This is a previous version of the article published in Environmental and Ecological Statistics. 2016, 23(4): 491-512. doi:10.1007/s10651-016-0350-2

On Conditional Skewness With Applications To Environmental Data

F. Belzunce
Dpto. Estadística e Investigación Operativa, Universidad de Murcia
Facultad de Matemáticas, Campus de Espinardo
30100 Espinardo, Murcia, SPAIN

J. Mulero
Dpto. Matemáticas, Universidad de Alicante
Facultad de Ciencias, Apartado de correos 99
03080 Alicante, SPAIN

J.M. Ruíz
Dpto. Estadística e Investigación Operativa, Universidad de Murcia
Facultad de Matemáticas, Campus de Espinardo
30100 Espinardo, Murcia, SPAIN

A. Suárez-Llorens
Dpto. Estadística e Investigación Operativa, Universidad de Cádiz
Facultad de Ciencias, Campus Puerto Real
11500, Puerto Real, Cádiz, SPAIN
alfonso.suarez@uca.es

May 25, 2016

Abstract

The statistical literature contains many univariate and multivariate skewness measures that allow two datasets to be compared, some of which are defined in terms of quantile values. In most situations, the comparison between two random vectors focuses on univariate comparisons of conditional random variables truncated in quantiles; this kind of comparison is of particular interest in the environmental sciences. In this work, we describe a new approach to comparing skewness in terms of the univariate convex transform ordering proposed by van Zwet (1964), associated with skewness as well as concentration. The key to these comparisons is the underlying dependence structure of the random vectors. Below we describe graphical tools and use several examples to illustrate these comparisons.

Keywords: Skewness, Right-Skewed Distributions, Convex Transform Ordering, Gini, Copula, Environmental Data.

1 Introduction

Estimating and studying the location and variability of a dataset is sometimes supplemented by an investigation of skewness as a measure of symmetry, or more precisely, as a measure of a lack of symmetry. Skewness is intended to represent the departure of a density from symmetry (or sometimes even departure from normality), whereby one tail of the density is more "stretched out" than the other. Although symmetry is natural for distributions with support $(-\infty, \infty)$, the notion of asymmetry also has a place in describing distributions with support $[0, \infty)$. For instance, environmental data are typically skewed, meaning that datasets are not symmetric around the mean or median and frequently have extreme values that stretch out more in one direction.

The study of skewed distributions has long attracted the attention of statisticians. Pearson (1895) considered the gamma distribution as a model for non-symmetric data and an alternative to the usual normal distribution. Pareto (1897) also considered a skewed distribution for the modelling of income distributions. Although skewness can be studied graphically through the box-whisker plot of a univariate dataset, many coefficients have been considered to measure it. We recall some of these in what follows.

Specifically, let X be a random variable with distribution function F_X and let $F_X^{-1}(p) = \inf\{x : F_X(x) \geq p\}$, for $p \in (0,1)$, denote the corresponding generalized quantile function. In basic statistics, we usually measure the skewness of a random variable in terms of the descriptive Fisher's measure

$$A_X = \frac{E[(X-\mu)^3]}{\sigma^3},$$

where μ and σ denote the mean and the standard deviation of X, respectively. A disadvantage of this measure is that it can be arbitrarily large. Others skewness measures based on quantile values assure a more stable and robust procedure in the event of outliers. For example, Boyley's coefficient is given by

$$b_1 = \frac{F_X^{-1}(0.75) + F_X^{-1}(0.25) - 2F_X^{-1}(0.5)}{F_X^{-1}(0.75) - F_X^{-1}(0.25)}.$$

Groeneveld and Meeden (1984) proposed three coefficients (among others) in terms of the quantile function. The first one is given by the curve

$$b_2(p) = \frac{F_X^{-1}(1-p) + F_X^{-1}(p) - 2F_X^{-1}(0.5)}{F_X^{-1}(1-p) - F_X^{-1}(p)},$$

for $0 < p < \frac{1}{2}$. The second one is defined as

$$b_3 = \frac{\int_0^{1/2}[F_X^{-1}(1-p) + F_X^{-1}(p) - 2F_X^{-1}(0.5)]dp}{\int_0^{1/2}[F_X^{-1}(1-p) - F_X^{-1}(p)]dp} = \frac{\mu - F_X^{-1}(0.5)}{E[|X - F_X^{-1}(0.5)|]}.$$

Finally, for X with finite interval support $I = (a,b)$, the third coefficient is defined as

$$b_4 = \lim_{p\to 0+} b_2(p) = \frac{b + a - 2F_X^{-1}(0.5)}{b-a}.$$

Obviously, the previous coefficients are equal to zero when X is symmetric.

When comparing the skewness of two populations, an alternative to skewness coefficients is to use a stochastic ordering, which captures the essence of what is meant when "a distribution function is less skewed than another one" (for further details on stochastic orderings, see, *e.g.*, Shaked and Shanthikumar, 2007). An excellent tool is the univariate convex transform ordering introduced by van Zwet (1964). Given two random variables X_1 and X_2 with distribution functions F_{X_1} and F_{X_2}, it is well known that $\psi(x) = F_{X_2}^{-1}(F_{X_1}(x))$ is an increasing function that maps the pth quantile of X_1 to the corresponding pth quantile of X_2. Additionally, under certain regularity conditions, ψ also stochastically maps X_1 onto X_2, *i.e.*, $\psi(X_1) =_{st} X_2$. Clearly inspired by the fact that an increasing convex function takes on large values in intervals of the form (a, ∞), the random variable X_1 is said to be smaller than X_2 in the convex transform ordering, denoted by $X_1 \leq_c X_2$, if

$$\psi(x) = F_{X_2}^{-1}(F_{X_1}(x)) \text{ is convex on the support of } X_1. \tag{1}$$

In practice, an easy way to verify the convexity of ψ is to examine its plot, given by the so-called quantile-quantile plot (QQ-plot), which can be obtained as the plot of $(F_{X_1}^{-1}(p), F_{X_2}^{-1}(p))$, for all $0 < p < 1$ (see Müller and Stoyan, 2002).

Providing a meaningful interpretation of the convex transform ordering, Marshall and Olkin (2007, p. 70), succinctly stated: "imagine that the density f_{X_1} of a random variable X_1 is graphed on a sheet of rubber that becomes thinner and thinner toward the right, and thus more and more easily stretched toward the right. Now, grasp the right-hand edge of the rubber sheet, stretch it out, and watch the density change shape. If f_{X_1} was symmetric and unimodal before stretching, then f_{X_1} after stretching has become a new density f_{X_2} which is also unimodal, but which has a relatively long right-hand tail, *i.e.*, f_{X_2} is *skewed to the right*. The flexibility requirement of the rubber sheet simply means that the horizontal axis has been transformed by an increasing function ψ with increments increasing as one move to the right, *i.e.*, $\psi(x + \Delta) - \psi(x)$ is increasing in x. Thus that ψ is convex".

