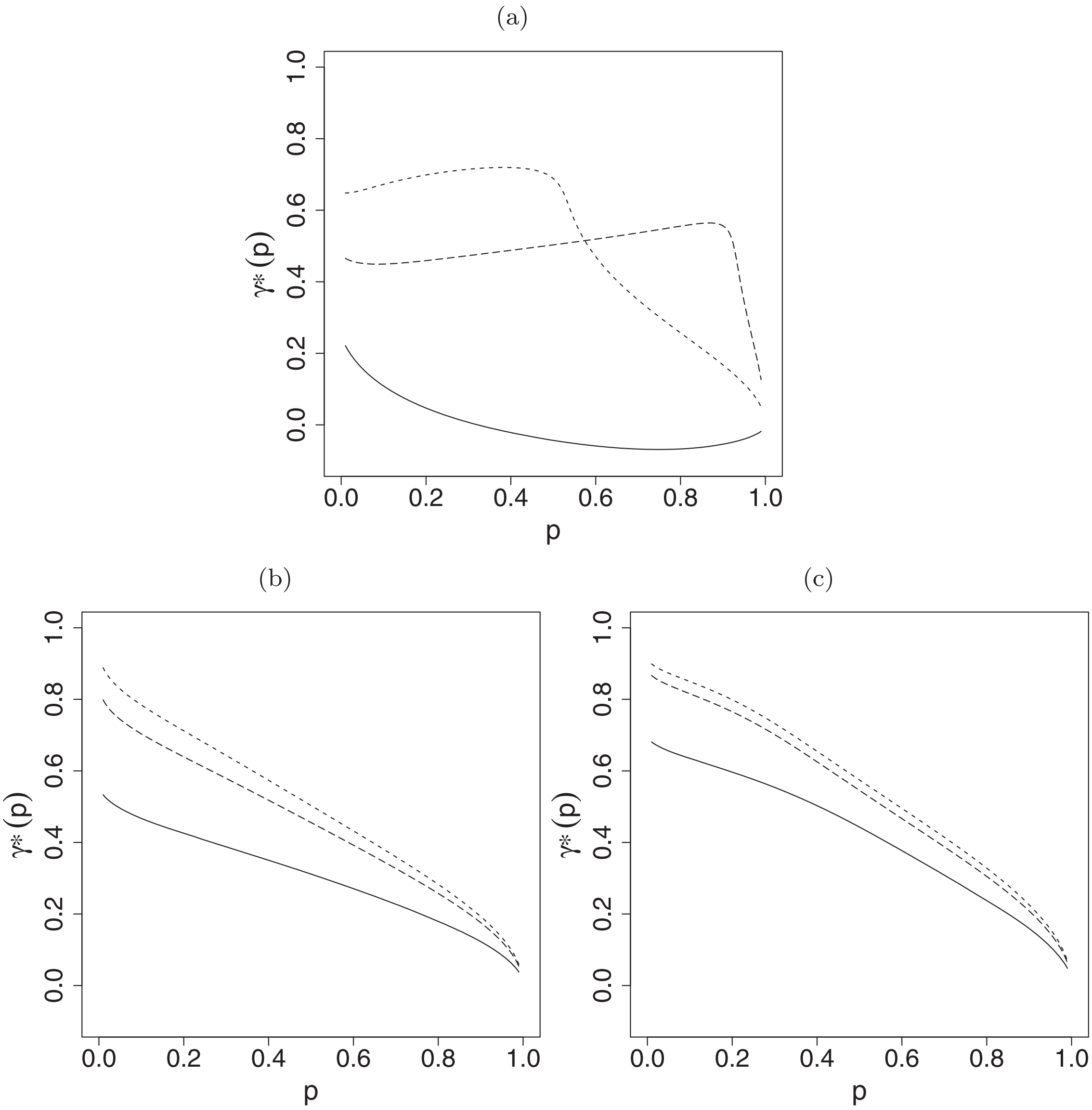


Fig. 3 The asymmetry function $\gamma^*(p)$ of density (5) for (**a**) $\nu = 0.35$, (**b**) $\nu = 2$ and (**c**) $\nu = \infty$ degrees of freedom and $\varepsilon = 1$ (*solid*), $\varepsilon = 2$ (*short dashed*), $\varepsilon = 3$ (*long dashed*)

increasing ε although, as was the case for the skewness measures portrayed in Fig. 2, for a given value of ν, γ_2 increases only very marginally as ε increases beyond 2.3 ($\rho > 0.7$).

Figure 5 displays the attainable region of skewness and excess kurtosis for the sinh-arcsinhed t family. Also included are the analogous regions for the Azzalini-type skew t and two-piece families. The former is almost entirely contained within the region for the sinh-arcsinhed t. The sinh-arcsinhed t (two-piece) family admits the wider skewness range for distributions with high (low) levels of kurtosis. Right-hand boundaries of the Azzalini-type and two-piece regions coincide where each corresponds to (their limiting) half t distributions. All three regions occupy substantial parts of the maximal region for unimodal distributions in which $\gamma_2 \geq \gamma_1^2 - 186/125$ (Klaassen et al. 2000).

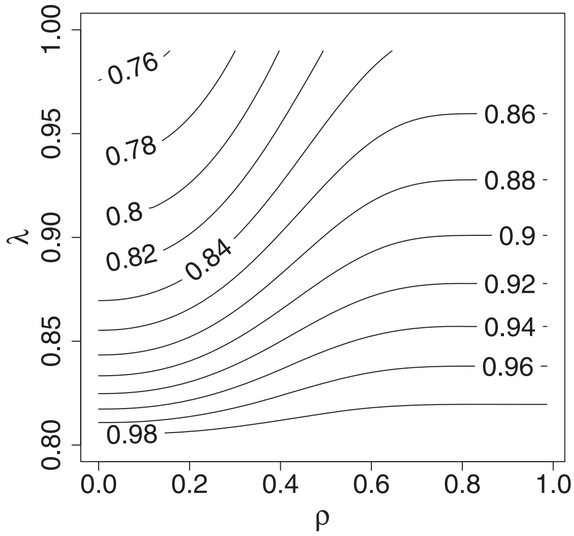


Fig. 4 Contour plot, for density (5) with $\varepsilon \geq 0$, of $\gamma_2/(1+\gamma_2)$ as a function of $\rho = \varepsilon/(1+\varepsilon)$ and $\lambda = \nu/(1+\nu)$

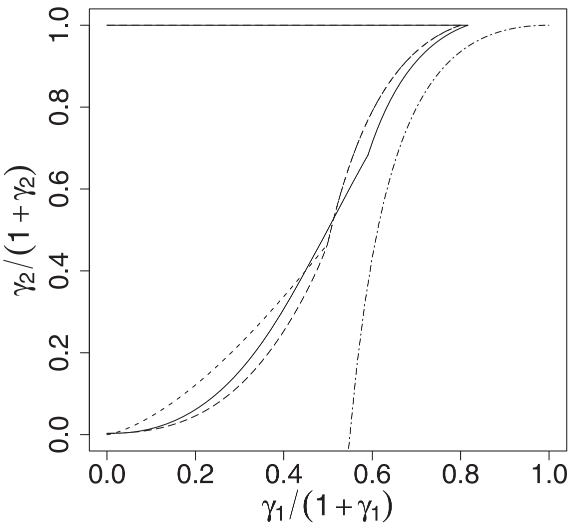


Fig. 5 Attainable regions of non-negative skewness and excess kurtosis, represented in terms of the measures $\gamma_1/(1+\gamma_1) \in [0, 1)$ and $\gamma_2/(1+\gamma_2) \in (0, 1)$, for the sinh-arcsinhed t (*solid*), Azzalini-type skew t (*short dashed*) and two-piece t (*long dashed*) distributions. Also included is Klaassen et al.'s (2000) lower bound for kurtosis as a function of skewness, $\gamma_2 \geq \gamma_1^2 - 186/125$, for unimodal distributions (*dot-dashed*)

As we show elsewhere (Jones et al. 2010), quantile based measures of kurtosis involving only (possibly scaled) differences between quantile function values of the form $Q(u) - Q(1-u)$ have, for all sinh-arcsinh distributions, the appealing property of being invariant to the value taken by the skewness parameter ε. Thus, for example, the quantile based kurtosis measure of Moors (1988),

$$M_{\varepsilon,\nu} = \frac{Q_{\varepsilon,\nu}(7/8) - Q_{\varepsilon,\nu}(5/8) + Q_{\varepsilon,\nu}(3/8) - Q_{\varepsilon,\nu}(1/8)}{Q_{\varepsilon,\nu}(6/8) - Q_{\varepsilon,\nu}(2/8)},$$

reduces to $M_\nu = (Q_\nu(7/8) - Q_\nu(5/8))/Q_\nu(6/8)$. A standardised version of it, $M_\nu^* = M_\nu/(1 + M_\nu)$, is portrayed in panel (a) of Fig. 6 and represents the Moors kurtosis signature of the t, and hence sinh-arcsinhed t, family. In particular, for $\nu = 0.35, 1, 2, \infty$, $M_\nu = 7.11, 2, 1.52, 1.23$, respectively. Note the minor decrease in M_ν as ν increases from 2 to ∞. Similarly, Hosking's (1990) L-kurtosis measure is

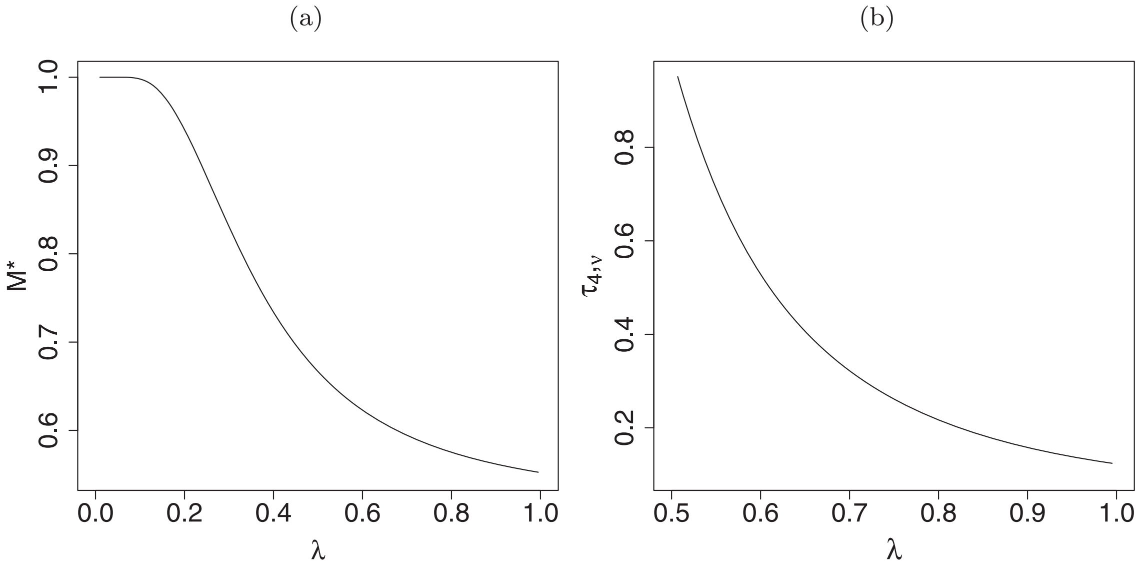


Fig. 6 Plots, for density (5) with arbitrary ε, of (**a**) $M_\nu^* = M_\nu/(1+M_\nu)$ and (**b**) $\tau_{4,\nu}$ as functions of $\lambda = \nu/(1+\nu)$. For $\nu < 1$, $\tau_{4,\nu}$ is undefined

also invariant to the value taken by the skewness parameter ε, being given by

$$\tau_{4,\nu} = \frac{\int_0^1 Q_\nu(u) P_3^*(u)\,\mathrm{d}u}{\int_0^1 Q_\nu(u) P_1^*(u)\,\mathrm{d}u}, \qquad \nu > 1,$$

where $P_1^*(u) = 2u - 1$ and $P_3^*(u) = 20u^3 - 30u^2 + 12u - 1$. The constraint, $\nu > 1$, is required to ensure the existence of the mean and hence $\tau_{4,\nu}$. As panel (b) of Fig. 6 attests, the shape of $\tau_{4,\nu}$ is very similar to that of M_ν^* for $\nu > 1$.

2.7 The simplest skew-symmetric sinh-arcsinhed t class and related topics

The simplest symmetric t distribution is the t_2 distribution (Jones 2002). Similarly, the simplest class of three-parameter skew-symmetric distributions within the sinh-arcsinhed t family is the sinh-arcsinhed t_2 class. Its density, distribution and quantile functions have the simple forms

$$\begin{aligned} f_{\varepsilon,2}(x) &= \frac{C_\varepsilon(x)}{\sqrt{1+x^2}(1+C_\varepsilon^2(x))^{3/2}}, \\ F_{\varepsilon,2}(x) &= \frac{1}{2}\left(1 + \frac{S_\varepsilon(x)}{\sqrt{2+S_\varepsilon^2(x)}}\right) \end{aligned} \tag{14}$$

and

$$Q_{\varepsilon,2}(u) = S_{-\varepsilon}\left(\frac{2u-1}{\sqrt{2u(1-u)}}\right).$$

Examples of $f_{\varepsilon,2}(x)$ for four different ε-values are presented in panel (c) of Fig. 1.

Using (13) together with (15.3.3) and (17.3.10) of Abramowitz and Stegun (1965), the mean of the distribution is given by

$$E[T_{\varepsilon,2}] = \pi \sinh(\varepsilon)_2F_1(3/2, 1/2; 1; -1) = \sinh(\varepsilon)\,\mathrm{E}(-1) \approx 1.9101 \sinh(\varepsilon),$$

where E denotes the complete elliptic integral of the second kind. Its variance and higher moments do not exist. However, quantile based measures of dispersion, skewness and kurtosis can easily be calculated due to the simple form of the quantile function. The values taken by Bowley's coefficient correspond to $\lambda = 2/3$ in panel (b) of Fig. 2 and range between 0 and approximately 0.35. Panel (b) of Fig. 3 portrays the asymmetry functions associated with it for three positive ε-values. As stated previously, the Moors kurtosis measure is 1.52 for this entire class, and Hosking's L-kurtosis measure is 0.375.

Even simpler but very similar distributions to (5) arise if the sinh-arcsinh transformation (4) is applied to the scaled random variable $Y_\nu = T_\nu/\sqrt{\nu}$ rather than to T_ν itself. Development of our ideas in this context has been eschewed because then the symmetric special cases of our model will be differently scaled t_ν distributions rather than the t_ν distributions themselves.

In fact, if the original sinh-arcsinh transformation (3) is applied to Y_2 then an existing four-parameter family of distributions is obtained, namely the L_U distributions introduced by Tadikamalla and Johnson (1982) (with their δ equal to twice our δ and their $\gamma = -2\epsilon$). L_U distributions are, from Tadikamalla and Johnson but in our notation, the distributions of $\mathcal{L} \equiv \sinh[\delta^{-1}\{(L/2) + \epsilon\}]$ where L is a standard logistic random variable. This matches with our construction $\mathcal{L} = \sinh[\delta^{-1}\{\sinh^{-1}(Y_2) + \epsilon\}]$ since Y_2 is distributed as $\sinh(L/2)$. Tadikamalla and Johnson (1982) provide the classical properties of the L_U family, whilst Jones (2004, Sect. 6.2) considers relations between the L_U and other distributions. The L_U distributions are not skew t distributions, however, in that none but the scaled t_2 distribution is a symmetric member and the symmetric density with lightest tails is the logistic; they are, in a strong sense, skew logistic distributions.

3 Likelihood based inference

In practice one will generally be interested in fitting the location-scale extension of (5). Introducing the location and scale parameters, $\xi \in \mathbb{R}$ and $\eta > 0$, respectively, in the usual way, the density becomes $\eta^{-1} f_{\varepsilon,\nu}((x-\xi)/\eta)$ and the log-likelihood for a random sample, $X_1, \dots, X_n$, drawn from it is given by

$$\begin{aligned}\ell(\xi, \eta, \varepsilon, \nu) &= n(\log K_\nu - \log \eta) \\ &\quad + \sum_{i=1}^{n}\left[\log C_\varepsilon(Y_i) - \frac{1}{2}\log\left(1 + Y_i^2\right) - \frac{(\nu+1)}{2}\log\left(1 + \frac{S_\varepsilon^2(Y_i)}{\nu}\right)\right],\end{aligned}$$

where $Y_i = (X_i - \xi)/\eta$.

As is generally the case, numerical methods of optimisation must be used to identify the maximum likelihood solution. We have successfully employed the simplex method of Nelder and Mead (1965) which is the default option, for instance, of R's optimisation routines. Particularly for small sized samples, a detailed exploration of the likelihood surface is required in order to ensure the global maximum is properly identified. When n is small, the profile log-likelihood functions for some of the parameters, particularly that for the location parameter ξ, can be far from smooth. Also, estimates on the upper ν boundary of the parameter space can arise fairly frequently when n is small. Whilst infinite estimates of ν might be interpreted as a problem from an accounting perspective (which can be resolved, for example, by transforming to $\lambda = \nu/(1+\nu) \in (0, 1)$), their interpretation is clear cut. They simply indicate that some member of the skewed normal sub-class is the most likely model from within the sinh-arcsinhed t family. It is important to note that these various issues are by no means exclusive to the optimisation problem considered here (see, e.g., Pewsey 2000).

The score equations and elements of the observed information matrix are given in the Appendix. The expected information matrix, particularly useful in theoretical work, can be calculated from the latter with the aid of numerical integration to compute expected values. Either matrix can be employed together with standard asymptotic normal theory as the basis of large-sample inferential techniques such as confidence set construction and hypothesis testing. Alternatively, profile likelihood methods are always available. All this appears to be on much of a par with the situation for the Azzalini-type skew t distribution; on the other hand, the high level of parameter orthogonality available with the two-piece distribution (Jones and Anaya-Izquierdo 2011) is not available here.

Extensive simulation based investigations confirmed the strong correlations between the maximum likelihood estimates (MLEs) of the location and skewness parameters, ξ and ε, and those of the scale and tailweight parameters, η and ν, predicted in a related context by Jones and Anaya-Izquierdo (2011). Use of the reparametrisation $\eta_\nu = \eta(1+1/\nu)$ suggested by them reduces the correlations between the MLEs for η_ν and ν. For some parameter combinations, relatively strong correlations were also observed between the MLEs of η and ε (as well as η_ν and ε).

4 Application

As an illustrative application we reanalyse data on the breaking strengths of $n = 63$ glass fibres of length 1.5 cm collected by workers at the U.K. National Physical Laboratory. Previous analyses of these data have appeared in Smith and Naylor (1987), the University of Padua technical report version of Azzalini and Capitanio (2003), Jones and Faddy (2003), Ma and Genton (2004) and Jones and Pewsey (2009), amongst others. The middle three of these publications present fits for three different four-parameter skew t models directly comparable with the sinh-arcsinhed t distribution considered here. Of those fits, that for the Azzalini-type skew t has the highest maximised log-likelihood value (-11.70).

Table 1 contains the MLEs of the parameters for the full sinh-arcsinhed t family as well as those for its sinh-arcsinhed t_2 and symmetric subclasses. Their standard errors, calculated from the observed information matrices for the three models, appear between brackets. Likelihood-ratio tests judge the fit for the full family to be

Table 1 Parameter estimates and, between brackets, their standard errors for the fits to the glass fibre strengths of, reading from right to left, the full location-scale extension of the sinh-arcsinhed t family with density (5), its sinh-arcsinhed t_2 subclass (SASt_2) with density (14) and the symmetric subclass of (5), with $\varepsilon = 0$. The maximised log-likelihood, $\ell_{\max}$, and p-value for the chi-squared goodness-of-fit test are included as fit diagnostics

	Model		
	Symmetric	SASt_2	Full
Parameter			
ξ	1.58 (0.03)	1.68 (0.05)	1.68 (0.05)
η	0.19 (0.04)	0.16 (0.03)	0.19 (0.04)
ε	0	−0.63 (0.26)	−0.58 (0.23)
ν	2.34 (1.11)	2	3.18 (1.85)
Diagnostic			
$\ell_{\max}$	−14.97	−11.83	−11.40
p-value	0.11	0.39	0.32

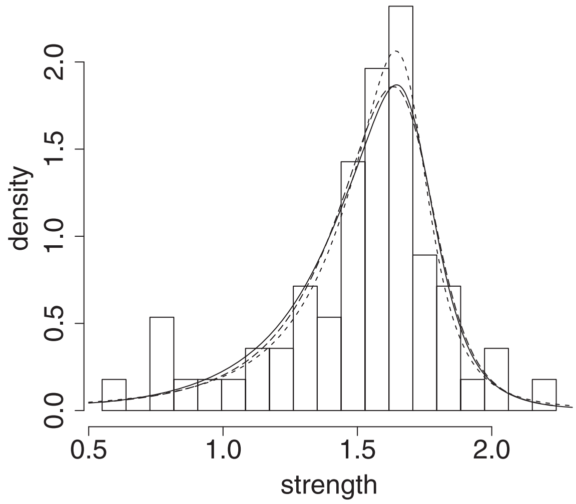


Fig. 7 Histogram of the glass fibre strength data together with fitted densities for the location-scale extensions of the sinh-arcsinhed t distribution with density (5) (*solid*), the Azzalini-type skew t distribution with density (1) (*long dashed*) and the sinh-arcsinhed t_2 distribution with density (14) (*dashed*)

superior to that for its symmetric subclass (p-value = 0.01) but not to that for its sinh-arcsinhed t_2 subclass (p-value = 0.35). Nevertheless, the p-values for the chi-squared goodness-of-fit test indicate that all three models provide adequate fits to the data. With a maximised log-likelihood value of -11.40, the fit for the full family is judged to be superior to that for any of the other three four-parameter skew t models referred to above. The MLE of the degrees of freedom in the Azzalini-type skew t fit is 2.73, close to the fitted values for ν for all three models represented in Table 1. A histogram of the data together with the densities for the sinh-arcsinhed t, Azzalini-type skew-t and the sinh-arcsinhed t_2 fits is presented in Fig. 7. The first two densities are very similar and suggest an underlying heavy-tailed and negatively skew distribution. The density for the sinh-arcsinhed t_2 fit is suggestive of a more peaked underlying distribution with slightly lighter flanks.