In line with the above, it is commonly accepted that any single skewness measure should be preserved by the univariate convex transform ordering (see, *e.g.*, MacGillivray, 1986; Arnold and Groeneveld, 1995). In other words, if $\gamma(X)$ represents a measure of skewness of a random variable X, we can expect that

$$X_1 \leq_c X_2 \Rightarrow \gamma(X_1) \leq \gamma(X_2).$$

This is the case, for instance, of A_X or b_1.

It is also known that, in the case of nonnegative random variables, the convex transform ordering also preserves the Gini coefficient (GI):

$$X_1 \leq_c X_2 \Rightarrow \mathrm{GI}(X_1) \leq \mathrm{GI}(X_2).$$

The Gini coefficient is a widely used income inequality indicator that recently has featured in several environmental science papers (see, *e.g.*, Cullis and van Koppen, 2007; and Chen et al.,2012).

One can also find many skewness measures for multivariate data that can broadly be divided into three groups. The first group is made up of measures based on joint moments of the random variable (Mardia, 1970; Mori et al., 1993). An alternative approach, proposed by Malkovich and Afifi (1973), projects the random variable onto a line and defines the multivariate skewness as the square of the skewness value maximizing some value of univariate skewness. Finally, the third group uses volumes of simplexes (Oja, 1983).

Our aim is to study and compare skewness between two random variables, given some additional information in terms of the conditional truncation of other explanatory variables, and to demonstrate applications in the environmental sciences. The comparisons are based on using the univariate convex transform ordering as proposed by van Zwet (1964) and involve modelling the joint distribution of several random phenomena. Since copulas are mathematical objects that are particularly suitable for modelling multivariate dependence between random variables, independently of their marginal distributions they play a fundamental role in such comparisons, as we will see later on.

Although this work is methodological in nature, the new notions are illustrated below with examples of applications using environmental data on drought, air quality and sunshine and humidity data.

The paper is organized as follows. In Section 2, we define the cs ordering, study its main properties and present some examples involving the notion of copula. In Section 3, we illustrate the interest of skewness comparisons through different datasets of real and simulated data. Finally, Section 4 includes a commentary and our conclusions.

In relation to notation, "increasing" means "non-decreasing" and "decreasing" means "non-increasing". We consider absolutely continuous random variables with conditional distribution functions with interval supports. Given a random variable X, we denote the cumulative distribution function by F_X, the survival function by $\overline{F}_X$ and the corresponding generalized quantile function by F_X^{-1}. Given an event A, we will denote by $[\mathbf{X}|A]$ a random vector, or random variable, whose distribution is the conditional distribution of $\mathbf{X}$ given A.

2 Comparison of skewness for conditioned random variables

In the environmental sciences, there are many situations where we may be interested in analysing the conditional random variable $[X|Y]$. In particular, an important issue to be addressed is the marginal behaviour of the random variable X under an adverse event. In the case of a bivariate random variable (X, Y) with a positive dependence structure, an adverse event typically refers to an unusually large value for Y. In hydrology, for instance, data associated with events such as peak flood or drought severity are computed given that an explanatory variable values like flood volume or drought duration exceed a certain threshold (see, *e.g.*, Shiau, 2003; 2006).

As mentioned, random variables are generally dependent in hydrological problems. Different combinations of rainfall intensity and storm duration may generate storms with quite different characteristics, for instance, or river management may strongly depend upon the joint features of flood peak and flood volume. Therefore, it is often crucial to be able to relate the marginal distributions of different variables in order to obtain a joint law describing the main features of the observed hydrological events. Copulas appear to be the most suitable tool for studying this kind of dependency between two or more variables.

We recall that a bivariate copula $C : [0,1]^2 \to [0,1]$ is a cumulative distribution function with uniform margins on $[0,1]$. The notion of copula was introduced by Sklar (1959) and the main purpose of a bivariate copula is to describe the interrelation between two random variables. Given a random vector $\mathbf{X} = (X, Y)$ with margins F_X, F_Y, there exists a copula $C_{\mathbf{X}}$ such that

$$F_{\mathbf{X}}(x, y) = P[X \le x, Y \le y] = C_{\mathbf{X}}(F_X(x), F_Y(y)).$$

Sklar (1959) showed that any multivariate distribution function inherently embodies a copula function. Furthermore, any copula correctly evaluated with two marginal distributions in the right way leads to a multivariate distribution function. A similar procedure is possible for the joint survival function of a random vector, leading to the notion of survival copula; more precisely, given the joint survival function $\overline{F}$ of a random vector $\mathbf{X}$, there exists a copula $\overline{C}_{\mathbf{X}}$, called survival copula of $\mathbf{X}$, such that

$$\overline{F}_{\mathbf{X}}(x, y) = P[X > x, Y > y] = \overline{C}_{\mathbf{X}}(\overline{F}_X(x), \overline{F}_Y(y)).$$

It is important to note that a random vector has both a copula and a survival copula and that these can be different. For a random vector $\mathbf{X}$, the copula and the survival copula satisfy that

$$C_{\mathbf{X}}(u, v) = u + v - 1 + \overline{C}_{\mathbf{X}}(1 - u, 1 - v).$$

Detailed properties for several types of copulas are described in Nelsen (1999) and Salvadori et al. (2007). Copulas, which are being increasingly used in the environmental sciences, were first used in hydrology by De Michele and Salvadori (2003) and by Favre et al. (2004).

Given $\mathbf{X}_1 = (X_1, Y_1)$ and $\mathbf{X}_2 = (X_2, Y_2)$, two bivariate random vectors, it is natural to wonder about the comparison between X_1 and X_2, when Y_1 and Y_2, respectively, exceed some risk values. Many valuable contributions can be found in this regard in the literature. Roy (2002) and Belzunce et al. (2012) considered stochastic comparisons between the conditional variables $[X_1|Y_1 > y]$ and $[X_2|Y_2 > y]$, for all $y \in \mathbb{R}$, in terms of dispersion and concentration, describing a number of applications in reliability and finance. Given that many situations require the explanatory variable to exceed a risk value given by a quantile, Khaledi and Kochar (2005) proposed a dispersive comparison between $[X_1|Y_1 > F_{Y_1}^{-1}(p)]$ and $[X_2|Y_2 > F_{Y_2}^{-1}(p)]$, for all $p \in (0,1)$. A recent work on this topic is Sordo et al. (2015), who inspired our proposal to compare the conditional distribution functions of $[X_1|Y_1 > F_{Y_1}^{-1}(p)]$ and $[X_2|Y_2 > F_{Y_2}^{-1}(p)]$ (or, equivalently, $[X_1|F_{Y_1}(Y_1) \in (p,1)]$ and $[X_2|F_{Y_2}(Y_2) \in (p,1)]$) in the univariate convex transform ordering. The most important tool applied in our research, called *conditionally more skewed to the right ordering* (cs ordering), is described in what follows.

Definition 2.1. *Let* $\mathbf{X}_1 = (X_1, Y_1)$ *and* $\mathbf{X}_2 = (X_2, Y_2)$ *be two bivariate random vectors. We say that* $\mathbf{X}_2$ *is conditionally more skewed to the right than* $\mathbf{X}_1$*, for short cs ordering and denoted by* $\mathbf{X}_1 \leq_{cs} \mathbf{X}_2$*, if*

$$[X_1|Y_1 > F_{Y_1}^{-1}(p)] \leq_c [X_2|Y_2 > F_{Y_2}^{-1}(p)], \text{ for all } p \in (0,1).$$

Given that interest is normally in one of the margins of the random vector, it should be noted that the cs ordering depends on the permutations of the components of the random vector. Obviously, a stronger order can be defined if we additionally require the comparison on interchanging the role played by the components in Definition 2.1. It is also clear that the cs ordering is reflexive and transitive.

From (1) and just considering the definition of the univariate convex transform ordering, it is apparent that $\mathbf{X}_1 \leq_{cs} \mathbf{X}_2$ holds if the function ψ_p defined as

$$\psi_p(x) = F^{-1}_{[X_2|Y_2 > F_{Y_2}^{-1}(p)]}\left(F_{[X_1|Y_1 > F_{Y_1}^{-1}(p)]}(x)\right), \tag{2}$$

is convex in the support of X_1, for all $p \in (0,1)$.

At this point, it is natural to wonder about the relationship between the function $\psi_p(x)$ given in (2) and the concept of copula. Given a bivariate random vector $\mathbf{X} = (X, Y)$ with copula $C_{\mathbf{X}}$, if we define

$$l_{p,C_{\mathbf{X}}}(u) = \frac{u - C_{\mathbf{X}}(u,p)}{1-p}, \text{ for all } u \in [0,1], p \in (0,1), \tag{3}$$

it is easy to see that

$$P[X \leq x | Y > F_Y^{-1}(p)] = l_{p,C_{\mathbf{X}}}(F_X(x)), \text{ for all } p \in (0,1).$$

Therefore, given two bivariate random vectors $\mathbf{X}_1 = (X_1, Y_1)$ and $\mathbf{X}_2 = (X_2, Y_2)$ with copulas C_1 and C_2, respectively, an equivalent way of expressing (2) is as

$$\psi_p(x) = F_{X_2}^{-1}(l_{p,C_2}^{-1}(l_{p,C_1}(F_{X_1}(x)))), \tag{4}$$

for all x in the support of X_1. As a first consequence, it is clear that the cs ordering does not depend on the distribution functions of the second random variables.

Both the expressions (2) and (4) can be used in practice to check the cs ordering. Intuitively, it seems that (2) is more mathematically tractable when the bivariate distribution functions are given, whereas (4) can be useful for constructing many possible examples, once copulas are identified and we can compute the inverse of (3).

Of the existing types of copula (or survival copula), the Archimedean type is widely used in hydrology (Shiau and Shen, 2001; Favre et al., 2004; Genest and Favre, 2007; Zhang and Singh, 2007). These (survival) copulas are given by

$$\overline{C}_\phi(u,v) = \phi^{-1}(\phi(u) + \phi(v)), \text{ for all } u, v \in (0,1), \tag{5}$$

where ϕ is a continuous, convex and decreasing function called the copula generator. It is also easy to see that ϕ^{-1} is a survival function. As pointed out in Nelsen (1999), many standard bivariate distributions (such as those in the Clayton-Oakes, Gumbel and Frank families) are special cases of this class.

An interesting observation is that when the survival copula belongs to the Archimedean family, then the function $l_{p,C}$ is invertible and closed form expressions for (3) can be obtained. Therefore, it is possible to verify whether the conditions for the cs ordering hold or not. If a random vector $\mathbf{X}$ has an Archimedean survival copula $\overline{C}_\phi$, from (3) and (5) we easily have that

$$l_{p,C_\phi}(u) = 1 - \frac{\phi^{-1}(\phi(1-u) + \phi(1-p))}{1-p}.$$

Now, let $\mathbf{X}_1 = (X_1, Y_1)$ and $\mathbf{X}_2 = (X_2, Y_2)$ be two random vectors with Archimedean survival copulas $\overline{C}_{\phi_1}$ and $\overline{C}_{\phi_2}$, and let, for $i = 1, 2$,

$$\begin{aligned} R_i(x) &= \phi_i(\overline{F}_{X_i}(x)), \text{ for all } x \in \mathbb{R}, \\ W_{p,i}(x) &= \phi_i^{-1}(x + \phi_i(1-p)), \text{ for all } p \in (0,1), x \in \mathbb{R}. \end{aligned}$$

Note that $R_i(x)$ and $W_{p,i}(x)$, $i = 1, 2$, are non-decreasing functions. From (4), the following result is obtained.

Proposition 2.2. *Under the previous notation, if* $\mathbf{X}_1 = (X_1, Y_1)$ *and* $\mathbf{X}_2 = (X_2, Y_2)$ *are two random vectors with Archimedean survival copulas* $\overline{C}_{\phi_1}$ *and* $\overline{C}_{\phi_2}$*, then*

$$\psi_p(x) = R_2^{-1}\left(W_{p,2}^{-1}\left(W_{p,1}(R_1(x))\right)\right). \tag{6}$$

Proof. We have that

$$\begin{aligned} \psi_p(x) &= F_{X_2}^{-1}\left(l_{p,C_{\phi_2}}^{-1}(l_{p,C_{\phi_1}}(F_{X_1}(x)))\right) \\ &= F_{X_2}^{-1}\left(1 - \phi_2^{-1}(\phi_2(\phi_1^{-1}(\phi_1(1 - F_{X_1}(x)) + \phi_1(1-p)) - \phi_2(1-p)))\right) \\ &= R_2^{-1}\left(W_{p,2}^{-1}(W_{p,1}(R_1(x)))\right). \end{aligned} \tag{7}$$

■

Due to well known fact that the composition of non-decreasing convex functions is also a convex function, we have the following result.