Individual nominally 95% confidence intervals for ξ, η, ε and ν, calculated from their profile log-likelihoods together with asymptotic chi-squared theory, are $(1.57, 1.77)$, $(0.11, 0.29)$, $(-1.11, -0.16)$ and $(1.27, 52.68)$, respectively. Their analogues calculated using the inverse of the observed information matrix are $(1.58, 1.78)$, $(0.10, 0.28)$, $(-1.02, -0.14)$ and $(0, 6.81)$. Clearly the intervals obtained using the two approaches are very similar, apart from those for the tailweight

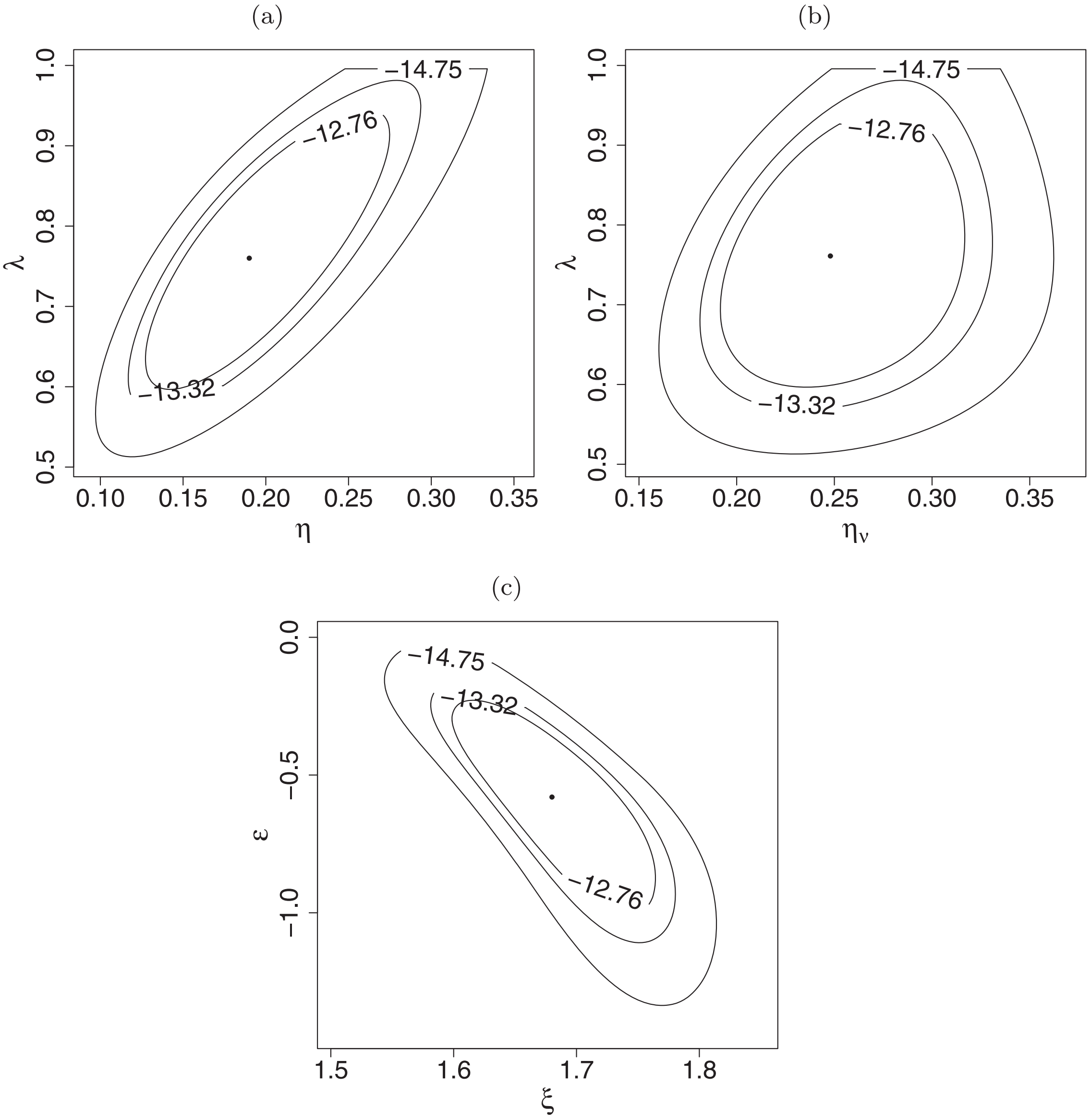


Fig. 8 Nominally 90%, 95%, and 99% profile log-likelihood based confidence regions for (**a**) (η, λ), (**b**) (η_ν, λ) and (**c**) (ξ, ε), where $\lambda = \nu/(1+\nu)$, for the glass fibre strength data. The contours of the three regions lie at $\chi_2^2(0.1)/2 = 2.305$, $\chi_2^2(0.05)/2 = 2.995$ and $\chi_2^2(0.01)/2 = 4.605$, respectively, below the log-likelihood value for the maximum likelihood solution identified by the filled circle. In panels (**a**) and (**b**) the upper boundary corresponds to the limiting value of $\nu = \infty$

parameter, ν; the profile log-likelihood based interval admits the possibility of an underlying distribution with close to normal tails. Figure 8 presents nominally 90%, 95% and 99% confidence regions for (η, λ), (η_ν, λ) and (ξ, ε), where $\lambda = \nu/(1+\nu)$. These regions were calculated from the joint profile log-likelihood functions of the different pairs of parameters together with standard asymptotic chi-squared theory. The shapes of the contours for (η, λ) and (ξ, ε), in panels (a) and (c), reflect the strong linear relationships that exist between their MLEs referred to previously in Sect. 3. As is evident from the contours in panel (b), reparametrisation of the scale parameter, η, to η_ν succeeds in breaking this strong dependence. The analogous confidence regions for the other four-parameter pairings are portrayed in Fig. 9, and confirm that the correlations between their respective MLEs are all low. Considered in combina-

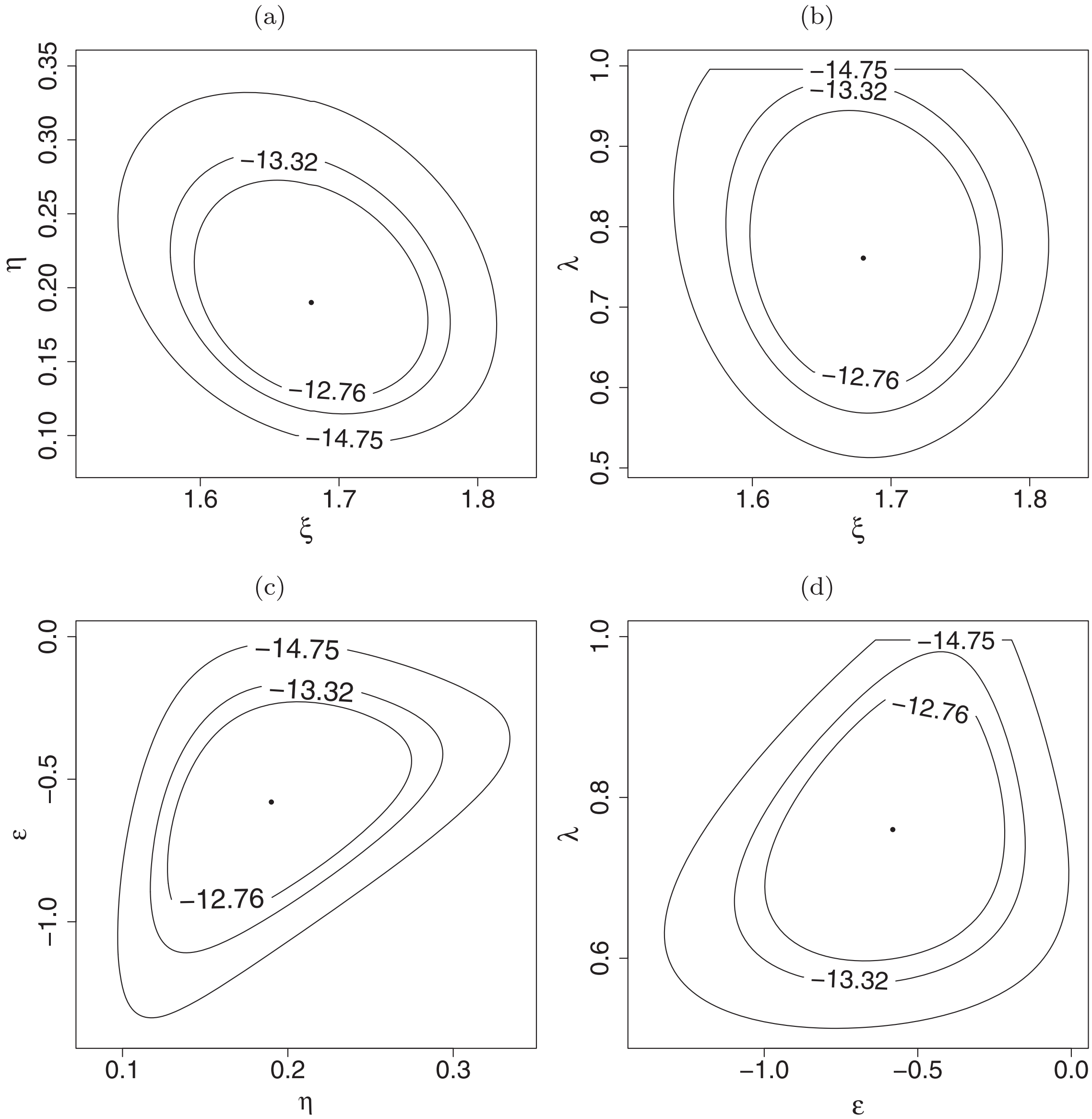


Fig. 9 Nominally 90%, 95%, and 99% profile log-likelihood based confidence regions for (**a**) (ξ, η), (**b**) (ξ, λ), (**c**) (η, ε) and (**d**) (ε, λ), where $\lambda = \nu/(1+\nu)$, for the glass fibre strength data. The contours of the three regions lie at $\chi_2^2(0.1)/2 = 2.305$, $\chi_2^2(0.05)/2 = 2.995$ and $\chi_2^2(0.01)/2 = 4.605$, respectively, below the log-likelihood value for the maximum likelihood solution identified by the filled circle. In panels (**b**) and (**d**) the upper boundary corresponds to the limiting value of $\nu = \infty$

tion, the confidence regions provide additional support for an underlying distribution that is unimodal and negatively skewed with heavier than normal tails.

As soon as one entertains the possibility of an underlying distribution that is asymmetric, the question as to which measure of central location is of real interest, or indeed meaningful, immediately arises. The mean is generally not a sensible measure of the "centre" of a skewed distribution, while the mode is not explicitly available in this case. The median is meaningful and available. For the fibre strength data, the maximum likelihood estimate of the median is $\hat{\xi} + \hat{\eta}\sinh(\hat{\varepsilon}) = 1.56$ and a 95% profile likelihood based confidence interval for it is given by the relatively tight interval (1.49, 1.63).

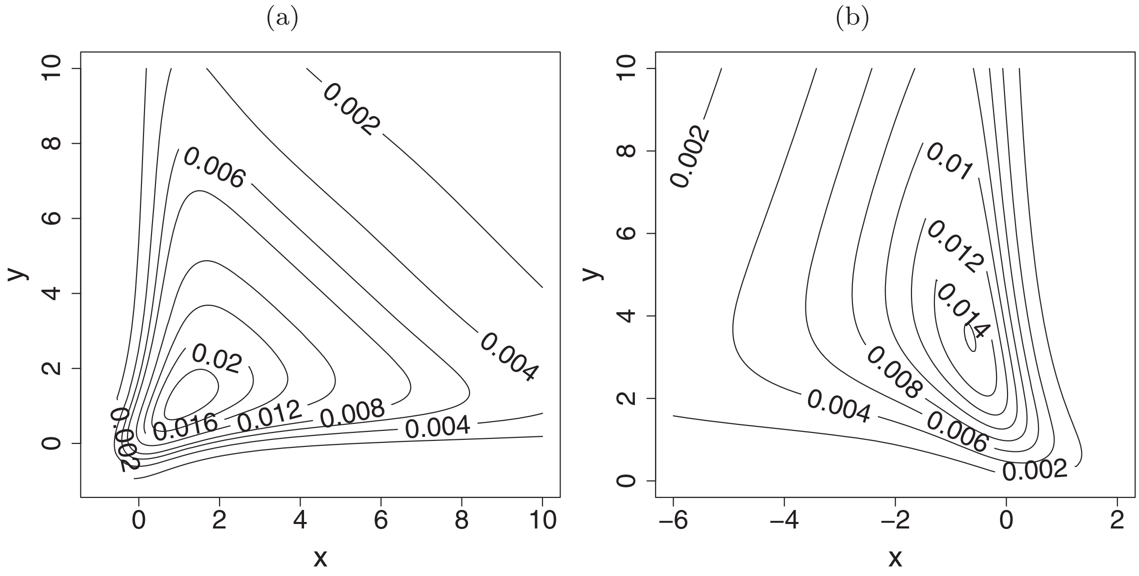


Fig. 10 Contour plots of $f_{\varepsilon_1,\varepsilon_2,\nu}(x, y)$ for $\nu = 2$ and (**a**) $\varepsilon_1 = \varepsilon_2 = 2$ and (**b**) $\varepsilon_1 = -1.5$ and $\varepsilon_2 = 3$

A referee is right to remind us that these data might be even better modelled by distributions outside the skew t class: examples with log-likelihoods of -10.00 and -10.02, respectively, can be found in Jones and Pewsey (2009) and Fischer and Vaughan (2010).

5 Possible multivariate extensions

There is a natural extension of the sinh-arcsinhed t distribution to the d-dimensional case and that is as the distribution of $T_{\varepsilon_j,\nu,j}$, $j = 1, \dots, d$, where

$$T_{\nu,j} = S_{\varepsilon_j}(T_{\varepsilon_j,\nu,j}) = \sinh\bigl(\sinh^{-1}(T_{\varepsilon_j,\nu,j}) - \varepsilon_j\bigr), \quad j = 1, \dots, d,$$

and $T_{\nu,j}$, $j = 1, \dots, d$ follow the multivariate t distribution with ν degrees of freedom (see e.g. Kotz and Nadarajah 2004). This distribution allows d different skewness parameters but only a single degrees of freedom parameter. When $d = 2$, its density is relatively simply

$$f_{\varepsilon_1,\varepsilon_2,\nu}(x, y) = \frac{1}{2\pi} \frac{C_{\epsilon_1}(x)}{\sqrt{1+x^2}} \frac{C_{\epsilon_2}(y)}{\sqrt{1+y^2}} \left(1 + \frac{S^2_{\epsilon_1}(x) + S^2_{\epsilon_2}(y)}{\nu}\right)^{-(\nu/2)-1},$$

$-\infty < x, y < \infty$. (This is the canonical case into which a location vector and non-identity scale matrix can be introduced in the usual way.) Two examples of $f_{\varepsilon_1,\varepsilon_2,\nu}$, for $\nu = 2$, are shown in Fig. 10; the values of ε_1 and ε_2 being equal in panel (a) and being different and with different signs in panel (b). The difference between ε_1 and ε_2 is what drives the asymmetry of the contours in panel (b). We hope to pursue this multivariate extension in further work, but for now a property that is easy to see is that

$$\mathrm{Cov}(T_{\varepsilon_1,\nu,1}, T_{\varepsilon_2,\nu,2}) = \sinh\epsilon_1 \sinh\epsilon_2 \mathrm{Cov}\Bigl(\sqrt{1+T^2_{\nu,1}}, \sqrt{1+T^2_{\nu,2}}\Bigr).$$

The sinh-arcsinhed t distribution shares with the Azzalini-type skew t distribution (and perhaps not the two-piece skew t distribution) a 'naturalness' of multivariate extension. It is not clear, however, that 'naturalness' necessarily equates to 'most usefulness' (for example, it is responsible for the single degree of freedom parameter above and in Azzalini-type multivariate distributions, unless a more general, and less accessible, form of multivariate t distribution is employed). Instead, generalised univariate families of distributions can always be extended to the multivariate case using some general scheme, the most obvious of which is to marginally transform copulas (see e.g. Nelsen 2006) using the distribution function (7). This allows d skewing and d degrees of freedom parameters. (And a t copula, Demarta and McNeil 2005, would be a natural choice; the distribution in the previous paragraph is a special case of this approach with equal degrees of freedom parameters.) Other general schemes which allow full flexibility are also available (see e.g. Ferreira and Steel 2007).

6 Conclusions

Student's t is a well-known and popular distribution. We consider the sinh-arcsinhed t family to provide an appealing extension of it, above all because many of its properties effectively have the same order of complexity as their Student t counterparts. This is particularly true of the quantile function and measures based upon it which, unlike classical moment based measures, are available for all degrees of freedom. The parameters of the sinh-arcsinhed t distribution have clear interpretations. Limiting distributions as shape parameters tend to their extreme values are especially appealing.

In practice, maximum likelihood inference reduces to an optimisation problem which is not devoid of complications, particularly when the sample size is small. However, such complications are inherent in all but the most basic of applications of the maximum likelihood method. Reparametrisation as suggested by Jones and Anaya-Izquierdo (2011) reduces the correlation that exists between the scale and tailweight parameters. In our analysis of the glass fibre strength data, in Sect. 4, we explored the fit of competing four-parameter skew t distributions, the sinh-arcsinhed t distribution being found to fit the data best. However, as explained in the same section, Jones and Pewsey (2009) report an even better fit to these data, obtained using their four-parameter sinh-arcsinhed normal distribution. Although both distributions model the heavy tails of the data equally well, the sinh-arcsinhed normal distribution models the data around the mode more closely. Thus, at least for the glass fibre strength data, there is evidence that the use of a standard normal Z in (3), with δ controlling tailweight, leads to greater flexibility in the modelling of the overall shape of the data distribution than the use of T_ν in (4), with ν controlling tailweight. Of course, in other applications involving heavy-tailed data sets one might expect this relation to be reversed.

Finally, we have compared our sinh-arcsinhed t distribution throughout with the Azzalini-type skew t distribution and the two-piece t distribution whose densities are given by (1) and (2), respectively. Of course, each has its advantages and disadvantages. Azzalini-type skew t distributions, in particular, have interesting genesis as marginal distributions of multivariate t distributions truncated on another variable, and as distributions of order statistics of multivariate t marginal variables. Where such modelling assumptions can be justified, Azzalini-type distributions are necessarily the distributions of choice. However, in purely empirical modelling terms, and with some emphasis on tractability, we judge the numbers of advantages of the sinh-arcsinhed t distribution to be on a par with the two-piece t distribution, both having more advantages than the Azzalini-type skew t distribution.

Acknowledgements Financial support for the research which led to the production of this paper was received by Rosco in the form of grant PRE08028 from the Junta de Extremadura, and by Pewsey in the form of grant MTM2010-16845 from the Spanish Ministry of Science and Education and grant PRI08A094 from the Junta de Extremadura. We are grateful to the referees for positive and helpful remarks.

Appendix

Score equations:

$$\frac{\partial \ell}{\partial \xi} = \frac{1}{\eta}\sum_{i=1}^{n}\left[-\frac{S_\varepsilon(Y_i)}{\sqrt{1+Y_i^2}C_\varepsilon(Y_i)} + \frac{Y_i}{1+Y_i^2} + \frac{\nu+1}{\nu}\frac{S_\varepsilon(Y_i)C_\varepsilon(Y_i)}{\sqrt{1+Y_i^2}(1+\nu^{-1}S_\varepsilon^2(Y_i))}\right] = 0,$$

$$\frac{\partial \ell}{\partial \eta} = -\frac{n}{\eta} + \frac{1}{\eta}\sum_{i=1}^{n}\left[-\frac{Y_i S_\varepsilon(Y_i)}{\sqrt{1+Y_i^2}C_\varepsilon(Y_i)} + \frac{Y_i^2}{1+Y_i^2}\right.$$
$$\left. + \frac{\nu+1}{\nu}\frac{Y_i S_\varepsilon(Y_i)C_\varepsilon(Y_i)}{\sqrt{1+Y_i^2}(1+\nu^{-1}S_\varepsilon^2(Y_i))}\right] = 0,$$

$$\frac{\partial \ell}{\partial \varepsilon} = \sum_{i=1}^{n}\left[-\frac{S_\varepsilon(Y_i)}{C_\varepsilon(Y_i)} + \frac{\nu+1}{\nu}\frac{S_\varepsilon(Y_i)C_\varepsilon(Y_i)}{1+\nu^{-1}S_\varepsilon^2(Y_i)}\right] = 0,$$

$$\frac{\partial \ell}{\partial \nu} = n(\log K_\nu)' + \sum_{i=1}^{n}\left[-\frac{1}{2}\log(1+\nu^{-1}S_\varepsilon^2(Y_i)) + \frac{\nu+1}{2\nu^2}\frac{S_\varepsilon^2(Y_i)}{1+\nu^{-1}S_\varepsilon^2(Y_i)}\right] = 0.$$

Elements of the observed information matrix:

$$-\frac{\partial^2 \ell}{\partial \xi^2} = \frac{1}{\eta^2}\sum_{i=1}^{n}\left[-\frac{1}{(1+Y_i^2)C_\varepsilon^2(Y_i)} + \frac{Y_i S_\varepsilon(Y_i)}{(1+Y_i^2)^{3/2}C_\varepsilon(Y_i)} + \frac{1-Y_i^2}{(1+Y_i^2)^2}\right.$$
$$+ \frac{\nu+1}{\nu}\frac{2S_\varepsilon^2(Y_i)+1}{(1+Y_i^2)(1+\nu^{-1}S_\varepsilon^2(Y_i))} - \frac{\nu+1}{\nu}\frac{Y_i S_\varepsilon(Y_i)C_\varepsilon(Y_i)}{(1+Y_i^2)^{3/2}(1+\nu^{-1}S_\varepsilon^2(Y_i))}$$
$$\left. - \frac{2(\nu+1)}{\nu^2}\frac{S_\varepsilon^2(Y_i)C_\varepsilon^2(Y_i)}{(1+Y_i^2)(1+\nu^{-1}S_\varepsilon^2(Y_i))^2}\right],$$

$$-\frac{\partial^2 \ell}{\partial \xi \partial \eta} = \frac{1}{\eta^2}\sum_{i=1}^{n}\left[-\frac{Y_i}{(1+Y_i^2)C_\varepsilon^2(Y_i)} + \frac{Y_i^2 S_\varepsilon(Y_i)}{(1+Y_i^2)^{3/2}C_\varepsilon(Y_i)} + \frac{Y_i(1-Y_i^2)}{(1+Y_i^2)^2}\right.$$

$$+\frac{\nu+1}{\nu}\frac{Y_i(2S_\varepsilon^2(Y_i)+1)}{(1+Y_i^2)(1+\nu^{-1}S_\varepsilon^2(Y_i))}-\frac{\nu+1}{\nu}\frac{Y_i^2S_\varepsilon(Y_i)C_\varepsilon(Y_i)}{(1+Y_i^2)^{3/2}(1+\nu^{-1}S_\varepsilon^2(Y_i))}$$

$$-\frac{2(\nu+1)}{\nu^2}\frac{Y_iS_\varepsilon^2(Y_i)C_\varepsilon^2(Y_i)}{(1+Y_i^2)(1+\nu^{-1}S_\varepsilon^2(Y_i))^2}\Bigg],$$

$$-\frac{\partial^2\ell}{\partial\xi\partial\varepsilon}=\frac{1}{\eta}\sum_{i=1}^n\Bigg[-\frac{1}{\sqrt{1+Y_i^2}C_\varepsilon^2(Y_i)}+\frac{\nu+1}{\nu}\frac{2S_\varepsilon^2(Y_i)+1}{\sqrt{1+Y_i^2}(1+\nu^{-1}S_\varepsilon^2(Y_i))}$$

$$-\frac{2(\nu+1)}{\nu^2}\frac{S_\varepsilon^2(Y_i)C_\varepsilon^2(Y_i)}{\sqrt{1+Y_i^2}(1+\nu^{-1}S_\varepsilon^2(Y_i))^2}\Bigg],$$

$$-\frac{\partial^2\ell}{\partial\xi\partial\nu}=\frac{1}{\eta\nu^2}\sum_{i=1}^n\Bigg[\frac{S_\varepsilon(Y_i)C_\varepsilon(Y_i)}{\sqrt{1+Y_i^2}(1+\nu^{-1}S_\varepsilon^2(Y_i))}-\frac{\nu+1}{\nu}\frac{S_\varepsilon^3(Y_i)C_\varepsilon(Y_i)}{\sqrt{1+Y_i^2}(1+\nu^{-1}S_\varepsilon^2(Y_i))^2}\Bigg],$$

$$-\frac{\partial^2\ell}{\partial\eta^2}=\frac{n}{\eta^2}+\frac{1}{\eta^2}\sum_{i=1}^n\Bigg[-\frac{Y_i^2}{(1+Y_i^2)C_\varepsilon^2(Y_i)}+\frac{Y_i^3S_\varepsilon(Y_i)}{(1+Y_i^2)^{3/2}C_\varepsilon(Y_i)}+\frac{Y_i^2(1-Y_i^2)}{(1+Y_i^2)^2}$$

$$+\frac{\nu+1}{\nu}\frac{Y_i^2(2S_\varepsilon^2(Y_i)+1)}{(1+Y^2)(1+\nu^{-1}S_\varepsilon^2(Y_i))}-\frac{\nu+1}{\nu}\frac{Y_i^3S_\varepsilon(Y_i)C_\varepsilon(Y_i)}{(1+Y_i^2)^{3/2}(1+\nu^{-1}S_\varepsilon^2(Y_i))}$$

$$-\frac{2(\nu+1)}{\nu^2}\frac{Y_i^2S_\varepsilon^2(Y_i)C_\varepsilon^2(Y_i)}{(1+Y_i^2)(1+\nu^{-1}S_\varepsilon^2(Y_i))^2}\Bigg],$$

$$-\frac{\partial^2\ell}{\partial\eta\partial\varepsilon}=\frac{1}{\eta}\sum_{i=1}^n\Bigg[-\frac{Y_i}{\sqrt{1+Y_i^2}C_\varepsilon^2(Y_i)}+\frac{\nu+1}{\nu}\frac{Y_i(2S_\varepsilon^2(Y_i)+1)}{\sqrt{1+Y_i^2}(1+\nu^{-1}S_\varepsilon^2(Y_i))}$$

$$-\frac{2(\nu+1)}{\nu^2}\frac{Y_iS_\varepsilon^2(Y_i)C_\varepsilon^2(Y_i)}{\sqrt{1+Y_i^2}(1+\nu^{-1}S_\varepsilon^2(Y_i))^2}\Bigg],$$

$$-\frac{\partial^2\ell}{\partial\eta\partial\nu}=\frac{1}{\eta\nu^2}\sum_{i=1}^n\Bigg[\frac{Y_iS_\varepsilon(Y_i)C_\varepsilon(Y_i)}{\sqrt{1+Y_i^2}(1+\nu^{-1}S_\varepsilon^2(Y_i))}-\frac{\nu+1}{\nu}\frac{Y_iS_\varepsilon^3(Y_i)C_\varepsilon(Y_i)}{\sqrt{1+Y_i^2}(1+\nu^{-1}S_\varepsilon^2(Y_i))^2}\Bigg],$$

$$-\frac{\partial^2\ell}{\partial\varepsilon^2}=\sum_{i=1}^n\Bigg[-\frac{1}{C_\varepsilon^2(Y_i)}+\frac{\nu+1}{\nu}\frac{2S_\varepsilon^2(Y_i)+1}{1+\nu^{-1}S_\varepsilon^2(Y_i)}-\frac{2(\nu+1)}{\nu^2}\frac{S_\varepsilon^2(Y_i)C_\varepsilon^2(Y_i)}{(1+\nu^{-1}S_\varepsilon^2(Y_i))^2}\Bigg],$$

$$-\frac{\partial^2\ell}{\partial\varepsilon\partial\nu}=\frac{1}{\nu^2}\sum_{i=1}^n\Bigg[\frac{S_\varepsilon(Y_i)C_\varepsilon(Y_i)}{1+\nu^{-1}S_\varepsilon^2(Y_i)}-\frac{\nu+1}{\nu}\frac{S_\varepsilon^3(Y_i)C_\varepsilon(Y_i)}{(1+\nu^{-1}S_\varepsilon^2(Y_i))^2}\Bigg],$$

$$-\frac{\partial^2\ell}{\partial\nu^2}=-n(\log K_\nu)''+\frac{1}{\nu^3}\sum_{i=1}^n\Bigg[\frac{S_\varepsilon^2(Y_i)}{(1+\nu^{-1}S_\varepsilon^2(Y_i))}-\frac{\nu+1}{2\nu}\frac{S_\varepsilon^4(Y_i)}{(1+\nu^{-1}S_\varepsilon^2(Y_i))^2}\Bigg],$$

where

$$(\log K_\nu)' = \frac{\mathrm{d}\log K_\nu}{\mathrm{d}\nu} = \frac{1}{2}\left\{\psi\left(\frac{\nu+1}{2}\right) - \psi\left(\frac{\nu}{2}\right) - \frac{1}{\nu}\right\},$$

$$(\log K_\nu)'' = \frac{\mathrm{d}^2\log K_\nu}{\mathrm{d}\nu^2} = \frac{1}{4}\left\{\psi^{(1)}\left(\frac{\nu+1}{2}\right) - \psi^{(1)}\left(\frac{\nu}{2}\right) + \frac{2}{\nu^2}\right\}$$

and $\psi(\cdot)$ and $\psi^{(1)}(\cdot)$ denote the digamma and trigamma functions, respectively.