Corollary 2.3. *If* R_1 *and* $W_{p,2}^{-1}W_{p,1}$ *are convex, and* R_2 *is concave, then* ψ_p *is convex, i.e.,* $\mathbf{X}_1 \leq_{cs} \mathbf{X}_2$.

Next we provide an interesting example related to the classical Archimedean survival copulas.

Example 2.4. *As a particular case we recall the bivariate Pareto distribution of the first kind, denoted by* $\mathbb{P}(I)$ *(see, e.g., Johnson et al., 1994). For* $\alpha, \beta \geq 0$ *and* $\theta > 0$, $\mathbf{X} = (X, Y) \sim \mathbb{P}(I)(\alpha, \beta, \theta)$ *if its joint survival function is given by*

$$\overline{F}_{\mathbf{X}}(x, y) = P(X > x, Y > y) = \left(\frac{x}{\alpha} + \frac{y}{\beta} - 1\right)^{-1/\theta}, \text{ for all } x > \alpha, y > \beta.$$

It is known that this bivariate distribution has Pareto margins, in particular,

$$\overline{F}_X(x) = \left(\frac{x}{\alpha}\right)^{-1/\theta}, \text{ for all } x > \alpha,$$

and a Clayton-Oakes survival copula, that is,

$$\overline{C}_\theta(u, v) = (u^{-\theta} + v^{-\theta} - 1)^{-1/\theta},$$

which is an Archimedean copula with generator $\phi(x) = x^{-\theta} - 1$.

Now, let $\mathbf{X}_1 = (X_1, Y_1) \sim \mathbb{P}(I)(\alpha_1, \beta_1, \theta_1)$ *and* $\mathbf{X}_2 = (X_2, Y_2) \sim \mathbb{P}(I)(\alpha_2, \beta_2, \theta_2)$*. In this case, we have*

$$\begin{aligned} R_i(x) &= \frac{x}{\alpha_i} - 1, \text{ for all } x > \alpha_i, \\ W_{p,i}(x) &= [x + \phi_i(1-p) + 1]^{-1/\theta_i}, \text{ for all } p \in (0,1), x > \alpha_i. \end{aligned}$$

From Corollary 2.3, $\mathbf{X}_1 \leq_{cs} \mathbf{X}_2$ *whenever* $\theta_1 < \theta_2$*. From* (6)*, the expression of* $\psi_p(x)$ *is given by*

$$\psi_p(x) = \alpha_2 \left\{ \left[\frac{x}{\alpha_1} + (1-p)^{-\theta_1} - 1\right]^{\theta_2/\theta_1} - (1-p)^{-\theta_2} + 1 \right\}, \text{ for all } x \geq \alpha_1.$$

In Figure 1, we plot the joint density functions of bivariate Pareto distributions for $\alpha_1 = \beta_1 = \alpha_2 = \beta_2 = 0.5$, $\theta_1 = 1/9$ *and* $\theta_2 = 1/3$*. Clearly, the bivariate Pareto with* $\theta_2 = 1/3$ *(b) is more skewed to the right than the bivariate Pareto with* $\theta_1 = 1/9$ *(a).*

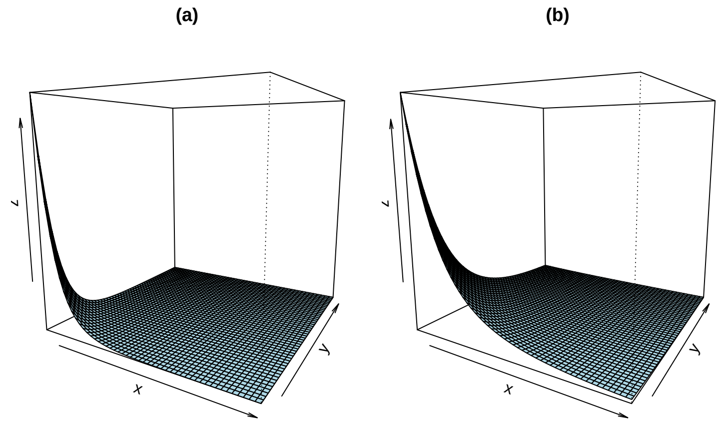

Figure 1: *Joint density function for the bivariate Pareto distributions for* $\alpha_1 = \beta_1 = \alpha_2 = \beta_2 = 0.5$ *and* $\theta_1 = 1/9$ *(a) and* $\theta_2 = 1/3$ *(b).*

Next we present an example where explicit computations of ψ_p are feasible using a non-Archimedean copula. It can be useful to construct many possible examples of the cs ordering by playing with different margins.

Example 2.5. *In this example, the non-Archimedean Farlie-Gumbel-Morgenstern copula is considered. This copula is given by*

$$C_\theta(u,v) = uv + \theta uv(1-u)(1-v),\ \theta \in [-1,1].$$

In this case, $l_{p,C}$ is a polynomial of degree two. Specifically,

$$l_{p,C_\theta}(u) = \theta p u^2 + (1-\theta p)u.$$

Given $\mathbf{X}_1 = (X_1, Y_1)$ and $\mathbf{X}_2 = (X_2, Y_2)$ having Farlie-Gumbel-Morgenstern copulas with parameters θ_1 and θ_2, respectively, by (4), we obtain that

$$\begin{aligned}\psi_p(x) &= F_{X_2}^{-1}(l_{p,C_{\theta_2}}^{-1}(l_{p,C_{\theta_1}}(F_{X_1}(x)))) \\ &= F_{X_2}^{-1}\left(\frac{-(1-\theta_2 p) + \sqrt{[1-\theta_2 p(1-2F_{X_1}(x))]^2 + p^2\theta_2(\theta_1-\theta_2)\left[(1-2F_{X_1}(x))^2 - 1\right]}}{2\theta_2 p}\right).\end{aligned}$$

We now ask ourselves about the relationship between cs ordering and comparisons of the marginal distributions. In particular, if two random vectors are ordered in the cs ordering, the first margins are also ordered in the univariate convex transform ordering.

Proposition 2.6. *Let $\mathbf{X}_1 = (X_1, Y_1)$ and $\mathbf{X}_2 = (X_2, Y_2)$ be two random vectors. If $\mathbf{X}_1 \leq_{cs} \mathbf{X}_2$, then $X_1 \leq_c X_2$.*

Proof. By taking the limit when p tends to 0 in (4), it is evident that $\psi_p(x) = F_{X_2}^{-1}(F_{X_1}(x))$, which concludes easily the proof. ■

Unfortunately, the reverse of Proposition 2.6 is not necessarily true. In order to study when the marginal behaviour determines the cs ordering, we need to fix the dependence structure between the random variables. A particularly interesting situation is to assume that both random vectors share the same dependence structure, *i.e.*, they have the same copula $C = C_1 = C_2$. Important contributions for this case have been made by Müller and Scarsini (2001), Khaledi and Kochar (2005), Belzunce et al. (2008) and Balakrishnan et al. (2011), among others. The result below shows that the cs ordering is reduced to the comparison between the first margins when the bivariate random vectors share the same copula.

Proposition 2.7. *Let $\mathbf{X}_1 = (X_1, Y_1)$ and $\mathbf{X}_2 = (X_2, Y_2)$ be two random vectors with a common copula $C = C_1 = C_2$. Then, $\mathbf{X}_1 \leq_{cs} \mathbf{X}_2$ if, and only if, $X_1 \leq_c X_2$.*

Proof. By (4), it is evident that

$$\psi_p(x) = F_{X_2}^{-1}(F_{X_1}(x)), \text{ for all } p \in (0,1).$$

whenever $C_1 = C_2$. The proof follows easily. ■

Example 2.8. *An interesting situation where two random vectors have the same copula arises when we consider consecutive values for record weather conditions. Chandler (1952) introduced the mathematical notion of record values to study, from a statistical point of view, sequences of record values that arise in practice. Let* $X_1, X_2, \ldots$ *be a sequence of i.i.d. random variables, which can be considered as independent observation of some random environmental data of interest* X*. At this point, we would like to emphasize that samples taken relatively close together in space or time are to some degree redundant, i.e., they are usually highly correlated. Therefore, this model is useful when the samples can be taken sufficiently distant in time in order to eliminate or weaken the autocorrelation.*

Let F *denote the distribution function of an absolutely continuous random variable* X *and let* f *denote the corresponding density function. Record values are defined by means of record times, so first let us recall the definition of record times. Given a sequence of i.i.d. random variables as above, record times are given by*

$$\begin{aligned} &L(1) = 1, \\ &L(j) = \min\{j > L(j-1) | X_j > X_{L(j-1)}\}, \quad j = 2, 3, \ldots \end{aligned}$$

The sequence of the first n ***record values*** *is defined as*

$$X_{(j)} = X_{L(j)}, \quad j = 1, 2, \ldots, n.$$

In this context, given two sequences of record values, e.g. record values based on a random variable measured in two different locations, denoted by X *and* Y*, it would be interesting to provide comparisons of the* r*th record values, i.e.,* $X_{(r)}$ *and* $Y_{(r)}$*, under some additional information regarding a previous record value, for any* $r = 2, \ldots, n$*. The cs notion we introduced above provides a tool for such kind of comparisons. In particular, we provide a result for the comparison in the cs ordering of* $(X_{(r)}, X_{(i)})$ *and* $(Y_{(r)}, Y_{(i)})$ *for all* $i, r \in \{1, \ldots, n\}$ *such that* $i < r$*.*

From the general framework of generalized order statistics explained in detail in Kamps (1995a) and (1995b), it is well known that random vectors given by the first n *record values based on different baseline distributions share a common copula and satisfy that*

$$F^{-1}_{Y_{(j)}}\left(F_{X_{(j)}}(x)\right) = F^{-1}_{Y}(F_X(x)), \quad j = 1, 2, \ldots, n. \tag{8}$$

Therefore, just using the marginalization property of copulas, the random vectors $(X_{(r)}, X_{(i)})$ *and* $(Y_{(r)}, Y_{(i)})$ *share a common copula, for all* $i, r \in \{1, \ldots, n\}$ *such that* $i < r$*.*

Therefore, from Proposition 2.7, we have that

$$\left[X_{(r)} \,\middle|\, X_{(i)} > F^{-1}_{X_{(i)}}(p)\right] \leq_c \left[Y_{(r)} \,\middle|\, Y_{(i)} > F^{-1}_{Y_{(i)}}(p)\right],$$

for all $i, r \in \{1, \ldots, n\}$ *such that* $i < r$*, for all* $p \in (0, 1)$*, if, and only if,* $X_{(r)} \leq_c Y_{(r)}$ *or, equivalently, from (8), if, and only if,* $X \leq_c Y$*.*

It is important to note that for record values, the first component is equally distributed as the distribution from which the record values arise. Consequently, $X_{(1)} \leq_c Y_{(1)}$ *is a sufficient condition for a comparison in the cs ordering between two consecutive record values from two populations.*

It is a well-known fact that a copula is preserved by increasing transformations of the marginal distributions. The following corollary is a direct consequence in case of identical copulas and it merely remains to compare the first margins. Proof has been omitted as being straightforward.

Corollary 2.9. *Let* $\mathbf{X}_1 = (X_1, Y_1)$ *and* $\mathbf{X}_2 = (\alpha(X_1), \beta(Y_1))$ *be two bivariate random vectors where* α *and* β *are strictly increasing. If* α *is a convex function, then* $\mathbf{X}_1 \leq_{cs} \mathbf{X}_2$.

To conclude this section, a meaningful interpretation in terms of bivariate quantile curves and exceedence events of the form $\{X > x, Y > y\}$ is provided.

Proposition 2.10. *Let* $\mathbf{X}_1 = (X_1, Y_1)$ *and* $\mathbf{X}_2 = (X_2, Y_2)$ *be two random vectors. If* $\mathbf{X}_1 \leq_{cs} \mathbf{X}_2$*, then*

$$\begin{aligned}&[X_1|Y_1 > F_{Y_1}^{-1}(p_1), X_1 > F^{-1}_{[X_1|Y_1>F_{Y_1}^{-1}(p_1)]}(p_2)]\\ &\qquad\leq_c [X_2|Y_2 > F_{Y_2}^{-1}(p_1), X_2 > F^{-1}_{[X_2|Y_2>F_{Y_2}^{-1}(p_1)]}(p_2)], \text{ for all } p_1, p_2 \in (0,1).\end{aligned}$$

Proof. Recall that given two random variables X_1 and X_2, it is well-known that the convex transform ordering is preserved under truncation on quantiles, *i.e.*,

$$X_1 \leq_c X_2 \Rightarrow [X_1|X_1 > F_{X_1}^{-1}(p)] \leq_c [X_2|X_2 > F_{X_2}^{-1}(p)], \text{ for all } p \in (0,1). \tag{9}$$