References

Aas K, Haff IH (2006) The generalized hyperbolic skew student's t-distribution. J Financ Econ 4:275–309

Abramowitz M, Stegun IA (eds) (1965) Handbook of mathematical functions. Dover, New York

Avérous J, Fougères AL, Meste M (1996) Tailweight with respect to the mode for unimodal distributions. Stat Probab Lett 28:367–373

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

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 65:367–389

Azzalini A, Genton MG (2008) Robust likelihood methods based on the skew-t and related distributions. Int Stat Rev 76:106–129

Bailey RW (1994) Polar generation of random variates with the t-distribution. Math Comput 62:779–781

Boshnakov GN (2007) Some measures for asymmetry of distributions. Stat Probab Lett 77:1111–1116

Branco MD, Dey DK (2001) A general class of multivariate skew-elliptical distributions. J Multivar Anal 79:99–113

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

Demarta S, McNeil AJ (2005) The t copula and related copulas. Int Stat Rev 73:111–129

Fernández C, Steel MFJ (1998) On Bayesian modelling of fat tails and skewness. J Am Stat Assoc 93:359–371

Ferreira JTAS, Steel MFJ (2007) A new class of skewed multivariate distributions with applications to regression analysis. Stat Sin 17:505–529

Fischer M, Vaughan D (2010) The beta-hyperbolic secant distribution. Austrian J Stat 39:245–258

Genton MG (ed) (2004) Skew-elliptical distributions and their applications; a journey beyond normality. Chapman and Hall/CRC, Boca Raton

Gradshteyn IS, Ryzhik IM (1994) Tables of integrals, series, and products. Academic Press, San Diego

Hansen B (1994) Autoregressive conditional density estimation. Int Econ Rev 35:705–730

Hosking JRM (1990) L-moments: analysis and estimation of distributions using linear combinations of order statistics. J R Stat Soc, Ser B 52:105–124

Johnson NL, Kotz S, Balakrishnan N (1994a) Continuous univariate distributions, 2nd edn, vol 1. Wiley, New York

Johnson NL, Kotz S, Balakrishnan N (1994b) Continuous univariate distributions, 2nd edn, vol 2. Wiley, New York

Jones MC (2002) Student's simplest distribution. J R Stat Soc, Ser D 51:41–49

Jones MC (2004) Families of distributions arising from distributions of order statistics (with discussion). Test 13:1–43

Jones MC (2007) Connecting distributions with power tails on the real line, the half line and the interval. Int Stat Rev 75:58–69

Jones MC (2008) The t family and their close and distant relations (with discussion). J Korean Stat Soc 37:293–311

Jones MC, Anaya-Izquierdo K (2011) On parameter orthogonality in symmetric and skew models. J Stat Plan Inference 141:758–770

Jones MC, Faddy MJ (2003) A skew extension of the t distribution, with applications. J R Stat Soc, Ser B 65:159–174

Jones MC, Pewsey A (2009) Sinh-arcsinh distributions. Biometrika 96:761–780

Jones MC, Rosco JF, Pewsey A (2010) Skewness-invariant measures of kurtosis. Under consideration

Klaassen CAJ, Mokveld PJ, van Es B (2000) Squared skewness minus kurtosis bounded by 186/125 for unimodal distributions. Stat Probab Lett 50:131–135

Klein I, Fischer M (2006) Skewness by splitting the scale parameter. Commun Stat, Theory Methods 35:1159–1171

Kotz S, Nadarajah S (2004) Multivariate *t* distributions and their applications. Cambridge University Press, Cambridge

Ma Y, Genton MG (2004) Flexible class of skew symmetric distributions. Scand J Stat 31:459–468

MacGillivray HL (1986) Skewness and asymmetry: measures and orderings. Ann Stat 14:994–1011

Moors JJA (1988) A quantile alternative for kurtosis. J R Stat Soc, Ser D 37:25–32

Nelder JA, Mead R (1965) A simplex method for function minimization. Comput J 7:308–313

Nelsen RB (2006) An introduction to copulas, 2nd edn. Springer, New York

O'Hagan A (1994) Kendall's advanced theory of statistics, vol 2B: Bayesian inference. Arnold, London

Pewsey A (2000) Problems of inference for Azzalini's skew-normal distribution. J Appl Stat 27:859–870

Rachev ST, Menn C, Fabozzi FJ (2005) Fat-tailed and skewed asset return distributions: implications for risk management, portfolio selection, and option pricing. Wiley, New York

Smith RL, Naylor JC (1987) A comparison of maximum likelihood and Bayesian estimators for the three-parameter Weibull distribution. Appl Stat 36:358–369

Tadikamalla PR, Johnson NL (1982) Systems of frequency curves generated by transformations of logistic variables. Biometrika 69:461–465

van Zwet WR (1964) Convex transformations of random variables. Mathematisch Centrum, Amsterdam

Skewness-Invariant Measures of Kurtosis

M. C. Jones, J. F. Rosco, and Arthur Pewsey

Measures of kurtosis, when applied to asymmetric distributions, are typically much affected by the asymmetry which muddies their already murky interpretation yet further. Certain kurtosis measures, however, when applied to certain wide families of skew-symmetric distributions display the attractive property of skewness-invariance. In this article, we concentrate mainly on quantile-based measures of kurtosis and their interaction with skewness-inducing transformations, identifying classes of transformations that leave kurtosis measures invariant. Further miscellaneous aspects of skewness-invariant kurtosis measures are briefly considered, these not being quantile-based and/or not involving transformations. While our treatment is as unified as we are able to make it, we do not claim anything like a complete characterization of skewness-invariant kurtosis measures but hope that our results will stimulate further research into the issue.

KEY WORDS: Asymmetry; Johnson distributions; Quantile measures; Sinh function; Sinh–arcsinh transformation.

1. INTRODUCTION

For many, the kurtosis of a random variable, X, is nothing other than its value of the moment-based measure $\alpha_4 = \mu_4/\sigma^4$, where $\mu_k = E[(X-\mu)^k]$, $\sigma^2 = E[(X-\mu)^2]$, and $\mu = E(X)$ (Thiele 1889; Pearson 1905); or perhaps its version calibrated relative to normality, the excess kurtosis $\alpha_4 - 3$. Despite its popularity as a measure of distributional shape, α_4 is well-known to have some important drawbacks. A first is that α_4 is undefined if any of the first four moments of X do not exist, which makes it inapplicable to heavy-tailed distributions.

Later 20th-century articles on kurtosis centered on two questions:

- What does kurtosis mean?
- What are alternative ways of measuring kurtosis?

M. C. Jones is Professor of Statistical Science, Department of Mathematics & Statistics, The Open University, Walton Hall, Milton Keynes MK7 6AA, U.K. (E-mail: *m.c.jones@open.ac.uk*). J. F. Rosco is Postgraduate Student (E-mail: *juanfranrn@unex.es*) and Arthur Pewsey is Titular de Universidad (E-mail: *apewsey@unex.es*), Department of Mathematics, Escuela Politécnica, University of Extremadura, 10003 Cáceres, Spain. The authors are very grateful to the Associate Editor and two referees for their helpful suggestions to improve readability and to Nick Cox for numerous detailed comments. Financial support for this research was received by Rosco in the form of grant PRE08028 from the Junta de Extremadura, and by Pewsey in the form of grant MTM2010-16845 from the Spanish Ministry of Science and Education and grant PRI08A094 from the Junta de Extremadura.

It is not our purpose to get embroiled in the first question, but we note that if a one-word alternative is desired that one word would probably be "peakedness." However, it is clear that peakedness only makes sense relative to the weights of the tails of the distributions, with highly peaked/leptokurtic distributions often having heavy tails . . . whose kurtosis cannot be measured by α_4! An appealing alternative description of kurtosis, at once refined and vague, is given by Balanda and MacGillivray (1988) as the "location- and scale-free movement of probability mass from the shoulders of a distribution into its center and tails." (We note in passing that much of the discussion of the meaning of kurtosis, since 1970, has taken place in the pages of *The American Statistician*. Balanda and MacGillivray (1988) provide an excellent review of that debate and of the multifarious alternative approaches to measuring kurtosis that had been proposed to that date.) Modern alternative measures of kurtosis mostly center attention on the quantile function of the distribution of X, $Q(u), 0 < u < 1$, and this article will be no exception. The quantile function has the great advantage that it always exists. Quantile-based kurtosis measures typically involve ratios of quantities based on differences of quantile values, ratios being necessary to afford scale-invariance; some examples will be given in Section 2.

A second drawback of α_4 is implicit in the well-known relationship $\alpha_4 \geq \alpha_3^2 + 1$, where $\alpha_3 = \mu_3/\sigma^3$ (Pearson 1916): higher skewness (as measured by α_3) inevitably leads to higher kurtosis (as measured by α_4). This and the further complications asymmetry might pose in interpretation of kurtosis have led to the vast majority of the kurtosis literature dealing only with symmetric distributions. Notable exceptions are Balanda and McGillivray (1990) and Blest (2003) both of which we shall return to later, the latter only briefly. But why should kurtosis be a concept reserved for symmetric distributions only? Ideas of "peakedness," "weights of tails," "movement of probability mass," and "shoulders of a distribution" remain as meaningful in the presence of asymmetry as for symmetric distributions. Indeed, while it is accepted that kurtosis measures should be invariant to location and scale, it actually seems entirely reasonable to us to ask a third question:

- Can kurtosis be measured in a manner invariant to skewness?

It is this question, of kurtosis measures invariant to skewness, that we address in this article. We point up a number of examples of such measures, presented in as unified a manner as we are able. Our main results, presented in Section 2, concern quantile-based measures of kurtosis and their interaction with skewness-inducing transformations. The latter afford a very general consideration of skew distributions. Further miscellaneous aspects of skewness-invariant kurtosis measures are briefly considered in Section 3. We do not pretend to provide a complete characterization of kurtosis measures invariant to skewness, but hope that the current article stimulates further research and results on the issue.

 DOI: 10.1198/tast.2011.10194

2. TRANFORMATIONS AND SKEWNESS-INVARIANT QUANTILE-BASED MEASURES OF KURTOSIS

2.1 Quantile-Based Measures of Kurtosis

Given their potential nonexistence, we eschew kurtosis measures involving conventional moments and focus instead on quantile-based measures. It seems sensible, and often done, to consider measures that are ratios of linear combinations of differences between quantiles of the form $Q(u) - Q(1-u)$, $0 < u < 1$. That is, we entertain measures of the general form

$$\frac{\sum_{i=1}^{n_1} c_i \{Q(u_i) - Q(1-u_i)\}}{\sum_{j=1}^{n_2} d_j \{Q(u_j) - Q(1-u_j)\}} \tag{1}$$

for positive integers n_1 and n_2 and constants $c_i : i = 1, \ldots, n_1$ and $d_j : j = 1, \ldots, n_2$. Typically, $n_2 = 1$ and, as already mentioned, the denominator is present in order to enforce scale invariance.

In case the reader is concerned about the arbitrariness of this definition, here is an apparently different argument leading to the same place. Balanda and MacGillivray (1990, section 2.3) approach the issue by "taking the kurtosis properties of a random variable X to be those of its symmetrized version" $Z_{\rm BM}$, say. The specific definition of $Z_{\rm BM}$ they assume is

$$Z_{\rm BM} = X - F^{-1}(1 - F(X)),$$

where $F = Q^{-1}$ is the distribution function associated with X. In quantile terms this means that the quantile function $Q_{\rm BM}$ associated with the distribution of $Z_{\rm BM}$ is given by

$$Q_{\rm BM}(u) = Q(u) - Q\big(1 - F(Q(u))\big) = Q(u) - Q(1-u).$$

That is, given that Balanda and MacGillivray, like us, are primarily concerned with quantile-based measures of kurtosis, they are really arguing that kurtosis measures should be based on (ratios of linear combinations of) $Q_{\rm BM}$, and hence on (the same functions of) $Q(u) - Q(1-u)$.

Particular cases of measures having the general form (1) include:

- The p indexed measure

$$t(p) = \frac{Q\big(\frac{1}{2}+p\big) - Q\big(\frac{1}{2}-p\big)}{Q\big(\frac{3}{4}\big) - Q\big(\frac{1}{4}\big)},$$

  $0 < p < \frac{1}{2}$. This was referred to by Balanda and MacGillivray (1988), where earlier references to special cases can be found.
- Moors's (1988) octile-based measure

$$M = \frac{(O_7 - O_5) + (O_3 - O_1)}{O_6 - O_2} = \frac{(O_7 - O_1) - (O_5 - O_3)}{O_6 - O_2}$$
$$= \frac{Q\big(\frac{7}{8}\big) - Q\big(\frac{5}{8}\big) + Q\big(\frac{3}{8}\big) - Q\big(\frac{1}{8}\big)}{Q\big(\frac{3}{4}\big) - Q\big(\frac{1}{4}\big)},$$

  where $O_i = Q(\frac{i}{8})$, $i = 1, \ldots, 7$, is the ith octile. (A similar measure based on more extreme quantiles had earlier been suggested by Inman (1952).)
- The quintile-based measure

$$J = \frac{A_4 - A_3 - 2(A_3 - A_2) + A_2 - A_1}{A_4 - A_1}$$
$$= \frac{(A_4 - A_1) - 3(A_3 - A_2)}{A_4 - A_1}$$
$$= \frac{Q\big(\frac{4}{5}\big) - 3Q\big(\frac{3}{5}\big) + 3Q\big(\frac{2}{5}\big) - Q\big(\frac{1}{5}\big)}{Q\big(\frac{4}{5}\big) - Q\big(\frac{1}{5}\big)},$$

  where $A_i = Q(\frac{i}{5})$, $i = 1, \ldots, 4$, denotes the ith quintile. This seems to us to be a natural extension to kurtosis (by third differencing) of the Bowley skewness measure (which takes second differences; Bowley 1902)

$$\frac{Q\big(\frac{3}{4}\big) - 2Q\big(\frac{1}{2}\big) + Q\big(\frac{1}{4}\big)}{Q\big(\frac{3}{4}\big) - Q\big(\frac{1}{4}\big)},$$

  although one could use an alternative quantile-difference scale measure in the denominator of our measure.

  More generally, but possibly less transparently, general form (1) might be extended to the integral form

$$\frac{\sum_{i=1}^{n_1} \int_0^1 c_i(u)\{Q(u) - Q(1-u)\}\,du}{\sum_{j=1}^{n_2} \int_0^1 d_j(v)\{Q(v) - Q(1-v)\}\,dv}. \tag{2}$$

Now, for any function c,

$$\int_0^1 c(u)\{Q(u) - Q(1-u)\}\,du = \int_0^1 a(v)Q(v)\,dv,$$

where $a(u) = c(u) - c(1-u)$ is odd about $\frac{1}{2}$. Therefore, (2) has the alternative representation

$$\frac{\sum_{i=1}^{n_1} \int_0^1 a_i(u)Q(u)\,du}{\sum_{j=1}^{n_2} \int_0^1 b_j(v)Q(v)\,dv},$$

where the a_i's and b_j's are odd functions.

- The prime example of (2) is the L-moment based kurtosis measure of Hosking (1990, 1992):

$$\tau_4 = \frac{\int_0^1 P_3^*(u)Q(u)\,du}{\int_0^1 P_1^*(v)Q(v)\,dv},$$

  where $P_1^*(u) = 2u - 1$ and $P_3^*(u) = 20u^3 - 30u^2 + 12u - 1$ are the first and third shifted Legendre polynomials.

However, integration imposes conditions on the existence of such measures, τ_4 existing only if the mean of the distribution does.

2.2 Invariance Under Transformations: Requirements

We now focus on the behavior of the forms of kurtosis measures in Section 2.1 for families of distributions obtained using a popular approach based on the transformation of a symmetric random variable. Let Z denote a continuous random variable from a distribution which is symmetric about 0, and define the random variable X_λ via the transformation $X_\lambda = T_\lambda(Z)$ where T_λ is a one-to-one function taken, without loss of generality, to be increasing. For the moment, $\lambda \in \mathbb{R}$ is a general shape parameter which will specialize to a skewness parameter in Section 2.3. The quantile function, Q_λ, of the distribution of X_λ is given immediately by $Q_\lambda(u) = T_\lambda(Q_Z(u))$, $0 < u < 1$, where Q_Z is the quantile function of the distribution of Z. As Z is assumed to be symmetric about 0, if $Q_Z(u) = z$ then $Q_Z(1-u) = -z$. Thus,

$$\begin{aligned} Q_\lambda(u) - Q_\lambda(1-u) &= T_\lambda(Q_Z(u)) - T_\lambda(Q_Z(1-u)) \\ &= T_\lambda(z) - T_\lambda(-z). \end{aligned}$$

The aim now is to identify those transformations which leave any measure of the form (1) or (2) invariant to the value of λ. This happens if $T_\lambda(z) - T_\lambda(-z)$ factorizes as

$$T_\lambda(z) - T_\lambda(-z) = t_1(\lambda)t_2(z), \tag{3}$$

where $t_1(\lambda)$ is a nonzero function of the parameter λ (not involving z), and $t_2(z)$ is a function of z (not involving λ). Without loss of generality, $t_1(\lambda)$ can be subsumed into T_λ by division, so that (3) reduces to

$$T_\lambda(z) - T_\lambda(-z) = t_2(z).$$

By considering this equation when $z = 0$ and when z is replaced by $-z$, we get that $t_2(0) = 0$ and that t_2 is an odd function of z.

To make more structured progress, it is now useful to concentrate on a further reduced version of (3). To this end, make the natural further requirement that *the family of transformations T_λ includes the identity transformation as a special case*. Without further loss of generality, we can take this case to correspond to $\lambda = 0$. Then, $t_2(z) = T_0(z) - T_0(-z) = 2z$ and so the final requirement is that

$$T_\lambda(z) - T_\lambda(-z) = 2z \tag{4}$$

for monotone increasing transformations T_λ.

Of course, because $T_\lambda(z) - T_\lambda(-z) = Q_\lambda(u) - Q_\lambda(1-u)$ and $z = Q_Z(u)$, (4) is nothing other than a reexpression and restandardization of Balanda and MacGillivray's (1990) requirement, referred to in Section 2.1, that, in the current notation,

$$Q_\lambda(u) - Q_\lambda(1-u) = 2Q_Z(u). \tag{5}$$

2.3 Invariance Under Skewness-Inducing Transformations: Solutions

Considerable progress can be made on identifying T_λ satisfying (4). Requirement (4) also arose in Jones (2011), but in a different context, and what follows is an improved version of the relevant (small) part of the work of that article.

Differentiating (4) gives

$$T'_\lambda(z) + T'_\lambda(-z) = 2 \tag{6}$$

and then

$$T''_\lambda(z) - T''_\lambda(-z) = 0. \tag{7}$$

Since T_λ is increasing, Equation (6) shows that, additionally, $0 \le T'_\lambda(z) \le 2$.

First, only partially successful, attempts at solutions of (4) are the antiderivatives of twice the distribution or survival functions of distributions symmetric about zero. Write $k(z) = \ell(z^2)$ and $K(z)$ for the density and distribution functions of such a symmetric distribution, and L for the antiderivative of ℓ. Taking $T'_\lambda(z) = 2K(\lambda z)$, $\lambda \in \mathbb{R}$, covers distribution ($\lambda > 0$), survival ($\lambda < 0$) and unit ($\lambda = 0$) functions, satisfies (6), and leads to candidate solutions of (4) of the form

$$T_\lambda(z) = 2\big[zK(\lambda z) - \{L(\lambda^2 z^2)/2\lambda\}\big].$$

However, only in some special cases corresponding to distribution functions, K, with extremely heavy tails does this T_λ produce transformations with range the whole of $\mathbb{R}$ (Jones 2011, section 3).

This last observation inspires the following more appealing alternative solution which leads directly to a well-defined skewing interpretation for the corresponding transformation. Continue to consider monotone T'_λ, nonmonotonicity, though possible, being detrimental to the retention of unimodality of the transformed density. Even when monotone, T'_λ does not have to be twice a distribution (resp. survival) function. Instead of starting from 0 (resp. 2) for $x \to -\infty$ and ending at 2 (resp. 0) for $x \to \infty$, T'_λ can start from c (resp. $2-c$) and end at $2-c$ (resp. c) where $0 < c < 1$. Equivalently, and covering both cases, take $T'_\lambda(z) = 1 - \lambda + 2\lambda K(z)$, $-1 < \lambda < 1$, and, correspondingly,

$$T_\lambda(z) = z\{1 - \lambda + 2\lambda K(z)\} - \lambda L(z^2), \qquad -1 < \lambda < 1. \tag{8}$$

Clearly $T_0(z) = z$.

Moreover, monotonicity of K implies convexity of T_λ in (8) when $\lambda > 0$ and concavity when $\lambda < 0$. This corresponds precisely to the parameter λ acting as a skewness parameter in the classical sense of van Zwet (1964), positive λ introducing positive skewness, negative λ negative skewness.

To summarize, quantile-based measures of kurtosis of distributions of $X_\lambda = T_\lambda(Z)$ where Z is from a symmetric distribution and T_λ is given by (8) are invariant to the value of λ, which is a true skewness parameter in the sense of van Zwet.

2.4 Illustrative Families of Distributions

Examples of symmetric distributions yield numerous examples of formula (8). For instance, if $K(z) = \Phi(z)$, the standard normal distribution function, $L(z^2) = -2\phi(z)$, where ϕ is the standard normal density function, and then (8) becomes

$$T_\lambda(z) = z\{1 - \lambda + 2\lambda\Phi(z)\} + 2\lambda\phi(z), \qquad -1 < \lambda < 1.$$

Examples like this abound but have the disadvantage, for some practical purposes, of the transformation not being explicitly invertible. Explicitly invertible transformations of the form (8) are far fewer. The degenerate case of $K(z) = I(z \ge 0)$ leads to the aesthetically unattractive two-piece distributions with a discontinuity in density at their join. Other distributions on finite support introduce discontinuities in derivative too.

2.4.1 *The Sinh–Arcsinh Transformation*

Returning to K defined on the whole of $\mathbb{R}$, we are aware of just one, special, distribution that leads to invertible T_λ of the form (8). Set K to be the distribution function of the following scaled t distribution on two degrees of freedom (e.g., Jones

2002), a distribution that is ubiquitously useful in distribution theory:

$$K(z) = \frac{1}{2}\left(1 + \frac{z}{\sqrt{1+z^2}}\right), \qquad L(z^2) = -\frac{1}{\sqrt{1+z^2}}.$$

Then,

$$\begin{aligned} T_\lambda(z) &= z + \lambda\sqrt{1+z^2} = \frac{\sinh(\epsilon + \sinh^{-1}(z))}{\cosh\epsilon}, \\ T_\lambda^{-1}(y) &= \sinh(-\epsilon + \sinh^{-1}(y\cosh\epsilon)), \end{aligned} \tag{9}$$

where $\epsilon = \tanh^{-1}\lambda$ is also necessarily a skewness parameter in the sense of van Zwet (1964).

Jones and Pewsey (2009) introduced skew-symmetric families of distributions generated by sinh–arcsinh transformation of the random variable Z which follows a density g which is symmetric about zero. The sinh–arcsinh transformation is given by $S_{\epsilon,\delta}(x) = \sinh(\delta\sinh^{-1}(x) - \epsilon)$, $\epsilon \in \mathbb{R}$, $\delta > 0$, where ϵ is a skewness parameter and δ controls tailweight. Jones and Pewsey explored the distributions of $X_{\epsilon,\delta} \sim f_{\epsilon,\delta;g}$, related to $Z \sim g$ by

$$\begin{aligned} X_{\epsilon,\delta} &= S^{-1}_{\epsilon,\delta}(Z) = S_{-\epsilon/\delta,1/\delta}(Z) \\ &= \sinh\big(\delta^{-1}(\epsilon + \sinh^{-1}(Z))\big). \end{aligned} \tag{10}$$

The corresponding densities are

$$f_{\epsilon,\delta;g}(x) = \{1+x^2\}^{-1/2}\delta C_{\epsilon,\delta}(x) g\{S_{\epsilon,\delta}(x)\}, \tag{11}$$

where $C_{\epsilon,\delta}(x) = \cosh\{\delta\sinh^{-1}(x) - \epsilon\} = \{1 + S^2_{\epsilon,\delta}(x)\}^{1/2}$. Specific examples include the sinh–arcsinhed normal distribution of Jones and Pewsey (2009) when $g = \phi$ and, when $\delta = 1$, the sinh–arcsinhed t distributions of Rosco, Jones, and Pewsey (2011) when g is the density of the t distribution on $\nu > 0$ degrees of freedom.

Transformation (9) can now be seen to be a rescaled version of the $\delta = 1$ special case, $S_{-\epsilon,1}$, of the sinh–arcsinh transformation given at (10). It was in the course of study of the sinh–arcsinhed t distributions that the skewness-invariant nature of certain kurtosis measures was noticed by the second author and initially disbelieved by the third!

Quantile-based kurtosis measures are, in fact, skewness-invariant for the general sinh–arcsinh transformation given at (10) as well as for the special scaled case given at (9). This is because

$$\begin{aligned} &S_{-\epsilon/\delta,1/\delta}(z) - S_{-\epsilon/\delta,1/\delta}(-z) \\ &\quad = 2\cosh(\delta^{-1}\epsilon)\sinh(\delta^{-1}\sinh^{-1}(z)), \end{aligned}$$

satisfying (3) for each fixed δ. The general transformation did not arise directly from the considerations of Section 2.3 because the identity transformation requires $\delta = 1$. The general transformation satisfies (4) if we think of it as T_λ (with an irrelevantly reparametrized skewness parameter) applied to the (symmetric) distribution of $P(Z) = S_{0,\delta}(Z) = \sinh(\delta(\sinh^{-1}(Z))$ rather than to Z itself.

The skewness-invariant quantile-based kurtosis measures are now illustrated for the particular case in which Z is standard normal so that $X_\lambda = X_{\epsilon,\delta}$ is sinh–arcsinhed normal and so follows density (11) with $g = \phi$ (Jones and Pewsey 2009). Write $z(p) = \Phi^{-1}(\frac{1}{2} + p)$ for $0 < p < 1/2$. The formulas for the kurtosis measures defined in Section 2.1, namely, the p indexed measure $t(p)$, Moors's M, the quintile-based J and the L-moment ratio τ_4, are in this case:

$$\begin{aligned} t(p) &= \frac{\sinh(\delta^{-1}\sinh^{-1}(z(p)))}{\sinh(\delta^{-1}\sinh^{-1}(z(1/4)))}, \\ M &= \frac{\sinh(\delta^{-1}\sinh^{-1}(z(3/8))) - \sinh(\delta^{-1}\sinh^{-1}(z(1/8)))}{\sinh(\delta^{-1}\sinh^{-1}(z(1/4)))}, \\ J &= 1 - 3\frac{\sinh(\delta^{-1}\sinh^{-1}(z(1/10)))}{\sinh(\delta^{-1}\sinh^{-1}(z(3/10)))}, \\ \tau_4 &= \frac{\int_0^{1/2}\sinh(\delta^{-1}\sinh^{-1}(\Phi^{-1}(u)))P_3^*(u)\,du}{\int_0^{1/2}\sinh(\delta^{-1}\sinh^{-1}(\Phi^{-1}(v)))P_1^*(v)\,dv}. \end{aligned}$$

Crucially, none of these functions varies with ϵ. Indeed, they equal the kurtosis measures of the symmetric distributions with densities

$$\begin{aligned} f_{0,\delta;\phi}(x) &= \{2\pi(1+x^2)\}^{-1/2} \\ &\quad \times \delta C_{0,\delta}(x)\exp\{-S^2_{0,\delta}(x)/2\}, \end{aligned} \tag{12}$$

$\delta > 0$, examples of which are displayed in figure 1(c) of Jones and Pewsey (2009).

The four measures are depicted in the panels of Figure 1. The horizontal variable in each frame of Figure 1 is $\beta_1 = \delta/(1+\delta)$, a transformation of $\delta > 0$ to $0 < \beta_1 < 1$ made for plotting convenience. The analogous transformation $M^* = M/(1+M)$ is made for Moors's measure in Figure 1(b). Figure 1(a) differs from Figure 1(b)–(d) in being a contour plot rather than a single function plot, because two variables affect the value of $t(p)$: p and β_1.

Figure 2 portrays the same measures for the sinh–arcsinhed t distribution with density (11) with $\delta = 1$ and g the density of the t distribution on ν degrees of freedom (Rosco, Jones, and Pewsey 2011, whose figure 6 displays a portion of the same information), that is, the quantile-based kurtosis measures associated with the t distribution itself. Note that the horizontal variable in all frames of Figure 2 is $\beta_2 = \nu/(1+\nu)$ rather than $\beta_1 = \delta/(1+\delta)$ as in Figure 1. Figure 2 can nonetheless be compared directly with Figure 1. These comparisons help in clarifying the effects on the various kurtosis measures due to the base symmetric distribution employed. By and large the various kurtosis measures are very similar in the two cases. Effects on kurtosis of being a little "larger for longer" for the t distribution than the (symmetric) sinh–arcsinhed normal distribution reflect the heavier tails available in the limit in the former case than the latter; conversely, kurtosis, and weight of tails, is smaller in the short-tailed limit for the sinh–arcsinhed normal distribution than for the limiting t (normal) distribution.

2.4.2 *And the Sinh Transformation?*

The alert reader will have realized that requirement (3) is also satisfied by the (simpler) sinh transformation itself, namely,

$$R_{\epsilon,\delta}(z) = \sinh(\delta^{-1}(\epsilon + z)). \tag{13}$$

In this case,

$$R_{\epsilon,\delta}(z) - R_{\epsilon,\delta}(-z) = 2\cosh(\delta^{-1}\epsilon)\sinh(\delta^{-1}z).$$

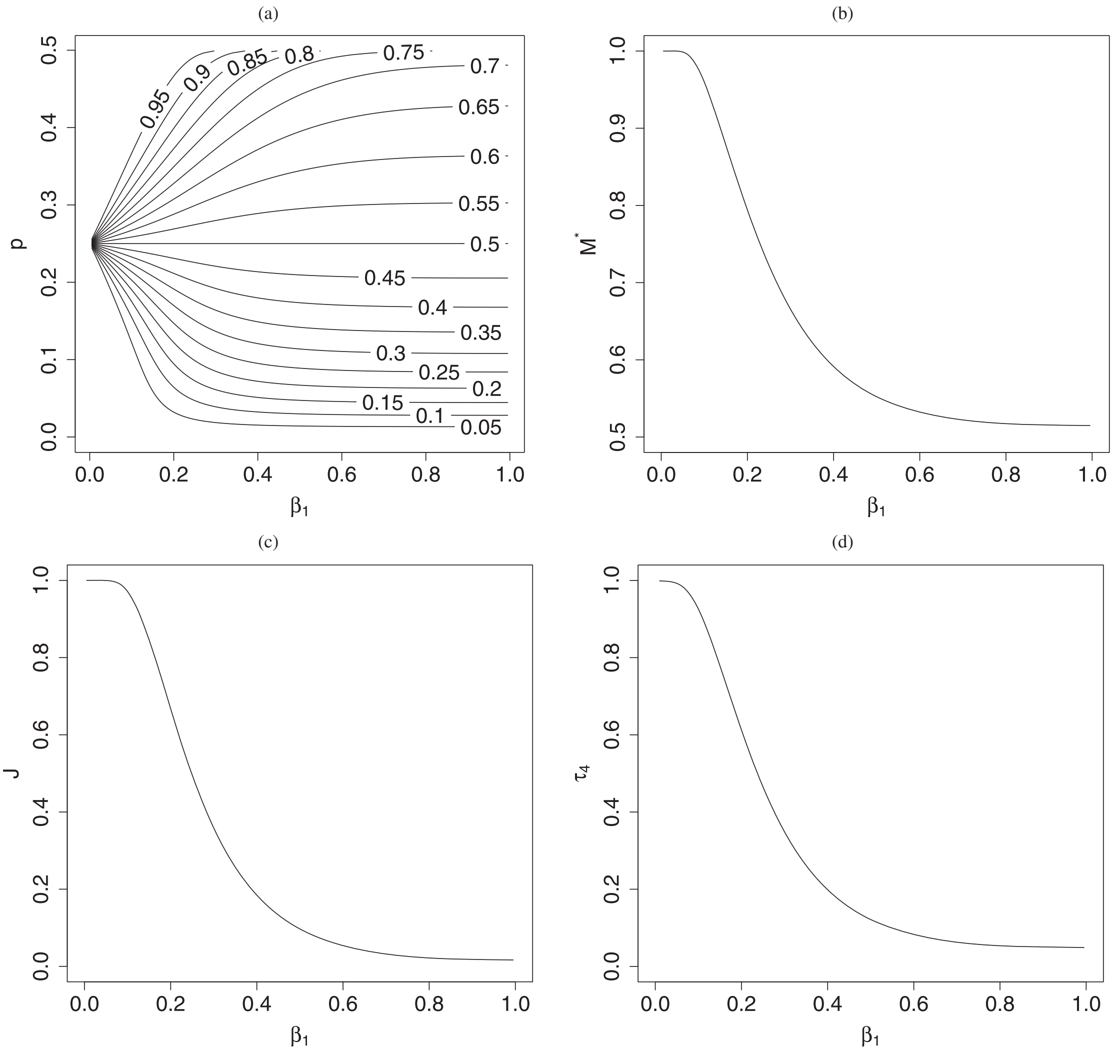


Figure 1. Panel (a) is a contour plot of $t(p)$ as a function of p and $\beta_1 = \delta/(1+\delta)$. Other frames are plots of (b) $M^* = M/(1+M)$, (c) J, and (d) τ_4 as functions of β_1. All four panels are for the sinh–arcsinhed normal distribution.

This transformation was introduced to statistics by Johnson (1949). When applied to normal Z, the resulting distribution is Johnson's S_U distribution. Johnson also suggested applying the transformation to other symmetric Z's—what we shall refer to as Johnson's unbounded family—of which, most prominently, the logistic distribution leads to the L_U distributions of Tadikamalla and Johnson (1982).

Reconciliation with the work of Section 2.3 comes, again, through the sinh–arcsinh transformation! Requirement (4) is not satisfied by the sinh function, which does not afford the identity transformation as a special case for any value of δ. Sinh transformation (13) is, however, a sinh–arcsinh transformation (10): $R_{\epsilon,\delta}(z) = S_{-\epsilon/\delta,1/\delta}(\sinh(z))$ and is therefore T_λ (with the same irrelevantly reparametrized skewness parameter) applied to $Q(Z) = \sinh(\delta Z)$. That is, Johnson's unbounded family has skewness-invariant quantile-based kurtosis measures because they can be thought of as the sinh–arcsinh transformation of the (symmetric) distribution of $\sinh(\delta Z)$.

We prepared plots of the same four kurtosis measures for the Johnson S_U distributions but they are not shown to save space. By skewness-invariance, they correspond, of course, to kurtosis measures for the symmetric S_U distributions with density

$$f_{0,\delta;\phi}(x) = \{\delta\sqrt{2\pi(1+x^2)}\}^{-1} \times \exp\bigl[-\{\sinh^{-1}(x)\}^2/2\delta^2\bigr], \qquad (14)$$

$\delta > 0$. The figure for these distributions is most similar to Figure 1 except for indications of failing to match the smallest kurtosis for the lightest tails of the sinh–arcsinh normal distribution.

Transformations (10) and (13) are clearly special cases of the general transformation

$$W_{\epsilon,\delta}(z) = \sinh\bigl(\delta^{-1}(\epsilon + h(z))\bigr), \qquad (15)$$

where h is an increasing odd function, all of which, when applied to symmetric random variables, afford skewness-invariant quantile-based measures of kurtosis. Only the choice $h(z) = \sinh^{-1}(z)$ allows $W_{0,1}(z) = z$, however, and hence incorporation of the generating distribution g into the heart of the ensuing family.

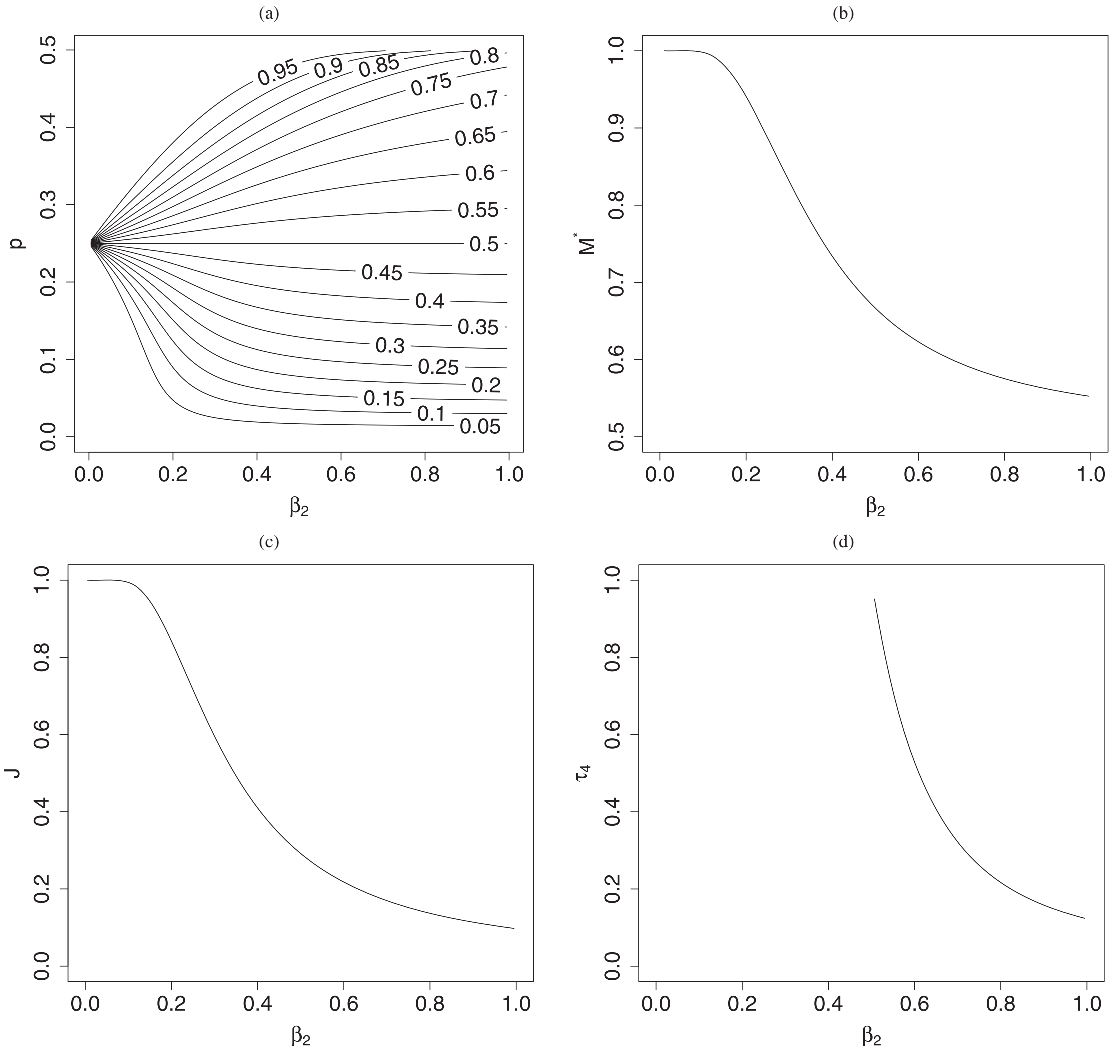


Figure 2. Panel (a) is a contour plot of $t(p)$ as a function of p and $\beta_2 = \nu/(1+\nu)$. Other frames are plots of (b) $M^* = M/(1+M)$, (c) J, and (d) τ_4 as functions of β_2. Note that τ_4 is undefined for $\nu \le 1$. All four panels are for the sinh–arcsinhed t distribution.

3. MORE ON SKEWNESS-INVARIANT MEASURES OF KURTOSIS

3.1 Direct Skewness-Invariant Quantile-Based Kurtosis

In this first subsection of Section 3, we continue to consider quantile-based measures of the form (1). A trawl through distributions defined in terms of their quantile functions (e.g., in Gilchrist 2000) yields just one such distribution with $Q(u) - Q(1-u)$ factorizable into a constant depending on a parameter introducing and, in some sense, controlling skewness, now denoted ϵ, times a function of u (and possibly other parameters). That one is the celebrated Tukey lambda distribution (Hastings et al. 1947; Tukey 1962) which has quantile function

$$Q_\epsilon(u) = \lambda^{-1}\{\epsilon u^\lambda - (1-u)^\lambda\}.$$

For this family of distributions,

$$Q_\epsilon(u) - Q_\epsilon(1-u) = (\epsilon+1)Q_1(u)$$

and so quantile-based measures of kurtosis are invariant to the value of ϵ.

3.2 Skewness-Invariant Density-Based Kurtosis

All the material in Sections 2 and 3.1 relates to quantile-based measures of kurtosis and that in Section 2 is intimately wrapped up with transformation of random variables. There are parallel relationships, described briefly here, for a form of density-based kurtosis and what the first author calls "transformation of scale."

Let the density, f, of X be *unimodal*, and define $x_R(p)$ and $x_L(p)$ to be the left-hand and right-hand solutions of the equation $f(x) = pf(x_0)$ where x_0 is the mode of f. A density-based "asymmetry function" (parallel to extended Bowley skewness in the quantile-based case) is

$$\gamma(p) = \frac{x_R(p) - 2x_0 + x_L(p)}{x_R(p) - x_L(p)}$$

(O'Hagan 1994, section 2.6; Avérous, Fougères, and Meste 1996; Boshnakov 2007; Critchley and Jones 2008). It is natural—in the spirit of Avérous, Fougères, and Meste (1996) but different from what is done by Critchley and Jones (2008)—to define a kurtosis-type measure as a ratio of linear combinations (or appropriate integrals) of quantities of the form

$x_R(p) - x_L(p)$ (for different p). For example, a measure mimicking J might be

$$\mathcal{J} = \frac{x_R(q) - x_L(q) - 3(x_R(p) - x_L(p))}{x_R(q) - x_L(q)}$$

for some fixed $1 > q > p > 0$.

Now, the analogues of skew-symmetric families of distributions based on transforming a symmetric random variable are the following families of distributions based on transformation of scale:

$$f(x) = 2g(2T_\lambda^{-1}(x)),$$

where g is again the density of a distribution symmetric about zero and λ is a parameter introducing skewness (Jones 2011). If g is unimodal, then f is also. But most importantly, for f to be a density, T_λ must satisfy (4) by proposition 1 of Jones (2011).

Now write $c_g(p) = \frac{1}{2}g^{-1}(pg(0)) > 0$ for one-half of the version of $x_R(p)$ associated with g. Then,

$$x_R(p) - x_L(p) = T_\lambda(c_g(p)) - T_\lambda(-c_g(p))$$

and kurtosis-type measures like $\mathcal{J}$ based on a ratio of linear combinations (or appropriate integrals) of these will be independent of λ for precisely the same transformations as in the quantile-based case in Sections 2.2 and 2.3. Essentially, therefore, *all* valid skew-symmetric transformation of scale distributions have (appropriately defined) density-based kurtosis independent of skewness.

3.3 Classical Measures

To close, a brief word on the classical case. Blest (2003) proposed an adjusted version of the classical fourth-moment measure α_4 by introducing a correction for asymmetry. His new coefficient arose out of focusing on what he termed the *meson*; that central value, ξ, about which the fourth moment of a distribution is minimum. Clearly, ξ is also that point about which the third moment is zero. Letting $\xi = \mu + k\sigma$, Blest proposed

$$\alpha_4^* = \alpha_4 - 3k^2(2 + k^2)$$

for an appropriate value of k, as a version of Pearson's coefficient of kurtosis adjusted for skewness. α_4^* proves not to be a kurtosis measure that is completely unaffected by skewness although it reduces the dependence of kurtosis on skewness somewhat; see Rosco, Pewsey, and Jones (2011) for an investigation of the performance of α_4 and of an adjusted version that we propose. In the context of the current article we just note the intriguing fact that k, as derived by Blest but written in different form here, is nothing other than another example of the appearance of the sinh–arcsinh transformation: in terms of (one half of) the classical third moment skewness measure α_3,

$$k = 2\sinh\left\{\frac{1}{3}\sinh^{-1}\left(\frac{\alpha_3}{2}\right)\right\}.$$

[Received October 2010. Revised March 2011.]

REFERENCES

Avérous, J., Fougères, A. L., and Meste, M. (1996), "Tailweight With Respect to the Mode for Unimodal Distributions," *Statistics and Probability Letters*, 28, 367–373. [94]

Balanda, K. P., and MacGillivray, H. L. (1988), "Kurtosis: A Critical Review," *The American Statistician*, 42, 111–119. [89,90]

——— (1990), "Kurtosis and Spread," *Canadian Journal of Statistics*, 18, 17–30. [89-91]

Blest, D. C. (2003), "A Measure of Kurtosis Adjusted for Skewness," *Australian and New Zealand Journal of Statistics*, 45, 175–179. [89,95]

Boshnakov, G. N. (2007), "Some Measures for Asymmetry of Distributions," *Statistics and Probability Letters*, 77, 1111–1116. [94]

Bowley, A. L. (1902), *Elements of Statistics* (2nd ed.), London: P. S. King. [90]

Critchley, F., and Jones, M. C. (2008), "Asymmetry and Gradient Asymmetry Functions: Density-Based Skewness and Kurtosis," *Scandinavian Journal of Statistics*, 35, 415–437. [94]

Gilchrist, W. G. (2000), *Statistical Modelling With Quantile Functions*, Boca Raton: Chapman & Hall/CRC. [94]

Hastings, C., Mosteller, F., Tukey, J. W., and Winsor, C. P. (1947), "Low Moments for Small Samples: A Comparative Study of Order Statistics," *Annals of Mathematical Statistics*, 18, 413–426. [94]

Hosking, J. R. M. (1990), "L-Moments: Analysis and Estimation of Distributions Using Linear Combinations of Order Statistics," *Journal of the Royal Statistical Society, Ser. B*, 52, 105–124. [90]

——— (1992), "Moments or L-Moments—An Example Comparing Two Measures of Distributional Shape," *The American Statistician*, 46, 186–189. [90]

Inman, D. L. (1952), "Measures for Describing the Size Distribution of Sediments," *Journal of Sedimentary Petrology*, 22, 125–145. [90]

Johnson, N. L. (1949), "Systems of Frequency Curves Generated by Methods of Translation," *Biometrika*, 36, 149–176. [93]

Jones, M. C. (2002), "Student's Simplest Distribution," *Journal of the Royal Statistical Society, Ser. D*, 51, 41–49. [92]

——— (2011), "On the Generation of Distributions by Transformation of Scale," unpublished manuscript, Open University, Milton Keynes, U.K. [91,95]

Jones, M. C., and Pewsey, A. (2009), "Sinh–Arcsinh Distributions," *Biometrika*, 96, 761–780. [92]

Moors, J. J. A. (1988), "A Quantile Alternative for Kurtosis," *Journal of the Royal Statistical Society, Ser. D*, 37, 25–32. [90]

O'Hagan, A. (1994), *Kendall's Advanced Theory of Statistics, Volume 2B: Bayesian Inference*, London: Edward Arnold. [94]

Pearson, K. (1905), "Skew Variation, a Rejoinder," *Biometrika*, 4, 169–212. [89]

——— (1916), "IX. Mathematical Contributions to the Theory of Evolution.—XIX. Second Supplement to a Memoir on Skew Variation," *Philosophical Transactions of the Royal Society, Ser. A*, 216, 429–457. [89]

Rosco, J. F., Jones, M. C., and Pewsey, A. (2011), "Skew-t Distributions via the Sinh–Arcsinh Transformation," *Test*, to appear. [92]

Rosco, J. F., Pewsey, A., and Jones, M. C. (2011), "On Blest's Measure of Kurtosis Adjusted for Skewness," unpublished manuscript, University of Extremadura, Caceres, Spain and Open University, Milton Keynes, U.K. [95]

Tadikamalla, P. R., and Johnson, N. L. (1982), "Systems of Frequency Curves Generated by Transformations of Logistic Variables," *Biometrika*, 69, 461–465. [93]

Thiele, T. N. (1889), *Forlæsinger over Almindelig Iagttagelseslære: Sandsynlighedsregning og Mindste Kvadraters Methode*, Copenhagen: C. A. Reitzel. [89]

Tukey, J. W. (1962), "The Future of Data Analysis," *Annals of Mathematical Statistics*, 33, 1–67. [94]

van Zwet, W. R. (1964), *Convex Transformations of Random Variables*, Amsterdam: Mathematisch Centrum. [91,92]

On Blest's measure of kurtosis adjusted for skewness

J.F. Rosco[a], Arthur Pewsey[a,*], M.C. Jones[b]

[a]*Department of Mathematics, Escuela Politécnica, University of Extremadura, 10003 Cáceres, Spain*
[b]*Department of Mathematics & Statistics, The Open University, Walton Hall, Milton Keynes MK7 6AA, UK*

Abstract

We reconsider the derivation of Blest's (2003) skewness adjusted version of the classical moment-based coefficient of kurtosis and propose an adaptation of it which generally eliminates the effects of asymmetry a little more successfully. Lower bounds are provided for the two skewness adjusted kurtosis moment measures as functions of the classical coefficient of skewness. The results from a Monte Carlo experiment designed to investigate the sampling properties of numerous moment-based estimators of the two skewness adjusted kurtosis measures are used to identify those estimators with lowest mean squared error for small to medium sized samples drawn from distributions with varying levels of asymmetry and tailweight.