The proof concludes with the calculation, using (9), of the univariate conditional distributions $[X_1|Y_1 > F_{Y_1}^{-1}(p)]$ and $[X_2|Y_2 > F_{Y_2}^{-1}(p)]$. ■

Belzunce et al. (2007) defined a bivariate vector-valued quantile notion that has been successfully applied in hydrology (see Chebana and Ouarda, 2011). Let (X, Y) be an absolutely continuous random vector and $p \in (0,1)$. The pth bivariate quantile set or bivariate quantile curve for the direction ϵ is defined as

$$Q_{X,Y}(p, \epsilon) = \{(x, y) \in \mathbb{R}^2 : F_\epsilon(x, y) = p\},$$

where $F_\epsilon(x, y)$ represents one of the following probabilities: $F_{\epsilon++}(x, y) = \Pr(X \geq x, Y \geq y)$, $F_{\epsilon--}(x, y) = \Pr(X \leq x, Y \leq y)$, $F_{\epsilon+-}(x, y) = \Pr(X \geq x, Y \leq y)$ and $F_{\epsilon-+}(x, y) = \Pr(X \leq x, Y \geq y)$. Chebana and Ouarda (2011) indicate that, of the four events described above, simultaneous exceedence $\{X \geq x, Y \geq y\}$ would be of particular interest in hydrology. This is mainly due to the positive correlation generally observed between the X and Y variables; specifically, those events are important when floods are considered.

Under regularity conditions, quantile curves can be described in a parametric way. It can be observed that $F_{\epsilon++}(x, y) = p$ represents a curve on the plane which can be expressed by means of the quantiles for the conditional distribution $[X|Y \geq y]$ as follows:

$$Q_{X,Y}(p, \epsilon_{++}) = \left\{(x_p(u), y_p(u)) = \left(F^{-1}_{[X|Y \geq F_Y^{-1}(u)]}\left(1 - \frac{p}{1-u}\right), F_Y^{-1}(u)\right) : u < 1 - p\right\}. \tag{10}$$

Figure 2 shows the survival quantile curve $Q_{X,Y}(p, \epsilon_{++})$ given by $F_{\epsilon++}(x, y) = p$. All points of the form $(x_p(u), y_p(u))$ in (10) represent exceedence events such that

$$F_{\epsilon++}(x_p(u), y_p(u)) = \Pr(X \geq x_p(u), Y \geq y_p(u)) = p.$$

Let $\mathbf{X}_1 = (X_1, Y_1)$ and $\mathbf{X}_2 = (X_2, Y_2)$ be two random vectors and $p \in (0, 1)$. Let us consider $(x_{1p}(u), y_{1p}(u))$ and $(x_{2p}(u), y_{2p}(u))$ all points of the form in (10) for $\mathbf{X}_1$ and $\mathbf{X}_2$, respectively. From Proposition 2.10, it is clear that $\mathbf{X}_1 \leq_{cs} \mathbf{X}_2$ implies that

$$[X_1 | X_1 > x_{1p}(u), Y_1 > y_{1p}(u)] \leq_c [X_2 | X_2 > x_{2p}(u), Y_2 > y_{2p}(u)], \ \forall p \in (0, 1) \text{ and } u < 1 - p.$$

Observe that the cs ordering leads us to compare the conditional distributions for all exceedence events given in the pth quantile survival curve.

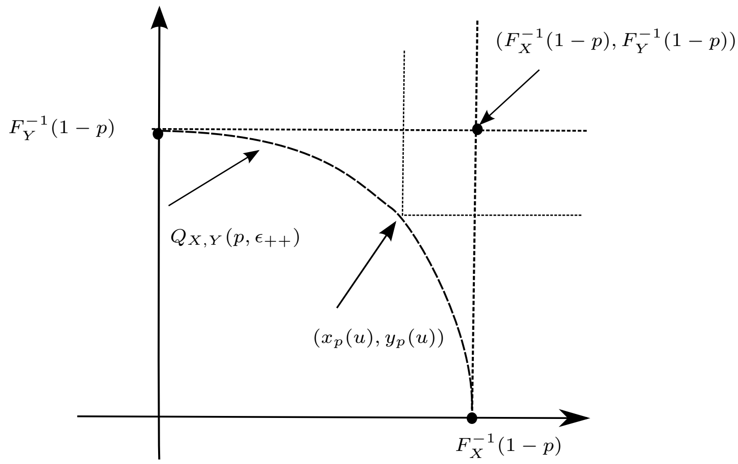

Figure 2: Survival curve.

3 Application to some datasets

The cs concept allows us to compare a whole bunch of examples in terms of skewness. In this section some examples with real data are presented. Our interest is not to develop a formal test but to provide reasonable empirical evidence of the cs ordering.

3.1 Datasets with the same copula

From Proposition 2.7, comparison in the cs ordering is simplified when random vectors have the same copula. Two different situations are described in what follows.

Droughts

Drought stemming from an absence of rainfall can affect humid and arid regions and may imply inadequate water supplies in urban and agricultural areas. Two of the most important features of a drought are duration and severity, defined in terms of the standardized precipitation index (SPI) introduced by McKee et al. (1993). A drought event is defined as a period with negative SPI values. Drought duration, denoted by D, is when the SPI is continuously negative, while drought severity, denoted by S, reflects cumulative values of SPI during a drought as given by

$$S = -\sum_{i=1}^{D} \mathrm{SPI}_i.$$

For convenience, drought severity is taken to be positive. For further information, see Patel et al. (2007). Drought severity and duration are usually abstracted from observed drought data and fitted by a probabilistic model. Of interest is not only the univariate distributions but also the dependence structure.

Let us consider the particular copula-based drought severity-duration study undertaken by Shiau and Modarres (2009), who analysed rainfall data for the period 1954-2003 collected from two stations located in Abadan and Anzali in Iran. In particular, the three-month SPI was calculated as described by Vicente-Serrano (2006). Under these assumptions, let $\mathbf{X}_1 = (S_1, D_1)$ and $\mathbf{X}_2 = (S_2, D_2)$ be two random vectors reflecting drought severity-duration for Abadan and Anzali, respectively. Shiau and Modarres (2009) fitted the univariate distributions and the bivariate copula and showed that drought severity S and duration D can be fitted to the gamma and exponential distributions with univariate density functions given by

$$\begin{aligned} f_S(s) &= \frac{s^{\alpha-1}}{\beta^\alpha \Gamma(\alpha)} e^{-s/\beta}, \text{ for all } s > 0, \\ f_D(d) &= \frac{1}{\lambda} e^{-d/\lambda}, \text{ for all } d > 0, \end{aligned}$$

respectively. By the maximum likelihood method, these authors also estimated the parameters (Table 1), for which the hypothesis of proposed gamma and exponential distributions to model drought severity and duration for Abadan and Anzali by the Kolmogorov-Smirnov test, respectively, could not be rejected.