Keywords: Asymmetry, Estimation, Lower bounds, Moment-based measures, Sinh-arcsinh transformation

1. Introduction

The classical fourth moment-based coefficient $\alpha_4 = \mu_4/\sigma^4$, where $\mu_k = E[(X-\mu)^k]$, $\sigma^2 = \mu_2$, $\mu = E(X)$ and X denotes a random variable (Thiele, 1889; Pearson, 1905), remains the best known and most widely applied measure of kurtosis. This is in spite of the fact that the coefficient does not exist if the fourth moment does not exist, a major limitation on its use with heavy-tailed distributions. Moreover, even for symmetric distributions, its interpretation can be far from obvious, and many alternatives have been proposed. For asymmetric distributions, it has long been known (Pearson, 1916) that $\alpha_4 \geq \alpha_3^2 + 1$, where $\alpha_3 = \mu_3/\sigma^3$ is the classical moment-based coefficient of skewness. Thus, higher skewness (as measured by α_3) is inevitably accompanied by higher kurtosis (as measured by α_4). These unappealing features of α_4 have stimulated considerable debate within the literature regarding exactly what 'kurtosis' is, what it measures (or should measure), and how best to measure it. For example, the Tukey school's view is summarised in their use of the word 'elongation', broadly 'tailweight' (Hoaglin et al., 1985, Chapters 10 and 11), while often

*Corresponding author
Email addresses: `juanfranrn@unex.es` (J.F. Rosco), `apewsey@unex.es` (Arthur Pewsey), `m.c.jones@open.ac.uk` (M.C. Jones)

'peakedness' is used, the two terms emphasising different aspects of what we will continue to call 'kurtosis'. van Zwet's celebrated, but quite different, approach (van Zwet, 1964) via transformation-based ordering of distributions is also relevant. An excellent review of the extensive related literature is provided by Balanda and MacGillivray (1988). Many of the alternative kurtosis measures that have been proposed are based on quantiles, which exist and are unique if the distribution function is continuous and strictly monotone.

In Jones et al. (2011), we investigated in a much more general setting the issue of whether kurtosis could be measured independently of skewness. In Section 1 of that paper, we argued that definitions of kurtosis make no reference whatsoever to the skewness of a distribution: 'peakedness' relates to the 'tightness' of the main body of a distribution, peakedness "only makes sense relative to the weights of the tails of the distribution", that is, 'elongation', while Balanda and MacGillivray (1988) define kurtosis as the "location-and scale free movement of probability mass from the shoulders of a distributions into its centre and tails." So the quest for skewness-invariant measures of kurtosis seems to make sense. Measures of kurtosis for use with asymmetric distributions were considered by Balanda and MacGillivray (1990) and in more detail by Jones et al. (2011).

A specific skewness adjusted version of the moment measure α_4 was proposed by Blest (2003). His proposal arose from consideration of what he termed the *meson*; that central value, ξ, about which the fourth moment of a distribution is minimum. Clearly, ξ is also that point about which the third moment is zero. Setting $\xi = \mu + k\sigma$,

$$k = \left(\sqrt{1+\tfrac{1}{4}\alpha_3^2}+\tfrac{1}{2}\alpha_3\right)^{1/3} - \left(\sqrt{1+\tfrac{1}{4}\alpha_3^2}-\tfrac{1}{2}\alpha_3\right)^{1/3}, \tag{1}$$

and thus

$$\mu_4 = \mu_4^* + 6\sigma^4 k^2 + 3\sigma^4 k^4,$$

where $\mu_4^* = E[(X-\xi)^4]$ denotes the minimum fourth moment. Given this relation, Blest proposed

$$\alpha_4^* = \mu_4^*/\sigma^4 = \alpha_4 - 3k^2(2+k^2), \tag{2}$$

as a moment-based measure of kurtosis adjusted for skewness, his clear intention being to try to eliminate the effects of skewness on α_4 noted earlier. Jones et al. (2011) note that k can be represented in terms of the sinh-arcsinh function as

$$k = 2\sinh(\tfrac{1}{3}\sinh^{-1}(\tfrac{1}{2}\alpha_3)) = 2S_{0,\frac{1}{3}}(\tfrac{1}{2}\alpha_3),$$

using the notation $S_{\varepsilon,\delta}(x) = \sinh(\delta\sinh^{-1}(x)-\varepsilon)$ of Jones and Pewsey (2009).

We reconsider the derivation of Blest's (2003) skewness adjusted version of the classical moment-based coefficient of kurtosis and propose an adaptation of it which generally eliminates the effects of asymmetry a little more successfully. We also consider estimation of the two skewness adjusted kurtosis measures. This paper takes the viewpoint that, whatever the advantages of non-moment-based measures of kurtosis, many researchers continue to equate 'skewness' and 'kurtosis' with α_3 and α_4, respectively. It is to those readers, who might be nudged in the direction of improving moment-based kurtosis as regards its relationship to skewness, and hence consider using Blest's approach, that the paper is addressed.

In Section 2, we reconsider the definition of α_4^* and propose our adaptation of it, $\alpha_4^\dagger$. In the same section, we show that neither α_4^* nor $\alpha_4^\dagger$ are moment-based kurtosis measures that are completely unaffected by skewness. We also provide lower bounds for the two skewness adjusted kurtosis measures. In Section 3 we consider the problem of how α_4^* and $\alpha_4^\dagger$ might be estimated, and present results of an extensive simulation study designed to explore the performance of various estimators based on popular estimators of the skewness measure α_3 and the kurtosis measure α_4. The paper ends with Section 4 where concluding remarks are drawn.

2. An alternative measure: comparative performance and bounds

2.1. An alternative skewness adjusted measure

It is easy to show that $\mu_2^* = E[(X-\xi)^2] = \sigma^2(1+k^2)$. This result raises the question as to why, in the definition of α_4^* in Equation (2), μ_4^* is divided by σ^4 and not $\sigma^4(1+k^2)^2$. We therefore propose the alternative moment-based skewness adjusted coefficient of kurtosis

$$\alpha_4^\dagger = \frac{\mu_4^*}{(\mu_2^*)^2} = \frac{\alpha_4^*}{(1+k^2)^2} = \frac{\alpha_4}{(1+k^2)^2} - \frac{3k^2(2+k^2)}{(1+k^2)^2}. \tag{3}$$

Like α_4 and α_4^*, $\alpha_4^\dagger$ does not exist if the fourth moment of X does not exist. As is the case for α_4^*, the new measure $\alpha_4^\dagger$ is a function of α_4 and a sinh-arcsinh transformation of the coefficient of skewness, α_3.

2.2. Performance of the skewness adjusted kurtosis measures

Although α_4^* and $\alpha_4^\dagger$ are generally less affected by skewness than α_4 is, they are not, however, skewness invariant measures. This fact is illustrated in Figures 1 and 2. Figure 1 represents all three measures for the popular skew-normal class of distributions of Azzalini (1985) with density

$$f_\alpha(x) = 2\phi(x)\Phi(\alpha x), \qquad -\infty < x, \alpha < \infty, \tag{4}$$

where ϕ and Φ are the density and distribution function, respectively, of the standard normal distribution. The parameter α is a shape parameter which affects both the skewness and kurtosis. The skew-normal distribution has shapes ranging from that of the normal distribution ($\alpha = 0$) to those of half-normal distributions ($\alpha = \pm\infty$). In Figure 1, both the measures and the shape parameter, constrained without loss of generality to be positive, have been transformed to put them on to $(0,1)$. (When $\rho = \alpha/(1+\alpha) = 0$, each kurtosis measure is $3/(1+3) = 0.75$, the α_4 kurtosis value of the normal distribution.) If the effects of asymmetry were eliminated completely for all members of the class, we would expect to see lines that were parallel with the horizontal axis in such a plot. Clearly they are not, but $\alpha_4^\dagger$ appears to do a better job than α_4^* at removing the effects of skewness for all but the most asymmetric of cases, in the neighbourhood of the half-normal ($\alpha = \infty$, $\rho = 1$) distribution.

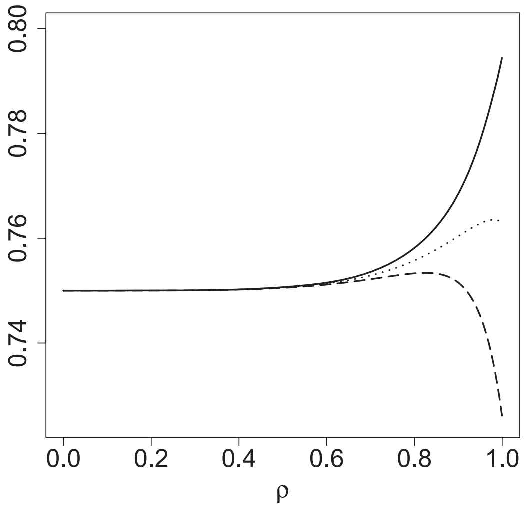


Figure 1: Moment-based kurtosis measures $\alpha_4/(1+\alpha_4)$ (solid), $\alpha_4^*/(1+\alpha_4^*)$ (dotted) and $\alpha_4^\dagger/(1+\alpha_4^\dagger)$ (dashed), as functions of $\rho = \alpha/(1+\alpha), \alpha > 0$, for the skew-normal distribution with density (4).

Panels (a)–(c) of Figure 2 present contour plots of $\alpha_4/(1+\alpha_4)$, $\alpha_4^*/(1+\alpha_4^*)$ and $\alpha_4^\dagger/(1+\alpha_4^\dagger)$, as functions of $\rho_1 = \varepsilon/(1+\varepsilon)$, $\varepsilon \geq 0$, and $\lambda_1 = \delta/(1+\delta)$, for the sinh-arcsinhed normal (or SAS-normal, for short) family of distributions of Jones and Pewsey (2009) with density

$$f_{\varepsilon,\delta}(x) = \{2\pi(1+x^2)\}^{-1/2}\delta C_{\varepsilon,\delta}(x)\exp\{-\tfrac{1}{2}S^2_{\varepsilon,\delta}(x)\}, \qquad -\infty < x, \varepsilon < \infty,\ \delta > 0, \tag{5}$$

where $C_{\varepsilon,\delta}(x) = \cosh\{\delta\sinh^{-1}(x) - \varepsilon\} = \{1 + S^2_{\varepsilon,\delta}(x)\}^{1/2}$. Here, δ is a tailweight parameter, while ε regulates the skewness of the distribution. The SAS-normal distribution has tails ranging from the extremely heavy ($\delta \simeq 0$), through those of the normal distribution ($\delta = 1$) to the extremely light ($\delta \to \infty$). Its densities are symmetric if $\varepsilon = 0$, and increasingly positively (negatively) skewed as $\varepsilon \to \infty$ ($\varepsilon \to -\infty$). In the contour plots of panels (a)–(c), we would expect to see contour lines that were parallel with the horizontal axis if the effects of asymmetry were eliminated completely. Here it is debatable which of the two forms of correction does best at removing the effects of asymmetry, effects that are not especially strong to start with in this case. Certainly for moderate levels of asymmetry and perhaps for high levels of asymmetry, $\alpha_4^\dagger$ performs best. However, for distributions with heavy tails ($\delta < 1, \lambda_1 < 1/2$) and low levels of asymmetry ($\varepsilon \simeq 0, \rho_1 \simeq 0$), α_4^* performs better.

Panels (d)–(f) of Figure 2 portray contour plots analogous to those in panels (a)–(c), now as functions of $\rho_1 = \varepsilon/(1+\varepsilon)$, $\varepsilon \geq 0$, and $\lambda_2 = \nu/(1+\nu)$, for the sinh-arcsinhed t distribution of Rosco et al. (2011) with density

$$f_{\varepsilon,\nu}(x) = \frac{K_\nu C_{\varepsilon,1}(x)}{\sqrt{1+x^2}(1+\nu^{-1}S^2_{\varepsilon,1}(x))^{(\nu+1)/2}}, \qquad -\infty < x, \varepsilon < \infty, \quad \nu > 0, \tag{6}$$

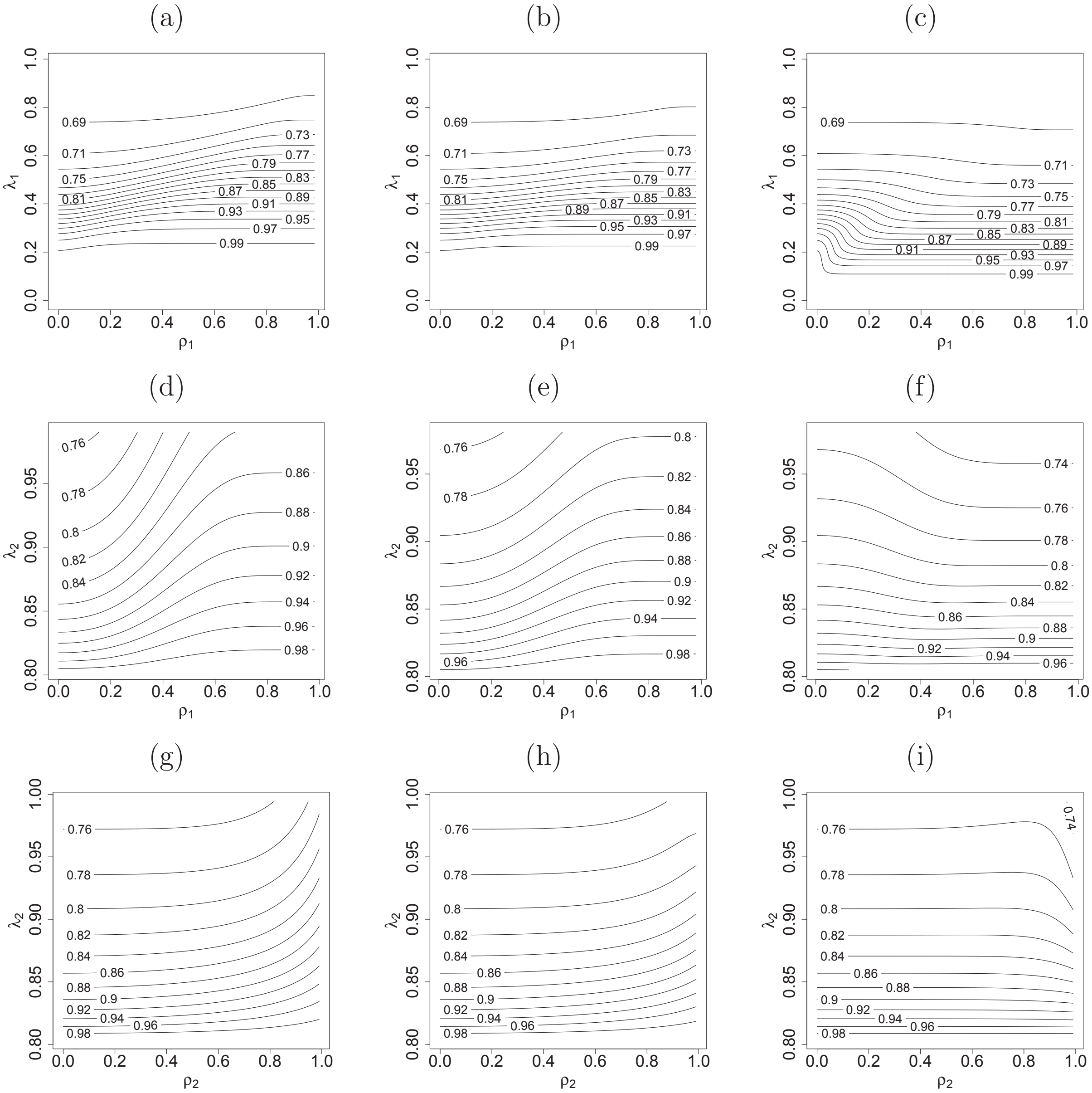


Figure 2: Contour plots of the three moment-based kurtosis measures $\alpha_4/(1+\alpha_4)$ (first column), $\alpha_4^*/(1+\alpha_4^*)$ (second column) and $\alpha_4^\dagger/(1+\alpha_4^\dagger)$ (third column). Panels (a)–(c) correspond to the SAS-normal distribution with density (5), as functions of $\rho_1 = \varepsilon/(1+\varepsilon), \varepsilon \geq 0$, and $\lambda_1 = \delta/(1+\delta)$. Panels (d)–(f) are their analogues for the SAS-t distribution with density (6), as functions of $\rho_1 \geq 0$ and $\lambda_2 = \nu/(1+\nu)$. Panels (g)–(i) correspond to the skew-t distribution with density (7), as functions of $\rho_2 = \alpha/(1+\alpha), \alpha \geq 0$, and λ_2.

where $K_\nu = \Gamma((\nu+1)/2)/(\sqrt{\nu\pi}\Gamma(\nu/2))$. As for the SAS-normal distribution, ε is the skewness regulating parameter. However, ν replaces δ as the tailweight parameter. The SAS-t distribution has tails ranging from the extremely heavy ($\nu \simeq 0$), through those of the Cauchy distribution ($\nu = 1$), all the way to those of the normal distribution ($\nu \to \infty$). However, the moment-based kurtosis measures are only defined for $\nu > 4$, or $\lambda_2 > 0.8$. For this family of distributions, $\alpha_4^\dagger$ can probably be judged to generally perform best.

Finally, panels (g)–(i) of Figure 2 provide analogous contour plots, now as functions of $\rho_2 = \alpha/(1+\alpha)$, $\alpha \geq 0$, and $\lambda_2 = \nu/(1+\nu)$, for the skew-t distribution of Azzalini and Capitanio (2003) with density

$$f_{\alpha,\nu}(x) = 2t_\nu(x)T_{\nu+1}\left\{\alpha x\left(\frac{\nu+1}{x^2+\nu}\right)^{1/2}\right\}, \qquad -\infty < x, \alpha < \infty, \quad \nu > 0, \tag{7}$$

where t_ν and T_ν denote the density and distribution function, respectively, of the t-distribution with ν degrees of freedom. Here, α is a skewness parameter (as for the skew-normal class) and ν is a tailweight parameter (as for the SAS-t family). Again, the moment-based kurtosis measures are only defined for $\nu > 4$, or $\lambda_2 > 0.8$. For this family, $\alpha_4^\dagger$ generally performs best, particularly for distributions with low to moderate levels of skewness.

Thus, although the ability of the moment-based kurtosis measures α_4^* and $\alpha_4^\dagger$ to remove the influence of skewness clearly depends on the family of distributions under consideration and the level of skewness, our findings for the four flexible families of unimodal distributions considered here indicate that $\alpha_4^\dagger$ generally outperforms α_4^*, if not by a huge amount. It is noteworthy that, in the examples of Figure 2, α_4^* actually makes little difference compared with α_4; on the other hand, $\alpha_4^\dagger$ makes more difference, although sometimes it seems to adjust α_4 a little bit too much.

2.3. Lower bounds

As stated in the Introduction, the standard moment-based kurtosis measure α_4 is bounded below by α_3^2+1. Here we consider lower bounds for the two moment-based skewness adjusted measures, α_4^* and $\alpha_4^\dagger$.

The key to obtaining a lower bound for α_4^* is the following simple bound for the 'symmetric' (actually, odd) sinh-arcsinh function $S_{0,\delta}(x) = \sinh(\delta\sinh^{-1}(x))$ when $0 \leq \delta \leq 1$ and $x \geq 0$: $S_{0,\delta}(x) \leq \delta x$. This follows because $S_{0,\delta}(0) = 0$, $S'_{0,\delta}(0) = \delta$ and, with just a little effort, $S_{0,\delta}(x)$ with $0 \leq \delta \leq 1$ can be shown to be concave on $x \geq 0$. It follows that $k = 2S_{0,\frac{1}{3}}(\alpha_3/2) \leq \alpha_3/3$ for $\alpha_3 \geq 0$ and hence, since k is an odd function of α_3,

$$k^2 \leq \tfrac{1}{9}\alpha_3^2.$$

(Blest (2003) notes essentially that $k \approx \alpha_3/3$ which is indeed a good approximation for small α_3.) Finally,

$$\alpha_4^* = \alpha_4 - 3k^2(2+k^2) \geq \alpha_3^2 + 1 - \tfrac{1}{3}\alpha_3^2\left(2 + \tfrac{1}{9}\alpha_3^2\right) = 1 + \tfrac{1}{3}\alpha_3^2 - \tfrac{1}{27}\alpha_3^4.$$

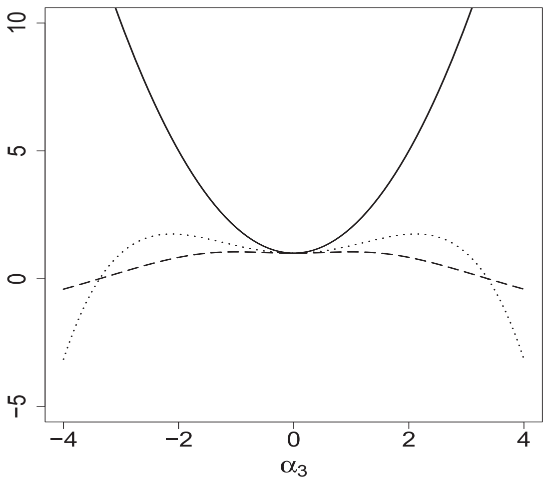


Figure 3: Lower bounds for α_4 (solid), α_4^* (dotted) and $\alpha_4^\dagger$ (dashed) as functions of the skewness measure α_3.

The same bound divided by $(1+\frac{1}{9}\alpha_3^2)^2$ clearly holds for $\alpha_4^\dagger = \alpha_4^*/(1+k^2)^2$. That is,

$$\alpha_4^\dagger \geq \frac{1+\frac{1}{3}\alpha_3^2 - \frac{1}{27}\alpha_3^4}{(1+\frac{1}{9}\alpha_3^2)^2} = 3\left(\frac{27+9\alpha_3^2-\alpha_3^4}{81+18\alpha_3^2+\alpha_3^4}\right).$$

Figure 3 portrays the lower bounds for α_4, α_4^* and $\alpha^\dagger$ as functions of α_3. All three lower bounds are clearly identical, and equal to one, if the underlying distribution is symmetric (and, indeed, for any distribution such that $\alpha_3 = 0$). The bounds for α_4^* and $\alpha_4^\dagger$ are not dissimilar for α_3 values within the plotted range. However, $\alpha_4^* \to -\infty$ as $|\alpha_3| \to \infty$, while $\alpha_4^\dagger \to -3$ as $|\alpha_3| \to \infty$. The lower bounds on the skewness-adjusted kurtosis measures are much less stringent than the classical lower bound on the value of α_4.

3. Estimation

When working with data using moment-based measures, it will of course be of interest to estimate the values of α_4^* and $\alpha_4^\dagger$, and this is the problem we consider here. Specifically, we focus on estimators of them based on popular estimators of α_3 and α_4. We introduce the underlying estimators of α_3 and α_4 in Section 3.1, and present the results from a simulation study designed to explore the performance of twelve estimators of each of α_4^* and $\alpha_4^\dagger$ in Section 3.2.

3.1. Estimators of α_3 and α_4

Let $X_1, ..., X_n$ denote a random sample from some unspecified distribution, and $\bar{X} = n^{-1}\sum_{i=1}^n X_i$, $M_k = n^{-1}\sum_{i=1}^n (X_i - \bar{X})^k$, $\tilde{\alpha}_3 = M_3/M_2^{3/2}$ and $\tilde{\alpha}_4 = g_2 + 3 = M_4/M_2^2$ denote the sample mean, the kth moment about the mean, and the classical sample moment estimators of α_3 and α_4, respectively. For data from a normal distribution, $\tilde{\alpha}_3$ is unbiased for α_3, whereas $\tilde{\alpha}_4$ is only asymptotically unbiased for α_4. For data from other distributions, the two estimators are asymptotically unbiased (see, for example, Đorić et al., 2009). $\tilde{\alpha}_3$ and $\tilde{\alpha}_4$ are the estimators of α_3 and α_4 implemented in the statistical software package STATA and the `moments` package of R.

Fisher (1930) proposed

$$\tilde{\alpha}_3' = \tilde{\alpha}_3 \frac{\sqrt{n(n-1)}}{n-2} \quad \text{and} \quad G_2 = \frac{n-1}{(n-2)(n-3)}\{(n+1)(\tilde{\alpha}_4 - 3) + 6\}$$

as estimators of α_3 and $\alpha_4 - 3$. We will denote the corresponding estimator of α_4 by $\tilde{\alpha}_4' = G_2 + 3$. For samples drawn from the normal distribution, $\tilde{\alpha}_3'$ and $\tilde{\alpha}_4'$ are unbiased. These are the estimators of α_3 and α_4 implemented within the packages SAS, SPSS and STATISTICA.

Making use of the unbiased estimators $M_2' = nM_2/(n-1)$, $M_3' = n^2 M_3/\{(n-1)(n-2)\}$ and

$$M_4' = \frac{n(n^2 - 2n + 3)}{(n-1)(n-2)(n-3)} M_4 - \frac{3n(2n-3)}{(n-1)(n-2)(n-3)} M_2^2,$$

of their population central moment counterparts, Cramér (1946) considered the estimators

$$\frac{M_3'}{(M_2')^{3/2}} = \tilde{\alpha}_3' \quad \text{and} \quad \tilde{\alpha}_4'' = \frac{M_4'}{(M_2')^2}. \tag{8}$$

As Đorić et al. (2009) explain, $\tilde{\alpha}_4''$ is biased with the same bias as $\tilde{\alpha}_4$ when the data are normal. More generally, $\tilde{\alpha}_3'$ and $\tilde{\alpha}_4''$ are biased but with smaller biases than $\tilde{\alpha}_3$ and $\tilde{\alpha}_4$.

The estimators of α_3 and $\alpha_4 - 3$ implemented in MINITAB, BMDP and the `timeDate` package of R are

$$\tilde{\alpha}_3'' = \frac{M_3}{(M_2')^{3/2}} = \tilde{\alpha}_3 \left(\frac{n-1}{n}\right)^{3/2} \quad \text{and} \quad b_2 = \frac{M_4}{(M_2')^2} - 3 = \tilde{\alpha}_4 \left(\frac{n-1}{n}\right)^2 - 3.$$

We will use $\tilde{\alpha}_4''' = b_2 + 3$ to denote the corresponding estimator of α_4. Like $\tilde{\alpha}_3'$, $\tilde{\alpha}_3''$ is a multiple of $\tilde{\alpha}_3$ and thus is also an unbiased estimator of $\alpha_3 = 0$ when the data are normal.

Joanes and Gill (1998) present results for the variances of the estimators $\tilde{\alpha}_3$, $\tilde{\alpha}_3'$ and $\tilde{\alpha}_3''$ and for the biases and variances of the estimators g_2, G_2 and b_2 for samples drawn from the normal distribution. They also summarise Monte Carlo based results for the bias and mean squared error (MSE) of the same estimators for data drawn from chi-squared distributions with varying levels of asymmetry, specifically, with 1, 10 and 50 degrees of freedom. They found all six estimators to be negatively biased for samples drawn from these positively skewed distributions, the bias decreasing with increasing sample size, n, and number of degrees of freedom. Based on their results, it can be concluded that $\tilde{\alpha}_3''$ and $\tilde{\alpha}_4'''$ have the smallest variances for samples drawn from the normal distribution, while $\tilde{\alpha}_3''$ and $\tilde{\alpha}_4$ have the smallest MSEs in the normal case. On the other hand, $\tilde{\alpha}_3'$ and $\tilde{\alpha}_4'$, for $n < 100$, and $\tilde{\alpha}_4$, for $100 \le n \le 200$, have the smallest MSEs for samples from a very skewed distribution like the chi-squared distribution with 1 degree of freedom.

3.2. Simulation study

There are twelve possible combinations of the three estimators $\tilde{\alpha}_3$, $\tilde{\alpha}'_3$ and $\tilde{\alpha}''_3$ of α_3 and the four estimators $\tilde{\alpha}_4$, $\tilde{\alpha}'_4$, $\tilde{\alpha}''_4$ and $\tilde{\alpha}'''_4$ of α_4 which one might contemplate substituting for α_3 and α_4 in (1)–(3) so as to obtain estimators of k, α_4^* and $\alpha_4^\dagger$. We identify these twelve combinations using the numbers: 1 for $(\tilde{\alpha}_3, \tilde{\alpha}_4)$, 2 for $(\tilde{\alpha}'_3, \tilde{\alpha}_4)$, 3 for $(\tilde{\alpha}''_3, \tilde{\alpha}_4)$, 4 for $(\tilde{\alpha}_3, \tilde{\alpha}'_4)$, 5 for $(\tilde{\alpha}'_3, \tilde{\alpha}'_4)$, 6 for $(\tilde{\alpha}''_3, \tilde{\alpha}'_4)$, 7 for $(\tilde{\alpha}_3, \tilde{\alpha}''_4)$, 8 for $(\tilde{\alpha}'_3, \tilde{\alpha}''_4)$, 9 for $(\tilde{\alpha}''_3, \tilde{\alpha}''_4)$, 10 for $(\tilde{\alpha}_3, \tilde{\alpha}'''_4)$, 11 for $(\tilde{\alpha}'_3, \tilde{\alpha}'''_4)$, 12 for $(\tilde{\alpha}''_3, \tilde{\alpha}'''_4)$. In order to study the small-sample bias and MSE properties of the twelve resulting estimators of α_4^* and of $\alpha_4^\dagger$, we carried out a simulation study.

In our study we generated samples of size $n = 10, 20, 50, 100$ and 200 from the SAS-normal distribution with density (5), the SAS-t distribution with density (6), and Azzalini and Capitanio's skew-t distribution with density (7). We chose these three models because of their unimodal flexibility. For each of the three families of distributions we considered values of their skewness parameters (ε for the first two, and α for the last) of $0, 0.5, 1$ and 10. For the two asymmetric t distributions we explored values of their tailweight parameter, ν, of $4.1, 10$ and ∞. (The $\nu = \infty$ cases correspond to the SAS-normal distribution with $\delta = 1$ and the skew-normal distribution, respectively.) And for the SAS-normal we investigated values for its tailweight parameter, δ, of $0.2, 0.5, 2, 5$ and 20. These parameter combinations correspond to ranges of α_4 of: $(2.14, 1154.60)$ for the SAS-normal; $(3, 266.18)$ for the SAS-t; $(3, 230.70)$ for the skew-t. For each distribution, sample size, asymmetry parameter value and tailweight parameter value combination we simulated $10,000$ samples, and from these samples we calculated the sample bias and MSE of each of the twelve estimators of α_4^* and each of the twelve estimators of $\alpha_4^\dagger$.

Consistent with the results quoted above from Joanes and Gill (1998) and there being relatively little difference between α_4 and α_4^*, the biases of all the estimates of α_4^* were found to be negative, the bias decreasing (in absolute value) with increasing sample size and as the tailweight tends to that of the normal distribution and, generally, as the skewness tends to 0 (i.e. to symmetry). With regard to the MSE of the twelve estimators of α_4^*, for distributions with normal or heavier tails we observed patterns which are well represented by panels (a) and (c) of Figure 4. For distributions with lighter than normal tails, patterns like those displayed in panel (e) of the same figure were obtained. As panels (a), (c) and (e) of Figure 4 illustrate, there is little or no difference between the MSEs of the twelve estimators of α_4^* for sample sizes of 100 or more.

The results obtained for the estimators of $\alpha_4^\dagger$ were very similar to those for the estimators of α_4^*, except that their biases and MSEs were generally found to be somewhat larger. The difference between their MSEs can be appreciated by comparing the panels corresponding to the estimates of α_4^* in the first column of Figure 4 with their counterparts for the estimates of $\alpha_4^\dagger$ in its second column.

For samples drawn from distributions with heavier than normal tails, for example, Figure 4(a),(b), the estimators of α_4^* and $\alpha_4^\dagger$ which generally had lowest MSEs were those based on the combinations 9 $(\tilde{\alpha}''_3, \tilde{\alpha}''_4)$, 7 $(\tilde{\alpha}_3, \tilde{\alpha}''_4)$ and 8 $(\tilde{\alpha}'_3, \tilde{\alpha}''_4)$ (ordered according to increasing MSE). The estimator $\tilde{\alpha}''_4$ which appears in all three of these combinations was not considered by Joanes and Gill (1998) as a potential estimator of α_4. For samples from distributions with normal-like tails, for example, Figure 4(c),(d), the estimators with lowest MSEs generally

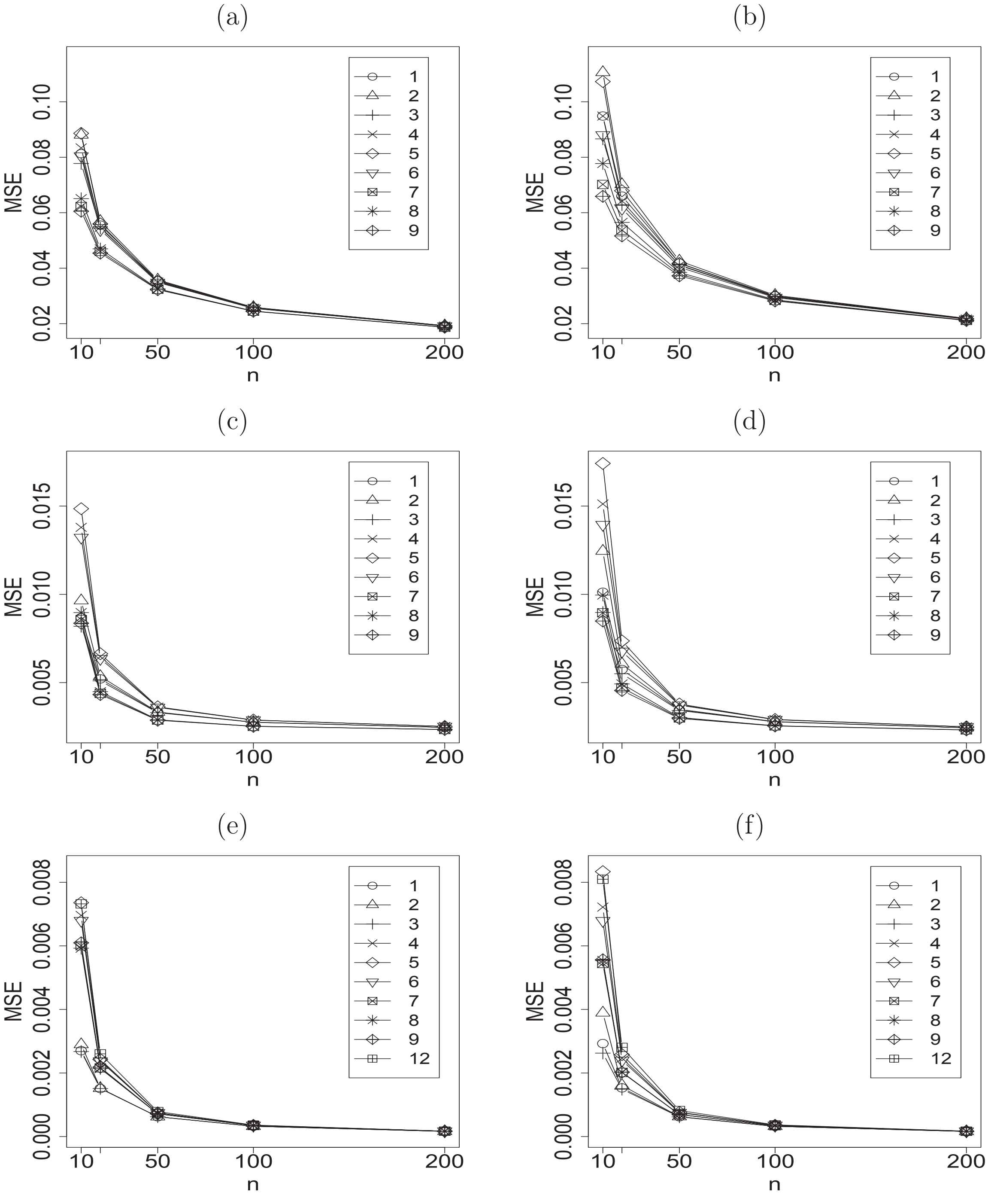


Figure 4: Empirical MSE, as a function of sample size, n, of estimators of $\alpha_4^*/(1+\alpha_4^*)$ (first column) and $\alpha_4^\dagger/(1+\alpha_4^\dagger)$ (second column) based on the combinations of the estimators of α_3 and α_4 identified in the keys. The rows correspond to data simulated from the: t distribution with $\nu = 4.1$ (first); skew-normal distribution with $\alpha = 1$ (second); SAS-normal distribution with $\delta = 20$ and $\varepsilon = 10$ (third). The results for those combinations producing the highest MSEs have been omitted so as to aid the identification of the combinations corresponding to the estimators with the lowest MSEs.

corresponded to the combinations 3 $(\tilde{\alpha}_3'', \tilde{\alpha}_4)$ and 9 $(\tilde{\alpha}_3'', \tilde{\alpha}_4'')$. Both of these combinations involve $\tilde{\alpha}_3''$ which was found by Joanes and Gill (1998) to be the estimator of α_3 with smallest MSE for data drawn from the normal distribution. Finally, for samples drawn from distributions with lighter than normal tails, for example, Figure 4(e),(f), the estimators based on the combinations 1 $(\tilde{\alpha}_3, \tilde{\alpha}_4)$, 2 $(\tilde{\alpha}_3', \tilde{\alpha}_4)$ and 3 $(\tilde{\alpha}_3'', \tilde{\alpha}_4)$ were found generally to be those with lowest MSEs. All three of these combinations contain the raw moment estimator $\tilde{\alpha}_4$ of α_4. Here a comparison with the results reported in Joanes and Gill (1998) is impossible because they did not investigate the performance of the different estimators for data simulated from light tailed distributions. The estimators corresponding to the combinations 11 $(\tilde{\alpha}_3', \tilde{\alpha}_4''')$ and 10 $(\tilde{\alpha}_3, \tilde{\alpha}_4''')$ were found consistently to be the ones with the largest MSEs, and this is the reason why the results for them have been omitted from Figure 4. Both combinations involve the estimator $\tilde{\alpha}_4'''$ of α_4.

The lessons gleaned from our simulation study are pulled together in Section 4 below.

4. Concluding remarks

In this paper we have proposed and investigated $\alpha_4^\dagger$, an adaptation of Blest's (2003) moment-based coefficient of kurtosis adjusted for skewness, α_4^*. For four flexible unimodal models considered in Section 2.2, $\alpha_4^\dagger$ was found generally to outperform α_4^*, though by relatively small amounts, in terms of its ability to remove the effects of asymmetry. Also, the lower bound for $\alpha_4^\dagger$ is closer to being constant than that for α_4^*.

Our Monte Carlo investigation of the MSEs of various moment-based estimators of α_4^* and $\alpha_4^\dagger$, reported in Section 3.2, identified the estimators corresponding to the combinations of any of the estimators of α_3 with $\tilde{\alpha}_4''$ as being the ones which generally performed best when working with samples drawn from distributions with heavier than normal tails. On the other hand, for samples drawn from distributions with lighter than normal tails, the estimators based on the combinations of any of the estimators of α_3 with $\tilde{\alpha}_4$ were found generally to perform best. In the intermediate case, for samples from distributions with close to normal tails, the estimators which generally performed best were those corresponding to the combinations 3 $(\tilde{\alpha}_3'', \tilde{\alpha}_4)$ and 9 $(\tilde{\alpha}_3'', \tilde{\alpha}_4'')$. It seems appropriate, therefore, to recommend use of $\tilde{\alpha}_3''$ throughout. The most appropriate estimator of α_4 depends on tailweight; $\tilde{\alpha}_4''$ would seem to be the more usual choice, as it is good for heavier and normal tails, but users should be aware that its performance is not so good for light tails. That said, of all the different combinations considered, only combination 1 involves estimators of both α_3 and α_4 — the classical moment estimators — which are readily available within all of the major statistical packages. These conclusions all apply to estimators of both α_4^* and $\alpha_4^\dagger$, but it has to be admitted that the performance of estimators of $\alpha_4^\dagger$ is generally a little inferior to those of α_4^*.

Like α_4 and α_4^*, $\alpha_4^\dagger$ is a moment-based measure which will not exist if the fourth moment does not exist. As stated in the Introduction, the potential non-existence of moment-based kurtosis measures has rightly led researchers to propose numerous alternative measures of kurtosis. In Jones et al. (2011), we identify two wide classes of quantile-based kurtosis measures which are skewness-invariant for certain families of distributions. Development of the ideas explored there we consider to warrant future investigation.

Acknowledgements

Financial support for this research was received by Rosco from the European Social Fund and by Pewsey in the form of grant MTM2010-16845 from the Spanish Ministry of Science and Education and grant GR10064 from the Junta de Extremadura.

References

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

Azzalini, A., Capitanio, A., 2003. Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t distribution. J. Roy. Statist. Soc. B 65, 367–389.

Balanda, K.P., MacGillivray, H.L., 1988. Kurtosis a critical review. Amer. Statist. 42, 111–119.

Balanda, K.P., MacGillivray, H.L., 1990. Kurtosis and spread. Canad. J. Statist. 18, 17–30.

Blest, D.C., 2003. A new measure of kurtosis adjusted for skewness. Aust. N. Z. J. Statist. 45, 175–179.

Cramér, H., 1946. Mathematical Methods of Statistics. Princeton University Press, Princeton.

Đorić, D., Nikolić-Đorić, E., Jevremović, V., Mališić, J., 2009. On measuring skewness and kurtosis. Qual. Quant. 43, 481–493.

Fisher, R.A., 1930. The moments of the distribution for normal samples of measures of departures from normality. Proc. R. Soc. Lond. A 130, 16–28.

Hoaglin, D.C., Mosteller, F., Tukey, J.W. (Eds.), 1985. Exploring Data Tables, Trends, and Shapes. Wiley, New York.

Joanes, D.N., Gill, C.A., 1998. Comparing measures of sample skewness and kurtosis. Statistician 47, 183–189.

Jones, M.C., Pewsey, A., 2009. Sinh-arcsinh distributions. Biometrika 96, 761–780.

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

Pearson, K., 1905. Skew variation, a rejoinder. Biometrika 4, 169–212.

Pearson, K., 1916. IX. Mathematical contributions to the theory of evolution.—XIX. Second supplement to a memoir on skew variation. Philos. Trans. R. Soc., A 216, 429–457.

Rosco, J.F., Jones, M.C., Pewsey, A., 2011. Skew-t distributions via the sinh-arcsinh transformation. Test 20, 630–652.

Thiele, T.N., 1889. Forlæsinger over Almindelig Iagttagelseslære: Sandsynlighedsregning og Mindste Kvadraters Methode. C.A. Reitzel, Copenhagen.

van Zwet, W., 1964. Convex Transformations of Random Variables. Mathematisch Centrum, Amsterdam.

Stat Papers
DOI 10.1007/s00362-012-0457-y

REGULAR ARTICLE

Measures of tail asymmetry for bivariate copulas

J. F. Rosco · Harry Joe

Received: 16 December 2011 / Revised: 25 April 2012

Abstract Three tail asymmetry measures for bivariate copulas are introduced using two different approaches—univariate skewness of a projection and distance between a copula and its survival/reflected copula. We compare the asymmetry measures based on certain desirable properties. Bounds for each measure are obtained and also copulas which attain these extreme values are identified. Two data examples show the amount of asymmetry that might be expected in practice.

Keywords Quantiles · Survival copula · Tail dependence · Univariate skewness

1 Introduction

Vine copulas, see Aas et al. (2009) and Kurowicka and Joe (2011), have been popular in recent years as a way to build multivariate copulas from bivariate marginal or conditional copulas. To help in deciding on appropriate bivariate copulas in the first level of the vine, diagnostics such as bivariate asymmetry measures, which can be applied to all pairs of variables, are useful. In this paper, we study several bivariate tail asymmetry measures, based on univariate skewness or distance between a copula and its reflected copula.

If multivariate data deviate from the multivariate normal copula, the usual forms of departure are in terms of the weight of the tails and/or asymmetry. When the asymmetry

J. F. Rosco (✉)
Department of Mathematics, Escuela Politécnica, University of Extremadura, Cáceres, Spain
e-mail: juanfranrn@unex.es

H. Joe
Department of Statistics, University of British Columbia, Vancouver, BC, Canada
e-mail: Harry.Joe@ubc.ca

Published online: 19 June 2012

cannot be neglected, the assumption of normality cannot be presumed and other models with more flexibility should be considered. Specifically, if tail asymmetry exists in the data, then the multivariate normal copula does not provide correct inferences on joint tail probabilities, and in quantitative risk analysis in finance and insurance, it is important to have models that provide good estimates of joint tail probabilities. Thus, copula families with a wide range of tail behaviour are useful for statistical modelling. Although the multivariate Gaussian and t copula families have wide dependence ranges, they are not appropriate when there is reflection or tail asymmetry.

The concept of asymmetry in the univariate context has been extended to the bivariate and multivariate distributions in many ways. Here we consider asymmetry in the copula dependence structure which can be separated from the univariate margins (by Sklar's theorem). For bivariate asymmetry, we focus on the class $\mathscr{C}$ of all bivariate copulas with Uniform(0, 1) margins. Given $C \in \mathscr{C}$ we define C_R as the survival copula, i.e., if $(U, V) \sim C$, then C_R is the copula associated with the distribution of $(1-U, 1-V)$ and given by the expression $C_R(u, v) = u + v - 1 + C(1-u, 1-v)$. The R subscript stands for 'reflection' since the univariate margins of C_R corresponds to the reflection of the univariate margins of C across the point $(1/2, 1/2)$. We say that a copula C is reflection symmetric if $C(u, v) = C_R(u, v)$ for all $u, v \in [0, 1]$; this concept is also called 'radial symmetry' in Nelsen (2006).

Regarding the dependence of multivariate models, there are the tail dependence parameters. Given $C \in \mathscr{C}$, we define the lower and upper tail parameters (see, for example Joe 1993), as the limits, when they exist,

$$\lambda_L = \lim_{u \downarrow 0} \frac{C(u, u)}{u}, \quad \lambda_U = \lim_{u \uparrow 1} \frac{C_R(1-u, 1-u)}{1-u},$$

respectively. And we say that C has lower tail dependence if $\lambda_L \in (0, 1]$ and has no lower tail dependence if $\lambda_L = 0$. Similarly with λ_U and upper tail dependence. A drawback of these parameters is that, as they are defined using limits, they cannot be estimated well from data.

More recently, in Hua and Joe (2011), the concept of tail order is introduced in a multivariate context to study a range of tail behaviour. The lower tail order is κ_L if $C(u, u) \sim \ell_L(u) u^{\kappa_L}$ as $u \to 0$ where $\ell_L(u)$ is a slowly varying function (such as a constant or a power of $-\log u$). If $C(u, u) = 0$ for all $0 < u < u_0$ for some positive u_0, then define $\kappa_L = \infty$. Similarly the upper tail κ_U is such that $C_R(1-u, 1-u) \sim \ell_U(u) u^{\kappa_U}$ as $u \to 0$. A property is that $\kappa_L \geq 1$ and $\kappa_U \geq 1$ with a smaller value corresponding to more dependence in the tail (more probability in the corner). Thus, the strongest tail dependence corresponds to $\kappa_L = 1$ or $\kappa_U = 1$. For a comonotonic (perfect positive dependence) tail, $\kappa_L = \kappa_U = 1$ and $\ell_L(u) = \ell_U(u) = 1$ for strongest tail dependence. For a countermonotonic (perfect negative dependence) tail, $\kappa_L = \kappa_U = \infty$ because there is no probability in the upper and lower corners. These tail orders also provide a simple condition to establish the direction of reflection asymmetry, namely: if $\kappa_L > \kappa_U$ ($\kappa_L < \kappa_U$) then C has reflection asymmetry skewed to the upper (lower) tail; and if $C(u, u) \sim \lambda_L u^\kappa$ and $C_R(u, u) \sim \lambda_U u^\kappa$ as $u \to 0$ with $\lambda_U > \lambda_L > 0$ ($\lambda_L > \lambda_U > 0$), then C has reflection asymmetry skewed to the upper (lower) tail.

Another tail asymmetry approach in Dobric et al. (2010) and Nikoloulopoulos et al. (2012), that has been applied to financial returns data, is based on the difference between the (Spearman or Pearson) correlations of the upper $[1-p, 1]^2$ tail and the lower $[0, p]^2$ tail, where $0 < p \le 1/2$. Specifically, if $(U, V) \sim C$, the difference of conditional correlations is

$$\mathrm{Corr}\,(U, V|U > 1-p, V > 1-p) - \mathrm{Corr}\,(U, V|U < p, V < p)\,. \tag{1}$$

A summary of the remainder of this paper is the following. Section 2 identifies desirable properties for measures of tail asymmetry. Section 3 analyzes three measures of tail asymmetry; two based on univariate skewness and one based on a distance measure of C and its reflected/survival copula C_R. Dehgani et al. (2011) study measures based on the L_p distance, and we overlap with them with the measure based on the L_∞ distance. The results show that most tail asymmetry tends to occur at intermediate strength of positive or negative dependence and that the copulas attaining extreme tail asymmetry depend on the measure. Section 4 has two data examples that illustrate the amount of tail asymmetry that might be expected in practice. Section 5 concludes with some discussion.

2 Desirable properties for measures of bivariate tail asymmetry

In this section, we present a list of appealing conditions that a measure of tail asymmetry, ς, should satisfy. Given $C \in \mathscr{C}$ and $(U, V) \sim C$, let $C_P(u, v) = C(v, u)$ denote the distribution of the permutation (V, U).

Mimicking what has been done in Durante et al. (2010) for measures of bivariate nonexchangeability, we present five axioms that are reasonable for measures of tail asymmetry to satisfy:

(i) there exists $K \in \mathbb{R}$ such that, $|\varsigma(C)| \le K$ for every $C \in \mathscr{C}$;
(ii) $\varsigma(C) = 0$ if C is reflection symmetric;
(iii) $\varsigma(C) = -\varsigma(C_R)$ for every $C \in \mathscr{C}$;
(iv) $\varsigma(C) = \varsigma(C_P)$ for every $C \in \mathscr{C}$;
(v) if $C \in \mathscr{C}$ and $\{C_n\}_{n\in\mathbb{N}}$ is a sequence of copulas such that $C_n \to C$ uniformly, then $\varsigma(C_n) \to \varsigma(C)$.

Our axioms were developed independently of Dehgani et al. (2011). The stronger version of axiom (ii) with "if and only if" is considered there, but this statement would only be satisfied for measures that are distances between C and C_R, and we think that would be too restrictive for some applications. Also, Dehgani et al. (2011) do not account for the direction of the asymmetry as they only consider non-negative measures.

In the following sections we study three tail asymmetry measures; two of them using an approach based on univariate skewness as a measure of asymmetry and another based on the L_∞ distance. For the latter, we obtain results that are not given in Dehgani et al. (2011).

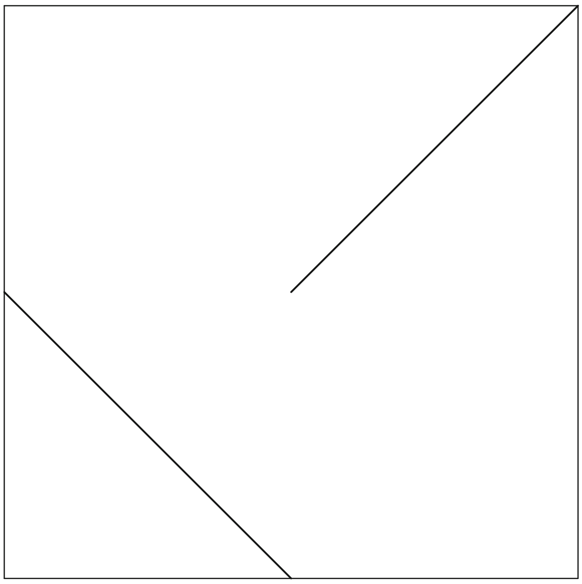


Fig. 1 Support of$C(\cdot; 1/2)$

3 Tail asymmetry measures

If $(U, V) \sim C$ for a reflection symmetric copula, then $(1 - U, 1 - V) \stackrel{d}{=} (U, V)$ and $U + V - 1 \stackrel{d}{=} (1 - U) + (1 - V) - 1 = 1 - U - V$ so that $U + V - 1$ is symmetric about 0. If C has tail asymmetry skewed to the upper (lower) tail, then $U + V - 1$ is right-skewed (left-skewed). There is much literature on univariate skewness measures and in the first two subsections, we use skewness measures based on moments and quantiles respectively. Note that any skewness measure (a function γ such that $\gamma(X) = -\gamma(-X)$ for a random variable X and $\gamma(X) = 0$ for X symmetric about 0) applied to $U + V - 1$ satisfies properties (ii) and (iii); property (iv) is also satisfied because $U + V - 1 = V + U - 1$.