The Clayton-Oakes copula (see Example 2.4) was considered to model the dependence structure, with the corresponding parameters estimated using the method of inference function for margins (Joe, 1997). In particular, $\widehat{\theta}_{\mathbf{X}_1} = 1.527$ for Abadan and $\widehat{\theta}_{\mathbf{X}_2} = 1.497$ for Anzali (see Table 1).

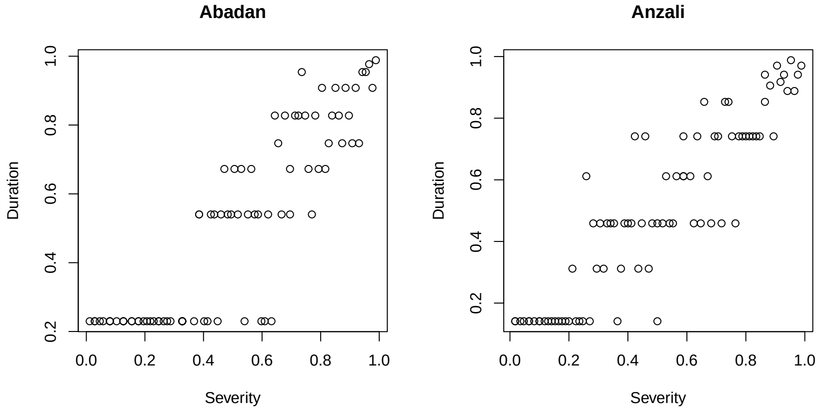

Figure 3: Empirical copulas for $\mathbf{X}_1$ and $\mathbf{X}_2$.

Station		Abadan	Anzali
	$\widehat{\theta}$	1.527	1.497
Severity (gamma)	$\widehat{\alpha}$ (shape)	0.737	0.917
	$\widehat{\beta}$ (scale)	2.796	3.277
Duration (exponential)	$\widehat{\lambda}$	2.125	3.129

Table 1: *Parameters for the margins and copulas for* $\mathbf{X}_1$ *and* $\mathbf{X}_2$*.*

Figure 3 shows estimates of the empirical copula for $\mathbf{X}_1$ and $\mathbf{X}_2$ given by the pseudo-observations. Observing the scatter plots of the empirical copulas and considering the Table 1 estimates of the dependency parameters, we can assume that $\mathbf{X}_1$ and $\mathbf{X}_2$ share the same copula. The fact that estimates of the classical Spearman's rho coefficient - a well-known concordance measure of dependency - take similar values ($\hat{\rho}_S(\mathbf{X}_1) = 0.877$ and $\hat{\rho}_S(\mathbf{X}_2) = 0.905$) reinforces this observation.

It is known that if X_1 and X_2 are distributed as gamma distributions with shape parameters α_1 and α_2, respectively, such that $\alpha_1 \geq \alpha_2$, then $X_1 \leq_c X_2$ (see van Zwet, 1964). In our case, the fact that the estimated shape parameter for S_1 ($\widehat{\alpha}_1 = 0.737$) is smaller than for S_2 ($\widehat{\alpha}_2 = 0.917$) is reasonable empirical evidence to affirm that $S_1 \geq_c S_2$. Hence, the previous fact combined with Proposition 2.7 would lead us to expect that $\mathbf{X}_1$ is more right-skewed than $\mathbf{X}_2$.

Air quality

Let us consider the following two real bivariate data vectors: the daily quantities of ozone (O_3) and nitrogen oxides (NO_X) for years 2011 and 2012 in Murcia (Spain), denoted by

$\mathbf{X}_1 = (O_3^{2011}, NO_X^{2011})$ and $\mathbf{X}_2 = (O_3^{2012}, NO_X^{2012})$, respectively. These data can be downloaded from the Murcia Air Quality website: http://www.carm.es/cmaot/calidadaire/portal/. Ozone (naturally produced in the atmosphere) is helpful in protecting us from the effects of the sun in the upper layer of the sky but can be dangerous when it occurs close to the earth. Nitrogen oxides NO_X produced from the reaction of nitrogen and oxygen gases in the air during combustion, most especially at high temperatures, are also dangerous for human health. The literature amply deals with the relationships between these pollutants (see, *e.g.*, Crutzen, 1970; Clapp and Jenkin, 2001; Wu et al., 2006). At this point, we might ask how the quantities of one pollutant affect concentrations of the other pollutant.

We obtained a bivariate sample of size $n = 365$ for each year. Straightforward computation shows that the time series of ozone and nitrogen oxides are highly auto-correlated, which is only to be expected as the time step is daily for the time series. Since we were not concerned with making predictions, we did not take the auto-correlation into account. For further research, however, it would be interesting to study how auto-correlation affects forecasts. Figure 4 depicts the bivariate plots for 2011 and 2012, showing not only a similar dependence structure, but also that the plot corresponding to 2012 seems to be obtained by expanding the plot corresponding to 2011.

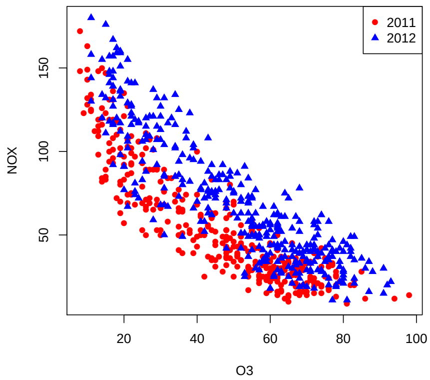

Figure 4: *Bivariate plots.*

As we have seen in Proposition 2.7, the copula approach simplifies the multivariate comparison of random vectors. In fact, copulas have been extensively used for climate studies because the dependence structures of random vectors are shared when dealing with different years or places. The analysis of environmental phenomena makes it possible to easily compare two random vectors.

Figure 5 shows estimates for the empirical copulas for $\mathbf{X}_1$ and $\mathbf{X}_2$ given by the pseudo-observations. From these scatter plots and from the Spearman's rho values of $\hat{\rho}_S(\mathbf{X}_1) = -0.890$ and $\hat{\rho}_S(\mathbf{X}_2) = -0.876$, we can assume that $\mathbf{X}_1$ and $\mathbf{X}_2$ share the same copula.