We next introduce a family of singular copulas parameterized by $0 \le a \le 1$:

$$C(u, v; a) = \begin{cases} \max\{0, u + v - a\}, & 0 \le u, v \le a, \\ \min\{u, v\}, & \text{otherwise.} \end{cases} \tag{2}$$

If $(U, V) \sim C(\cdot; a)$, then a stochastic representation is

$$V = a - U \quad \text{for} \quad 0 \le U < a \quad \text{and} \quad V = U \quad \text{for} \quad a \le U \le 1.$$

In Fig. 1 we show an example of the support of a member of (2) with $a = 1/2$. We shall also use the corresponding survival copulas, $\{C_R(u, v; a) : 0 \le a \le 1\}$; the stochastic representation is

$$V = U \quad \text{for} \quad 0 \le U < 1 - a \quad \text{and} \quad V = 2 - a - U \quad \text{for} \quad 1 - a \le U \le 1.$$

For $C(\cdot; a)$, the lower tail dependence parameter is 0, $\kappa_L = \infty$ and the upper tail dependence parameter is 1, regardless of the value of a. For the tail asymmetry measure in (1), the value of the upper bound is 2 for $C(\cdot; a)$ if $a/2 < p \leq \min\{a, 1/2\}$ and the value of the lower bound is -2 for $C_R(\cdot; a)$ if $a/2 < p \leq \min\{a, 1/2\}$.

The reason for considering (2) as a main example is that we think this family of copulas attains the maximum or extreme values of different reflection asymmetry measures such as ς_1 in Sect. 3.1. However for particular asymmetry measures, there will also be other copulas that attain the extreme values.

Because comonotonic, countermonotonic and independence copulas are all reflection symmetric, we can expect that the copulas attaining the most reflection asymmetry have medium, positive or negative dependence. For reference of later results, we will use Blomqvist's β, defined as $4C(1/2, 1/2) - 1$, and Spearman's ρ_S, defined as $12\mathrm{E}(UV) - 3$, as measures of monotone association. For (2), $\rho_S = 1 - 2a^3$ and the value of β is 1 for $0 < a \leq 1/2$ and $3 - 4a$ for $1/2 < a < 1$.

3.1 A family of measures based on moments

We define the function

$$\begin{aligned} \mathscr{C} \times (1, \infty) &\longrightarrow (-1, 1) \\ (C, k) &\longmapsto \varsigma_1(C, k) = \varsigma_1(U, V; k) := \mathrm{E}[|U + V - 1|^k \mathrm{sign}(U + V - 1)], \end{aligned}$$

with $(U, V) \sim C$. The function ς_1 is well defined since the expectation of $|U + V - 1|^k$ exists whatever the copula C and the value of $k \in (1, \infty)$, and the case $k = 1$ is not considered because this leads to the constant function $\varsigma_1 = 0$. Since $|u + v - 1|^k \mathrm{sign}(u + v - 1) \in (-1, 1)$, we have that the range of ς_1 is $(-1, 1)$, thus ς_1 satisfies property (i). Because $\gamma(X) = |X|^k \mathrm{sign}(X)$ is a univariate skewness measure, then as indicated above, properties (ii), (iii) and (iv) are satisfied. It may be shown, using the dominated convergence theorem, that property (v) is also satisfied.

This function ς_1 is a generalized measure of asymmetry for bivariate copulas; it becomes an analogue of Fisher's coefficient of skewness when $k = 3$.

An outline of arguments that show that the family $\{C(\cdot; a)\}$ is extreme for ς_1 is given in the Appendix. In order to obtain the actual range for ς_1, we maximize ς_1 over (2). We only consider this maximization problem since the reflections of these copulas attain the minimum value. So, we have

$$\varsigma_1(C(u, v; a), k) = \begin{cases} -a(1 - a)^k - \int_a^{1/2} (2u - 1)^k \mathrm{d}u + \int_{1/2}^1 (2u - 1)^k \mathrm{d}u, & a < 1/2, \\ -a(1 - a)^k + \int_a^1 (2u - 1)^k \mathrm{d}u, & a \geq 1/2. \end{cases}$$

Hence

$$\frac{\partial \varsigma_1}{\partial a}(C(u, v; a), k) = -(1 - a)^k + ak(1 - a)^{k-1} - (2a - 1)^k. \tag{3}$$

This is a polynomial of order k; the roots cannot always be found analytically but can be obtained numerically. Table 1 shows the maximum values and where they are attained for different integer values of k. For instance, when $k = 2$ we obtain that the maximum is attained at $(4 + \sqrt{2})/7$ and when $k = 3$ the maximum is attained at $3/4$.

Table 1 Maximum value of $\varsigma_1[C(\cdot; a), k]$ for different values of k

k	2	3	4	5	6	7	8
max	0.0997	0.1055	0.0941	0.0815	0.0709	0.0623	0.0555
$a_{\max}$	0.7735	0.75	0.7358	0.7261	0.7190	0.7136	0.7094
β	−0.094	0	0.057	0.096	0.124	0.146	0.163
ρ_S	0.075	0.156	0.203	0.234	0.257	0.273	0.286

Also the point where the maximum is attained, $a_{\max}$, Blomqvist's β and Spearman's ρ_S are shown

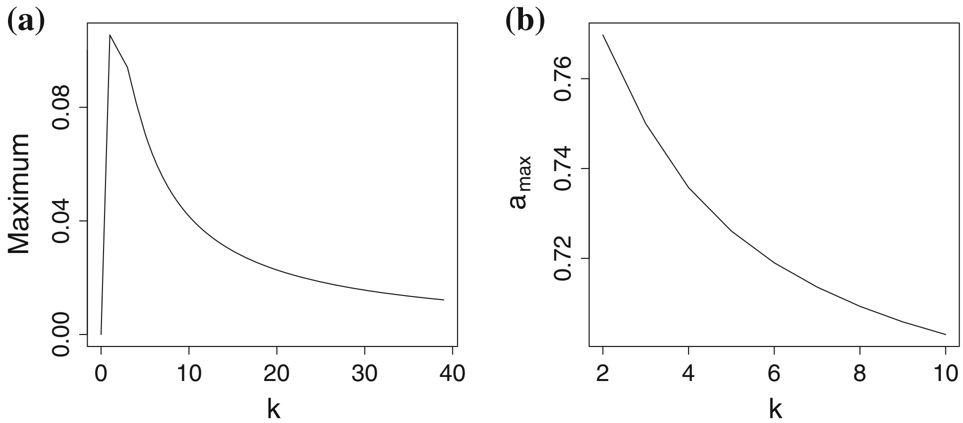


Fig. 2 Plots for **a** the maximum value of the asymmetry measure $\varsigma_1(\cdot, k)$ as a function of $1 < k \le 40$ and **b** the point at which the maximum is attained as a function of $1 < k \le 10$

In Fig. 2 we show a plot of the maxima of $\varsigma_1(C(\cdot; a), k)$ as a function of k in panel (a) and a plot of the points where those maxima are attained in panel (b). However, due to the flat shape of the function given by Eq. (3) and the use of numerical methods, we can only plot these points up to $k = 10$. Thus, the maximum value that can be attained for ς_1 increases with increasing k until k reaches $\simeq 2.61$, for which the maximum value is 0.107. Beyond that value ς_1 decreases with increasing k. For integer values of k, the maximum is attained when $k = 3$.

We are interested in the maximum values of $\varsigma_1(C(\cdot; a), k)$ when k is close to 1 and also when $k \to \infty$. It is easy to show that for every $\varepsilon > 0$, Eq. (3) decreases to 0 when $a = 2/3 - \varepsilon$ and $k \to \infty$. Also, it increases to 0 when $a = 2/3 + \varepsilon$ and $k \to \infty$. Thus, for every $\varepsilon > 0$, the maximum of $\varsigma_1(C(\cdot; a), k \to \infty)$ has to be attained in $[2/3 - \varepsilon, 2/3 + \varepsilon]$. Hence, the limit, when $k \to \infty$, of the points where the maximum is attained is $2/3$.

When $k \to 1$ we can use the very same argument to conclude that the limit of the points where the maximum is attained is around 0.82. This approximation was obtained as the root, using numerical methods, of $a + (2a-1)\log(1-a) + (1-2a)\log(2a-1)$, which is the coefficient of k in the Taylor series for Eq. (3) around 1. Figure 3 portrays plots of the maximum value in panel (a) and the points where the maximum is attained in panel (b) for $k \in (1, 2)$.

For absolutely continuous copulas, one that is the most asymmetric family, from the point of view of tail order (it satisfies $\kappa_L = 1, \lambda_L = 1$ and $\kappa_U = 2$), is the two-parameter family BB2 (see Joe 1997) given by

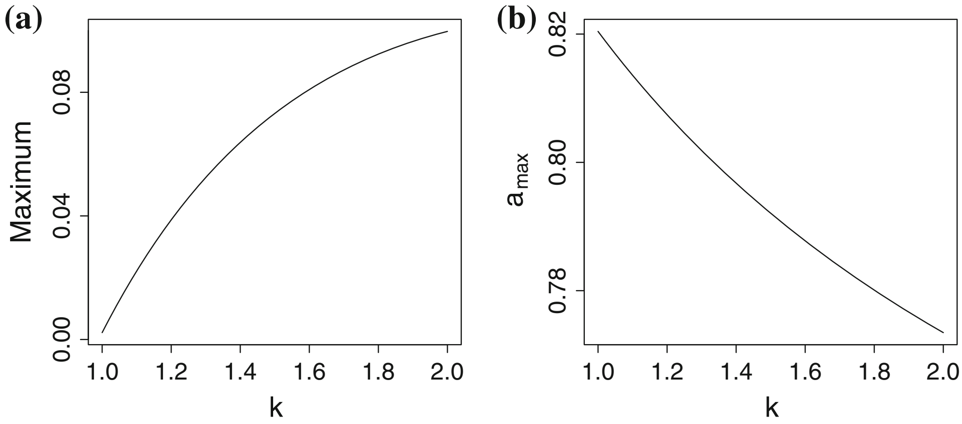


Fig. 3 Plots for **a** the maximum value of the asymmetry measure $\varsigma_1(\cdot, k)$ and **b** the points at which the maximum is attained as functions of $1 < k \leq 2$

$$C(u, v; \theta, \delta) = \left[1 + \delta^{-1} \log \left(e^{\delta(u^{-\theta}-1)} + e^{\delta(v^{-\theta}-1)} - 1\right)\right]^{-1/\theta},$$

where $\theta, \delta > 0$. For this family of copulas, we find numerically that the extreme asymmetry, as measured by ς_1 when $k = 3$, is -0.027 for the parameter values $\delta = 0.366$, $\theta = 1.198$. The value of Blomqvist's β for the copula with these parameters is 0.53.

Regarding the estimation of ς_1 for a set of data $(u_1, v_1), \ldots, (u_n, v_n)$, we can use any of the several moments estimators. Let $m_i = n^{-1} \sum_{j=1}^{n} (u_j + v_j - 1)^i$ denote the sample ith moment. When $k = 3$, we could just use the estimator $\hat{\varsigma}_1(3) = m_3$.

3.2 A family of measures based on quantiles

Another popular approach is to measure univariate asymmetry with quantiles. We define ς_2 as follows

$$\begin{aligned} &\mathscr{C} \times (0, 1/2) \longrightarrow [-1, 1] \\ &(C, p) \longmapsto \varsigma_2(C, p) := \frac{Q(1-p) - 2Q(1/2) + Q(p)}{Q(1-p) - Q(p)}, \end{aligned}$$

where Q denotes the quantile function of $U + V - 1$.

Since this measure is defined as a ratio where the denominator is always bigger than the numerator, property (i) is satisfied. As for property (v), this holds if the quantile function of $U + V - 1$ is continuous (and $Q(1-p) \neq Q(p)$).

The extreme value of $\varsigma_2(C, p) = 1$ can be attained for (2). For family (2), the cumulative distribution function of $U + V - 1$ satisfies $F_{U+V-1}(t) = 0$ for $t < a - 1$, $F_{U+V-1}(t) = a$ for $a - 1 \leq t < 2a - 1$ and $F_{U+V-1}(t) = (t + 1)/2$ for $2a - 1 \leq t \leq 1$. Hence for $0 < p < 1 - a$ and $a \geq 1/2$, the quantiles of $U + V - 1$ are $Q(p) = Q(1/2) = a - 1$ and $2a - 1 \leq Q(1 - p) < 1$, and $\varsigma_2(C(\cdot; a), p) = 1$. Also, -1 can be attained for the corresponding survival copulas.

For the BB2 copula, the extreme asymmetry, -0.168, is attained for $p = 0.102$ and the parameter values $\delta = 1.037$ and $\theta = 0.700$. The value of Blomqvist's β for the copula with these parameters is 0.49.

Regarding the sample version of this measure, $\hat{\varsigma}_2$, we have several choices when it comes to estimate quantiles. A comparison between estimators implemented in statistical packages can be found in Hyndman and Fan (1996). Another comparison between quantile estimators is made in Dielman et al. (1994).

3.3 A measure based on the distance between C and C_R

In Klement and Mesiar (2006) and Nelsen (2007), a measure of nonexchangeability is introduced as $\sup\{|C(u, v) - C(v, u)|\}$, where the supremum is taken over $(u, v) \in [0, 1]^2$. Following this approach, Dehgani et al. (2011) studied measures of (reflection) asymmetry based on the L_p distances. For the boundary case, $p = +\infty$, define the function ς_3 as

$$\varsigma_3(C) := \sup_{(u,v)\in[0,1]^2} \{|C(u, v) - C_R(u, v)|\}.$$

It is clear that $\varsigma_3 \geq 0$ since it is defined as the supremum of non-negative values, and it is bounded above by 2. It is easy to show that property (ii) is satisfied. Property (iii) cannot be satisfied by ς_3 since it only has non-negative values. Property (iv) is satisfied as $\{|C(u, v) - C_R(u, v)| : u, v \in [0, 1]\} = \{|C(v, u) - C_R(v, u)| : u, v \in [0, 1]\}$. Also, since $\varsigma_3 : (\mathscr{C}, \|\cdot\|_\infty) \longrightarrow (A \subset \mathbb{R}, |\cdot|)$ is a continuous function, we have that property (v) is satisfied.

It is shown in Dehgani et al. (2011) that $\varsigma_3 \leq 1/3$. More generally, we have the following result,

$$\sup_{(u,v)\in[0,1]^2} \{|C(u, v) - C_R(u, v)|\}$$
$$\leq \min_{(u,v)\in[0,1]^2} \left\{u, v, 1-u, 1-v, \left|\frac{1}{2} - u\right| + \left|\frac{1}{2} - v\right|\right\} \leq \frac{1}{3},$$

and the inequality is best-possible since $|C(1/3, 1/3; 2/3) - C_R(1/3, 1/3; 2/3)| = 1/3$. So we have $A = [0, 1/3]$. The proof of this inequality shall be given later as a particular case of a more general scenario.

Before we get to that, we emphasize that there are many copulas that attain the maximum value of ς_3. As shown in Dehgani et al. (2011), the maximum value of 1/3 can be attained by some copulas with support on three subsquares, each with total probability of 1/3:

(a) support on $[0, 1/3]^2$, $[1/3, 2/3] \times [2/3, 1]$, $[2/3, 1] \times [1/3, 2/3]$;
(b) support on $[0, 1/3] \times [1/3, 2/3]$, $[1/3, 2/3] \times [2/3, 1]$, $[2/3, 1] \times [0, 1/3]$;
(c) support on $[0, 1/3] \times [2/3, 1]$, $[1/3, 2/3] \times [0, 1/3]$, $[2/3, 1] \times [1/3, 2/3]$;
(d) support on $[0, 1/3] \times [1/3, 2/3]$, $[1/3, 2/3] \times [0, 1/3]$, $[2/3, 1]^2$.

All four of these could have lower and upper tail orders of infinity. For instance, the copulas in (b) and (c) have these tail orders for sure. Hence, ς_3 is not a strongly discriminating measure of tail asymmetry. For a similar reason the measure proposed in Klement and Mesiar (2006) and Nelsen (2007) is not strongly discriminating.

We return now to the upper bound for ς_3. Let $0 \le C(1/2, 1/2) = \alpha \le 1/2$ be a fixed value, then we have

$$\sup_{(u,v)\in[0,1]^2} \{|C(u,v) - C_R(u,v)|\} \le \min_{(u,v)\in[0,1]^2} \left\{u, v, 1-u, 1-v, \left|\frac{1}{2}-u\right| + \left|\frac{1}{2}-v\right|, \alpha^*_{(u,v)}, \left|\frac{1}{2}-\alpha^*_{(u,v)}\right| + \left|\frac{1}{2}-u\right|, \left|\frac{1}{2}-\alpha^*_{(u,v)}\right| + \left|\frac{1}{2}-v\right|\right\},$$

where $\alpha^*_{(u,v)} = 1/4 + (\alpha - 1/4)\,\mathrm{sign}\{(u-1/2)(v-1/2)\}$.

Let $C \in \mathscr{C}$ be a copula and $\alpha = C(1/2, 1/2)$. Then a straightforward calculation leads to $C_R(1/2, 1/2) = \alpha$ and by definition, for $(U, V) \sim C$, $C(u, v) = P(U \le u, V \le v)$ and $C_R(u, v) = P(U > 1-u, V > 1-v)$. We shall use the well-known inequalities

$$\max\{0, P(A) + P(B) - 1\} \le P(A \cap B) \le \min\{P(A), P(B)\}, \tag{4}$$

for different A, B; a special case is:

$$\max\{0, u+v-1\} \le P(U \le u, V \le v) \le \min\{u, v\}. \tag{5}$$

We divide the proof for the following four cases.

1. $u \le 1/2, v \le 1/2$
   We have $C(u,v) = \alpha - P(U \le 1/2, v \le V \le 1/2) - P(u \le U \le 1/2, V \le v)$ (see the left panel of Fig. 4). In order to obtain an upper bound for $C(u, v)$ we must find lower bounds for the probabilities. Using (4), we obtain

   $$P(U \le 1/2, v \le V \le 1/2) \ge 0, \quad P(u \le U \le 1/2, V \le v) \ge 0,$$

   thus, combining with (5), we have $C(u, v) \le \min\{u, v, \alpha\}$.
   Next we want to obtain a lower bound for $C_R(u, v) = \alpha - P(1/2 \le U \le 1, 1/2 \le V \le 1-v) - P(1/2 \le U \le 1-u, 1-v \le V \le 1)$, for which we must find upper bounds for the two probabilities. Using (4), we have

   $$P(1/2 \le U \le 1, 1/2 \le V \le 1-v) \le \min\{1/2, 1/2-v\} = 1/2 - v,$$

   $$P(1/2 \le U \le 1-u, 1-v \le V \le 1) \le \min\{1/2-u, v\},$$

   and, with (5), we have $C_R(u, v) \ge \max\{0, \alpha - 1/2, \alpha + u + v - 1\} = \max\{0, \alpha + u + v - 1\}$. Combining all of this we obtain

   $$C(u,v) - C_R(u,v) \le \min\{u, v, 1-u-v, \alpha, 1-u-\alpha, 1-v-\alpha\}. \tag{6}$$

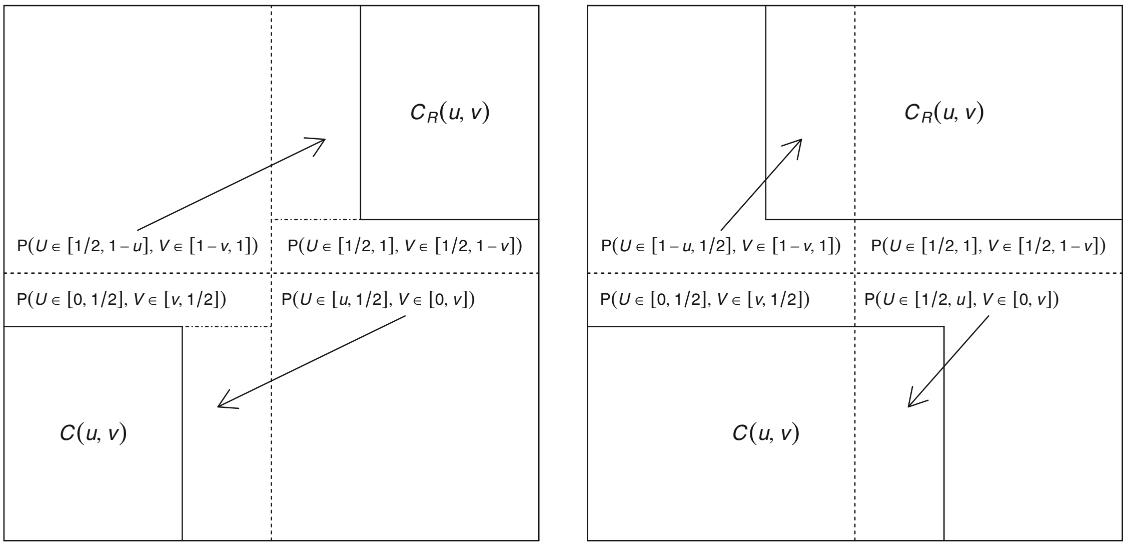


Fig. 4 Cases 1 and 3 of the proof

2. $u > 1/2, v > 1/2$
   A similar analysis to that above can be used. It turns out that the result is obtained by replacing $u \to 1-u$ and $v \to 1-v$ in the preceding case. That is,

$$C(u,v) - C_R(u,v) \le \min\{1-u, 1-v, u+v-1, \alpha, u-\alpha, v-\alpha\}. \quad (7)$$

3. $u > 1/2, v \le 1/2$
   We have $C(u,v) = \alpha - P(U \le 1/2, v \le V \le 1/2) + P(1/2 \le U \le u, V \le v)$ (see the right panel of Fig. 4). Next we find an upper bound for $C(u,v)$;

$$P(U \le 1/2, v \le V \le 1/2) \ge 0, \quad P(1/2 \le U \le u, V \le v) \le \min\{u - 1/2, v\}.$$

   So, combining with (5), we have $C(u,v) \le \min\{v, \alpha + u - 1/2\}$.
   For the survival copula we have $C_R(u,v) = \alpha - P(1/2 \le U \le 1, 1/2 \le V \le 1-v) + P(1-u \le U \le 1/2, 1-v \le V \le 1)$. Bounds for the probabilities are

$$P(1/2 \le U \le 1, 1/2 \le V \le 1-v) \le 1/2 - v,$$

$$P(1-u \le U \le 1/2, 1-v \le V \le 1) \ge 0.$$

   So, with (5), we have $C_R(u,v) \ge \max\{0, u+v-1, \alpha + v - 1/2\}$. Finally, we obtain

$$C(u,v) - C_R(u,v) \le \min\{1-u, v, u-v, 1/2-\alpha, \alpha+u-1/2, \alpha-v+1/2\}. \quad (8)$$

4. $u \le 1/2, v > 1/2$
   This case can be resolved by interchanging $u \leftrightarrow v$ in the preceding case. Hence, we obtain

$$C(u, v) - C_R(u, v) \le \min\{1 - v, u, v - u, 1/2 - \alpha, \alpha + v - 1/2, \alpha - u + 1/2\}. \tag{9}$$

We end the proof by noting that

$$\left|\frac{1}{2} - u\right| + \left|\frac{1}{2} - v\right| = \begin{cases} 1 - u - v, & u \le 1/2, v \le 1/2, \\ u - v, & u > 1/2, v \le 1/2, \\ v - u, & u \le 1/2, v > 1/2, \\ u + v - 1, & u > 1/2, v > 1/2, \end{cases}$$

and

$$\alpha^*_{(u,v)} = \begin{cases} \alpha, & u \le 1/2, v \le 1/2 \text{ or } u > 1/2, v \ge 1/2, \\ 1/2 - \alpha, & u \le 1/2, v > 1/2 \text{ or } u > 1/2, v < 1/2. \end{cases}$$

Using the last upper bound result in Fig. 5 we display the maximum values that ς_3 can attain when $C(1/2, 1/2) = \alpha$ is fixed. We see that we can differentiate between four cases:

i. $0 \le \alpha \le 1/6$; it follows that

$$a = 1/2 - \alpha, \quad u = 1 - v = 3/4 - \alpha/2, \quad |C(u, v) - C_R(u, v)| = \alpha/2 + 1/4;$$

ii. $1/6 \le \alpha \le 1/4$; it follows that

$$a = 1/2 - \alpha, \quad u = 2v = 1 - 2\alpha, \quad |C(u, v) - C_R(u, v)| = 1/2 - \alpha;$$

iii. $1/4 \le \alpha \le 1/3$; it follows that

$$a = \alpha, \quad u = 1 - 2v = 1 - 2\alpha, \quad |C(u, v) - C_R(u, v)| = \alpha;$$

iv. $1/3 \le \alpha \le 1/2$; it follows that

$$a = \alpha, \quad u = v = (1 - \alpha)/2, \quad |C(u, v) - C_R(u, v)| = (1 - \alpha)/2.$$

The importance of this result is that it shows that maximum asymmetry occurs for intermediate dependence—Blomqvist's beta is $\beta = 4\alpha - 1$, so extremes of ς_3 occur for $\beta = \pm 1/3$.

For the cases that attain the maximum of $\sup_{u,v} |C(u, v) - C_R(u, v)|$ given $\alpha = C(1/2, 1/2)$, some of the terms in (6–9) are equal. Matching the four cases above to the four cases of the proof, the following details are obtained.

i. $0 \le \alpha \le 1/6$
We have $u \ge 1/2$ and $v \le 1/2$ and hence we obtain $v = 1 - u = \alpha + u - 1/2 = \alpha - v + 1/2 = \alpha/2 + 1/4$ and $u - v = 1/2 - \alpha \ge \alpha/2 + 1/4$.

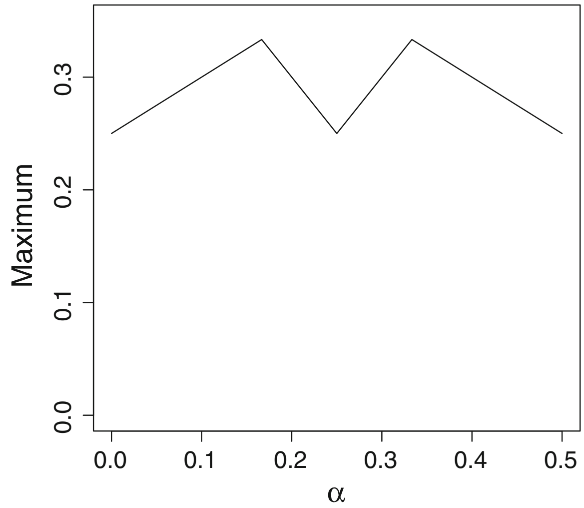


Fig. 5 Maximum value of ς_3 as a function of $\alpha = C(1/2, 1/2)$: $\alpha/2 + 1/4$ for $0 \le \alpha \le 1/6$; $1/2 - \alpha$ for $1/6 \le \alpha \le 1/4$; α for $1/4 \le \alpha \le 1/3$; $(1-\alpha)/2$ for $1/3 \le \alpha \le 1/2$

ii. $1/6 \le \alpha \le 1/4$
We have $u \ge 1/2$ and $v \le 1/2$ and we obtain $v = u - v = \alpha + u - 1/2 = 1/2 - \alpha$ and $\alpha - v + 1/2 = 1 - u > 1/2 - \alpha$.
iii. $1/4 \le \alpha \le 1/3$
We have $u \le 1/2$ and $v \le 1/2$ and we obtain $v = 1 - u - \alpha = 1 - u - v = \alpha$ and $u = 1 - u - v \ge \alpha$.
iv. $1/3 \le \alpha \le 1/2$
We have $u \le 1/2$ and $v \le 1/2$ and we obtain $u = v = 1 - u - \alpha = 1 - v - \alpha$ and $1 - u - v = \alpha \ge (1-\alpha)/2$.

As a comparison of a maximum value of 1/3 for ς_3 over all bivariate copulas, for the continuous bivariate BB2 copula we have that ς_3 has a maximum value of 0.092 when $\delta = 0.226$ and $\theta = 2.049$. The value of Blomqvist's β is 0.67.

4 Examples with real data

In order to illustrate the use of, and possible issues with, the sample versions for the tail asymmetry measures introduced above, we present two examples with real data; one considering insurance data and the other considering stock exchange data. For $\hat{\varsigma}_1(\cdot, k)$ we chose $k = 3$ and as the estimator of the third moment we used $n^{-1}\sum_{i=1}^{n}(x_i - \bar{x})^3$ where n is the sample size and $\bar{x} = n^{-1}\sum_{i=1}^{n} x_i$. Regarding $\hat{\varsigma}_2(\cdot, p)$, we chose $p = 0.05$ and used the type 8 method of the `quantile` function provided by the statistical software R for estimating the quantiles. The choice $k = 3$ is based on the fact that in the univariate case the third moment is regarded as a measure of asymmetry. The choice $p = 0.05$ was made instead of the more natural $p = 0.25$ because with the former value ς_2 gives more importance to the tails of the distribution. Regarding the estimators: the first was chosen because of its simplicity; the second, based upon the results given in Hyndman and Fan (1996). Regarding $\hat{\varsigma}_3$ as an estimator of ς_3, we

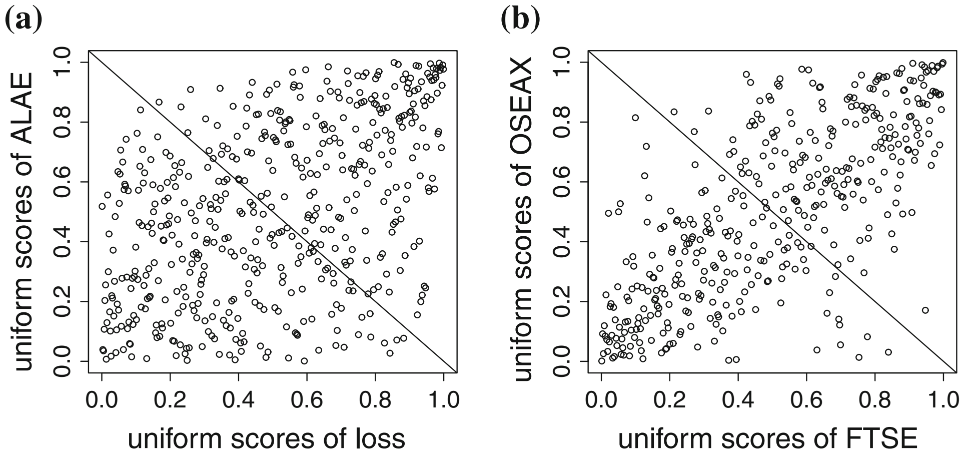


Fig. 6 Scatter plots of uniform scores of **a** the loss-ALAE (loss values were jittered to avoid ties in the plot) and **b** the FTSE-OSEAX data sets, together with the line $u + v - 1 = 0$

estimate the empirical bivariate copula as an initial step using the function `mecdf` from the homonym package available for the statistical software R.

For $\hat{\varsigma}_1$ and $\hat{\varsigma}_2$, the jackknife can be used to obtain the standard errors of the estimates. The asymptotics for $\hat{\varsigma}_3$ are not straightforward because maxima of a Gaussian process (the limit of empirical copula processes—see Fermanian et al. 2004) are involved. To know whether $\hat{\varsigma}_3$ is large or not, we compare it with its sampling distribution from replicated data sets of size n from three different reflection symmetric bivariate copulas (bivariate normal, Plackett and Frank copulas with the same Blomqvist's β value as that for the actual data). It turned out that the sampling distributions were very similar for the different choices of reflection symmetric bivariate copulas for a fixed β; the upper quantiles have a slow decreasing trend as $|\beta|$ increases.

4.1 Insurance data on losses and ALAEs

In order to illustrate tail asymmetry, our first set of data comprises 500 general liability claims which are a random subset of the uncensored cases of a data set which first appeared in Frees and Valdez (1998). Each claim consists of an indemnity payment (the loss) and an allocated loss adjustment expense (ALAE). A scatter plot of the data set is shown in Fig. 6. For these two variables, Spearman's ρ_S is 0.46. We can see that there is more probability mass in the right upper corner, suggesting positive tail asymmetry. This observation is backed up by the estimation of the measures shown in Table 2. In terms of magnitude, these skewness measures are in absolute value roughly half of the maximum absolute value for the BB2 copula family, but nowhere near the values for copula (2).

4.2 FTSE and OSEAX returns data

The second data set consists of 407 stock returns from the London FTSE index and the Oslo OSEAX index from 2007 to 2008. In order to eliminate possible serial dependence in the absolute returns, we applied a GARCH filter. For these two GARCH-filtered

Table 2 Loss-ALAE data: estimates for the different asymmetry measures

	ς_1	ς_2	ς_3
Estimate	0.014	0.089	0.058
Standard error	0.004	0.029	NA
Significantly non-zero	Yes	Yes	Yes (p-value$\ll$ 0.01)

For ς_1 and ς_2, standard errors were obtained using the jackknife, and the corresponding asymptotic 95 % confidence intervals do not include 0

Table 3 FTSE-OSEAX data: estimates of the different measures together with standard errors obtained using the jackknife

	ς_1	ς_2	ς_3
Estimate	−0.001	−0.019	0.027
Standard error	0.004	0.038	NA
Significantly non-zero	No	No	No (p-value$\approx$ 0.5)

For ς_1 and ς_2, 0 is inside the asymptotic 95 % confidence intervals (obtained using the jackknife)

indices, Spearman's ρ is 0.72. There is no obvious tail asymmetry but $\hat{\varsigma}_1$ and $\hat{\varsigma}_2$ are slightly negative (Table 3). The asymmetry is considerable lower in magnitude than that in the loss-ALAE data set and the asymmetry statistics are not significantly different from 0. There is theory, see for example Longin and Solnik (2001), that suggests more dependence in the lower tail during economic downturns, but perhaps tail-weighted dependence/asymmetry measures are needed to better detect the asymmetries.

5 Discussion

All three measures, ς_1, ς_2 and ς_3, have a finite range which is an advantage since a measure with arbitrarily large values can be hard to interpret. Related to this advantage, for each of the measures more than one copula attaining the maximum value exists. Thus, very different copulas can have the same value for each one of the measures.

A further advantage which is shared by ς_1 and ς_2 is that they provide not just one measure but a family of measures parameterized by a real number. Thus, we can choose between the various measures to suit the application. Moreover ς_1 and ς_2 measure not only the degree of asymmetry but also the direction of the asymmetry, whereas ς_3 only indicates the degree of asymmetry. A disadvantage of ς_3 is that there are copulas with upper and lower tail order of infinity which attain the maximum value, i.e., some copulas considered extreme according to ς_3 are not extreme according to other tail asymmetry concepts.

When it comes to estimate the three measures from a sample, ς_1 and ς_2 can be estimated straightforwardly using sample moments and sample quantiles, respectively, with the jackknife or bootstrap applied to obtain standard errors. For ς_3, the estimate $\hat{\varsigma}_3$ is available based on the empirical copula. The asymptotic behaviour of $\hat{\varsigma}_3$ is difficult to study as it would involve the limiting Gaussian process of the empirical copula. As it is unclear whether it will be more useful than the other two proposed measures, for the data examples we only compared $\hat{\varsigma}_3$ with the sampling distributions for some reflection symmetric copulas (for which the possible values will be positive and any fixed upper quantile decreasing to 0 as the sample size increases).

To ensure that $\varsigma_3(C) = 0$ it is not only sufficient but also necessary that C be reflection symmetric. This is not the case, however, for ς_1 and ς_2. For instance, in the univariate distributional case there exist asymmetric distributions with zero third central moment (see Ord 1968).

These advantages and disadvantages suggest that it is sensible to use more than one measure of tail asymmetry as each has quite different properties and, as we have seen, none of the three proposed measures satisfy all of the desirable properties stated in Sect. 1.

In this paper, we have further studied tail or reflection asymmetry, with the distance measure similar to Klement and Mesiar (2006) and Nelsen (2007). There are analogous measures for bivariate nonexchangeability based on univariate skewness measures applied to $U - V$ with $(U, V) \sim C$. Similar to tail asymmetry, having several bivariate nonexchangeability measures would be useful during data analysis. Our experience with real data is that tail asymmetry is more pronounced in bivariate uniform or normal scores plots than is nonexchangeability.

Acknowledgements The first author was supported by grant PRE08028 from the European Social Fund through the Junta de Extremadura and the second author was supported by an NSERC Discovery Grant. Thanks to Pavel Krupskii for the suggestion of comparing $\hat{\varsigma}_3$ for simulated distributions from reflection symmetric copulas. We also thank the referees for providing additional references on the topic.

Appendix: Outline of arguments supporting the extremeness of the copula family $\{C(\cdot, \cdot; a)\}$ for ς_1

We consider the family $\varsigma_1(C, k) = \varsigma_1(U, V; k)$ for $k > 1$ as a special case of $\mathrm{E}\,[\Psi(|U + V - 1|)\,\mathrm{sign}(U + V - 1)]$ where Ψ has domain [0, 1], is convex, increasing and satisfies $\Psi(0) = 0$. That is, the function x^k on [0, 1] for $k > 1$ satisfies these conditions.

We need the definitions of majorization for vectors of size 2 and supperadditivity. See Marshall et al. (2011) for much more on the theory of majorization. A convex, increasing function Ψ on [0, 1] satisfying $\Psi(0) = 0$ is superadditive.

Definition: Majorization for 2-vectors and a characterization. The vector (x_1, x_2) is majorized by (y_1, y_2) if $y_1 \le x_1 \le x_2 \le y_2$ are real numbers satisfying $x_1 + x_2 = y_1 + y_2$. A characterization is that if (x_1, x_2) is majorized by (y_1, y_2), then for every convex function f in the interval (a, b) we have $f(x_1) + f(x_2) \le f(y_1) + f(y_2)$.

Definition: Superadditivity. A real function f is superadditive if $f(x + y) \ge f(x) + f(y)$ for every $x, y, x + y$ in the domain of f.

Let $0 \le u_1 < u_2 \le 1$ and $0 \le v_1 < v_2 \le 1$ be such that (u_1, v_1), (u_1, v_2), (u_2, v_1), (u_2, v_2) are the four vertices of a rectangle in $[0, 1]^2$. Let Ψ denote a convex increasing function on [0, 1] with $\Psi(0) = 0$, and let $g(u, v) = \Psi(|u + v - 1|)\,\mathrm{sign}(u + v - 1)$.

We compare the sum $g(u_1, v_1) + g(u_2, v_2)$ over the two vertices on the main diagonal with the sum $g(u_1, v_2) + g(u_2, v_1)$ over the two vertices on the negative diagonal.

According to the position of the points relative to the line $u+v-1 = 0$, we consider the following six cases:

1. $u_i + v_j \geq 1$, $1 \leq i, j \leq 2$ so at least three vertices are strictly above the line $u+v-1 = 0$. Hence $(u_2+v_2-1)+(u_1+v_1-1) = (u_1+v_2-1)+(u_2+v_1-1)$ and $u_1+v_1-1 \leq u_1+v_2-1$, $u_2+v_1-1 \leq u_2+v_2-1$, and from majorization,

$$\Psi(u_1 + v_1 - 1) + \Psi(u_2 + v_2 - 1) \geq \Psi(u_1 + v_2 - 1) + \Psi(u_2 + v_1 - 1).$$

This means that with four points on or above the line $u + v - 1 = 0$ the sum is larger over the vertices on the main diagonal.

2. Only $u_1 + v_1 < 1$ and other three pairs are on or above the line $u + v - 1 = 0$. Hence $(1 - u_1 - v_1) + (u_1 + v_2 - 1) + (u_2 + v_1 - 1) = u_2 + v_2 - 1$, and using superadditivity,

$$\Psi(u_2 + v_2 - 1) - \Psi(1 - u_1 - v_1) \geq \Psi(u_1 + v_2 - 1) + \Psi(u_2 + v_1 - 1).$$

This means that with only one point below the line $u+v-1 = 0$ the sum is larger over the vertices on the main diagonal.

3. $u_1 + v_1 < 1, u_2 + v_1 \leq 1, u_1 + v_2 \geq 1$ and $u_2 + v_2 > 1$ The inequality depends on whether $u_1 + u_2 + v_1 + v_2 > 2$ or not. If the inequality holds, then $u_2+v_2-1 > 1-u_1-v_1$ and $u_1+v_2-1 > 1-u_2-v_1$, and using majorization,

$$\Psi(u_2 + v_2 - 1) - \Psi(1 - u_1 - v_1) \geq \Psi(u_1 + v_2 - 1) - \Psi(1 - u_2 - v_1).$$

If the inequality does not hold, then

$$\Psi(u_2 + v_2 - 1) - \Psi(1 - u_1 - v_1) \leq \Psi(u_1 + v_2 - 1) - \Psi(1 - u_2 - v_1).$$

This result means that in order to obtain a greater value of Ψ: the sum is larger over the vertices on the main diagonal if $u_1 + u_2 + v_1 + v_2 > 2$; the sum is larger over the vertices on the negative diagonal if $u_1 + u_2 + v_1 + v_2 < 2$; and the two sums are equal if $u_1 + u_2 + v_1 + v_2 = 2$.

4. $u_1 + v_1 < 1, u_2 + v_1 \geq 1, u_1 + v_2 \leq 1$ and $u_2 + v_2 > 1$. In parallel to the preceding case, if $u_1 + u_2 + v_1 + v_2 > 2$ then

$$\Psi(u_2 + v_2 - 1) - \Psi(1 - u_1 - v_1) \geq \Psi(u_2 + v_1 - 1) - \Psi(1 - u_1 - v_2)$$

and otherwise

$$\Psi(u_2 + v_2 - 1) - \Psi(1 - u_1 - v_1) \leq \Psi(u_2 + v_1 - 1) - \Psi(1 - u_1 - v_2).$$

5. Only $u_2 + v_2 > 1$ and other three vertices are on or below the line $u + v - 1 = 0$. Then $(1 - u_1 - v_2) + (1 - u_2 - v_1) + (u_2 + v_2 - 1) = 1 - u_1 - v_1$ and using superadditivity

$$\Psi(u_2 + v_2 - 1) - \Psi(1 - u_1 - v_1) \le -\Psi(1 - u_1 - v_2) - \Psi(1 - u_2 - v_1).$$

This means that with three points on or below the line $u + v - 1 = 0$ the sum is larger over the vertices on the negative diagonal.

6. $u_i + v_j \le 1, 1 \le i, j \le 2$; so at least three vertices are strictly below the line $u + v - 1 = 0$. Then $1 - u_2 - v_2 \le 1 - u_1 - v_2$, $1 - u_2 - v_1 \le 1 - u_1 - v_1$ and $(1 - u_2 - v_2) + (1 - u_1 - v_1) = (1 - u_1 - v_2) + (1 - u_2 - v_1)$. Hence using majorization,

$$-\Psi(1 - u_2 - v_2) - \Psi(1 - u_1 - v_1) \le -\Psi(1 - u_1 - v_2) - \Psi(1 - u_2 - v_1).$$

This means that if the four points are on or below the line $u + v - 1 = 0$ the sum is larger over the vertices on the negative diagonal.

The above cases indicate where mass or density can be put for uniform random variables (U, V) in order to make $\mathrm{E}[\Psi(|U + V - 1|)\,\mathrm{sign}(U + V - 1)]$ larger. If there is positive probability in a small rectangle about (u_1, v_1), (u_1, v_2), (u_2, v_1), (u_2, v_2) with $0 < u_1 < u_2 < 1$ and $0 < v_1 < v_2 < 1$ then a (small) constant density $\epsilon > 0$ can be added (subtracted) from (u_1, v_1), (u_2, v_2) and subtracted (added) to (u_1, v_2), (u_2, v_1) depending on one of the six cases. Note that such a shift of density would not affect the uniform margins.

For further analysis, consider the discretization of a copula with support on the n^2 points $(i/(n + 1), j/(n + 1))$ where $n \ge 2$ is an integer and $i, j \in \{1, \dots, n\}$. So below U, V are uniform on the set $\{1/(n + 1), \dots, n/(n + 1)\}$. Considering the four points $(i_1, j_1)/(n + 1)$, $(i_1, j_2)/(n + 1)$, $(i_2, j_1)/(n + 1)$ and $(i_2, j_2)/(n + 1)$ we can distinguish between the following cases:

a. There is positive probability on all four points. Then we can shift mass to the two points on one of the diagonals; the preferred diagonal depending on which of the six cases.
b. There is positive probability on three of the four points. In this case we have to be careful with the shifting of mass in order to preserve the uniform margins. For instance, suppose that there is positive probability on the points $(i_1, j_2)/(n + 1)$, $(i_2, j_1)/(n + 1)$ and $(i_2, j_2)/(n + 1)$. Then we can move a constant mass from the points $(i_1, j_2)/(n + 1)$, $(i_2, j_1)/(n + 1)$ to the points $(i_1, j_1)/(n + 1)$, $(i_2, j_2)/(n + 1)$ if this increases $\mathrm{E}[\Psi(|U + V - 1|)\,\mathrm{sign}(U + V - 1)]$. We cannot move mass from the point $(i_2, j_2)/(n + 1)$ to the point $(i_1, j_2)/(n + 1)$ because the uniformity of the margins would be lost. The other cases are analogous.
c. There is positive probability on two of the four points. We consider only the case when the two points are at the same diagonal, otherwise we cannot shift the mass. Again we can move mass to the other diagonal if this increases $\mathrm{E}[\Psi(|U + V - 1|)\,\mathrm{sign}(U + V - 1)]$.

For discretized uniform variables, the above suggests that, for some integer b with $2 < b < n$, the support points can be shifted to $(i/[n + 1], i/[n + 1])$ for $i = b, \dots, n$ and $(j/[n + 1], (b - j)/[n + 1])$ for $j = 1, \dots, b - 1$ with mass n^{-1} at each of these n points. But we show that for this family of distributions, mass can be shifted to $(i/[n + 1], i/[n + 1])$ for $i = b + 1, \dots, n$ and $(j/[n + 1], (b + 1 - j)/[n + 1])$

for $j = 1, \ldots, b$ unless $b > 2[n+1]/3$. This coincides with what is seen in Table 1 for the subset of convex functions $\Psi(x) = x^k$ for $k > 1$.

Suppose the support is as indicated above and let $s = b$. Consider the points $(u_1, v_1) = (1/[n+1], (s-1)/[n+1])$ and $(u_2, v_2) = (s/[n+1], s/[n+1])$ as points on a main diagonal. Note that $u_1 + v_1 = s/[n+1] < 1$, $u_1 + v_2 = (s+1)/[n+1] < 1$, and $u_2 + v_1 = (2s-1)/[n+1]$ or $u_2 + v_2 = 2s/[n+1]$ could be above or below 1. Based on cases 3–6, $\mathrm{E}[\Psi(|U + V - 1|)\,\mathrm{sign}(U + V - 1)]$ can be increased by moving the mass n^{-1} to (u_1, v_2) and (u_2, v_1) if $u_1 + u_2 + v_1 + v_2 = 3s/[n+1] \leq 2$ or $s = b \leq 2[n+1]/3$.

If the first move to $(1/[n+1], s/[n+1])$ and $(s/[n+1], (s-1)/[n+1])$ is made, then further moves can be made to shift mass to the line $(i+j)/[n+1] = (s+1)/[n+1]$. The next shift is from $(u_1', v_1') = (2/[n+1], (s-2)/[n+1])$ and $(u_2', v_2') = (s/[n+1], (s-1)/[n+1])$ to $(2/[n+1], (s-1)/[n+1])$ and $(s/[n+1], (s-2)/[n+1])$. Continue like this to locate mass on $(i/[n+1], (s+1-i)/[n+1])$ for $i = 1, \ldots, s-2$, and $(s/[n+1], 2/[n+1])$. Then the final move is from $(u_1'', v_1'') = ((s-1)/[n+1], 1/[n+1])$ and $(u_2'', v_2'') = (s/[n+1], 2/[n+1])$ to $((s-1)/[n+1], 2/[n+1])$ and $(s/[n+1], 1/[n+1])$.

References

Aas K, Czado C, Frigessi A, Bakken H (2009) Pair-copula constructions of multiple dependence. Insur Math Econ 44:182–198

Dehgani A, Dolati A, Úbeda-Flores M (2011) Measures of radial asymmetry for bivariate random vectors. Stat Papers. doi:10.1007/s00362-011-0425-y

Dielman T, Lowry C, Pfaffenberger R (1994) A comparison of quantile estimators. Commun Stat B 23: 355–371

Dobric J, Frahm G, Schmid F (2010) Dependence of stock returns in bull and bear markets. Technical report

Durante F, Klement EP, Sempi C, Úbeda-Flores M (2010) Measures of nonexchangeability for bivariate random vectors. Stat Papers 51:687–699

Fermanian J, Radulovic D, Wegkamp M (2004) Weak convergence of empirical copula processes. Bernoulli 10:847–860

Frees EW, Valdez EA (1998) Understanding relationships using copulas. N Am Actuar J 2:1–25

Hua L, Joe H (2011) Tail order and intermediate tail dependence for multivariate copulas. J Multivar Anal 102:1454–1471

Hyndman RJ, Fan Y (1996) Sample quantiles in statistical packages. Am Stat 50:361–365

Joe H (1993) Parametric families of multivariate distributions with given margins. J Multivar Anal 46:262–282

Joe H (1997) Multivariate models and dependence concepts. Chapman & Hall, London

Klement EP, Mesiar R (2006) How non-symmetric can a copula be. Comment Math Univ Carolin 47:141–148

Kurowicka D, Joe H (2011) Dependence Modeling: vine copula handbook. World Scientific, Singapore

Longin F, Solnik B (2001) Extreme correlation of international equity markets. J Financ 56:649–676

Marshall AW, Olkin I, Arnold BC (2011) Inequalities: theory of majorization and its applications, 2nd edn. Springer, New York

Nelsen RB (2006) An introduction to copulas, 2nd edn. Springer, New York

Nelsen RB (2007) Extremes of nonexchangeability. Stat Papers 48:329–336

Nikoloulopoulos AK, Joe H, Li H (2012) Vine copulas with asymmetric tail dependence and applications to financial return data. Comp Stat Data Anal 56:3659–3673

Ord JK (1968) The discrete student's *t* distribution. Ann Math Stat 39:1513–1516