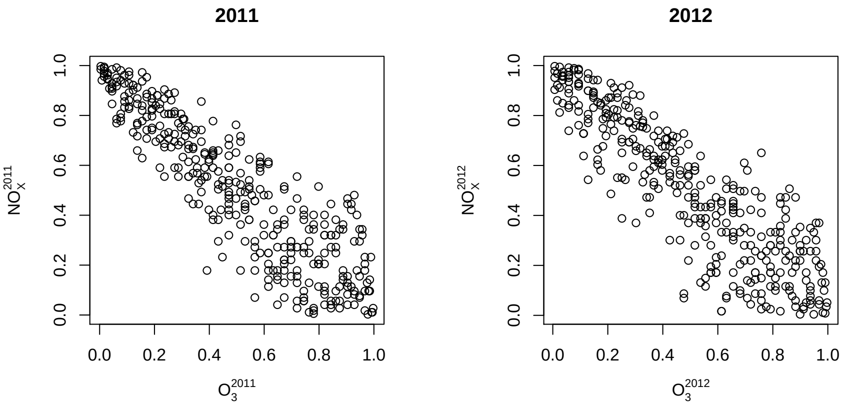

Figure 5: Empirical copulas for $\mathbf{X}_1$ and $\mathbf{X}_2$.

Hence, using Proposition 2.7, a sufficient condition for the cs ordering is given by the comparison of the underlying marginal distributions in the convex order. Figure 6 depicts the classical QQ-plot for O_3^{2011} and O_3^{2012}. Although the properties of the QQ-plot estimate would need to be studied in greater detail in order to develop a formal test for the convex order, Figure 6 shows reasonable empirical evidence that the QQ-plot is convex, *i.e.*, O_3^{2012} is less in the univariate convex transform ordering than O_3^{2011}. Recalling again Proposition 2.7, this fact suggests that $\mathbf{X}_1 \geq_{cs} \mathbf{X}_2$, *i.e.*, $\mathbf{X}_1$ is more right-skewed than $\mathbf{X}_2$.

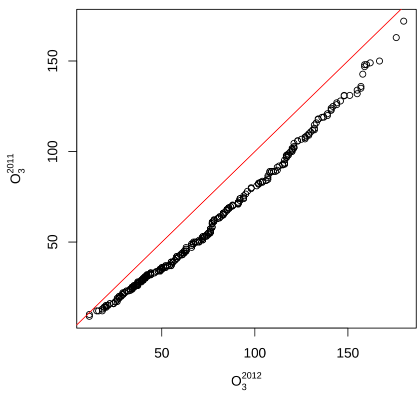

Figure 6: *QQ-plot for* O_3^{2011} *and* O_3^{2012}.

3.2 Datasets with different copulas

Below we provide two graphical tools to study the more complicated case of the cs ordering between two random vectors with different copulas.

As pointed out earlier, convex transform ordering compares skewness between two datasets. Arnold and Groeneveld (1995) suggested that any skewness measure has to be preserved by this order. The convex transform ordering can be considered a concentration order because it preserves the Gini coefficient (see Section 1). Given two bivariate random vectors $\mathbf{X}_1 = (X_1, Y_1)$ and $\mathbf{X}_2 = (X_2, Y_2)$, we denote by GI_i and γ_i, $i = 1, 2$, the following real functions

$$\mathrm{GI}_i : (0,1) \mapsto [0,1], \mathrm{GI}_i(p) = \mathrm{GI}\left([X_i|Y_i > F_{Y_i}^{-1}(p)]\right),$$

$$\gamma_i : (0,1) \mapsto \mathbb{R}, \gamma_i(p) = \gamma\left([X_i|Y_i > F_{Y_i}^{-1}(p)]\right).$$

In practice, the Gini coefficient function is defined as twice the area between the 45 degree line and the Lorenz curve, and it can be easily computed. In particular, we use the `reldist` package in the statistical software `R`. Of the many skewness measures that can be considered, Joanes and Gill (1998) discussed three methods. In particular, they estimated $\gamma = m_3/m_2^{3/2}$, where m_r are the sample moments of order r, as a skewness measure that can be computed using the `e1071` package in `R`.

Following the same procedure as in (1), the sets

$$\{(p_k, \widehat{\mathrm{GI}}_i(p_k)), \text{ for } k = 1, \ldots, m\} \text{ and } i = 1, 2, \tag{11}$$

$$\{(p_k, \widehat{\gamma}_i(p_k)), \text{ for } k = 1, \ldots, m\} \text{ and } i = 1, 2, \tag{12}$$

provide non-parametric estimations of the graphs for GI_i and γ_i, $i = 1, 2$. Note that $\widehat{\mathrm{GI}}_i$ and $\widehat{\gamma}_i$ represent non-parametric estimators of the Gini and skewness indexes, respectively, based on the empirical distribution, and that p_k, $k = 1, \ldots, m$, are univariate values in $(0, 1)$. From the graphical plots of (11) and (12) for $i = 1, 2$ we can easily compare the Gini indexes or skewness coefficients for the conditional distributions of $\mathbf{X}_1$ and $\mathbf{X}_2$, respectively.

The results obtained would indicate these numerical methods to be feasible paths to studying the cs ordering between two bivariate random vectors (the `R`-code is provided in Belzunce et al., 2015). An example follows.

Sunshine and humidity

To illustrate the bivariate analysis in the case of different copulas, we used daily data on sunshine and humidity for Alicante in Spain and Berlin-Tegel in Germany downloaded from the website of the European Climate Assessment and Dataset project (http://www.ecad.eu/; for further information, see Klein-Tank et al., 2002). As in Example 3.1, we are not interested in making predictions, so we assumed that both bivariate samples are representative of random quantities of sunshine and humidity. Let $\mathbf{X}_1 = (X_1, Y_1)$ and $\mathbf{X}_2 = (X_2, Y_2)$ be the sunshine and humidity observations for Alicante and Berlin-Tegel, respectively.

Figure 7 depicts the scatter plots for the empirical copulas, revealing important differences. In this case, the Spearman's rho values ($\hat{\rho}_S(\mathbf{X}_1) = -0.311$ and $\hat{\rho}_S(\mathbf{X}_2) = -0.849$) were not similar as happened with the previous cases, so we can assume that $\mathbf{X}_1$ and $\mathbf{X}_2$ do not share the same copula.

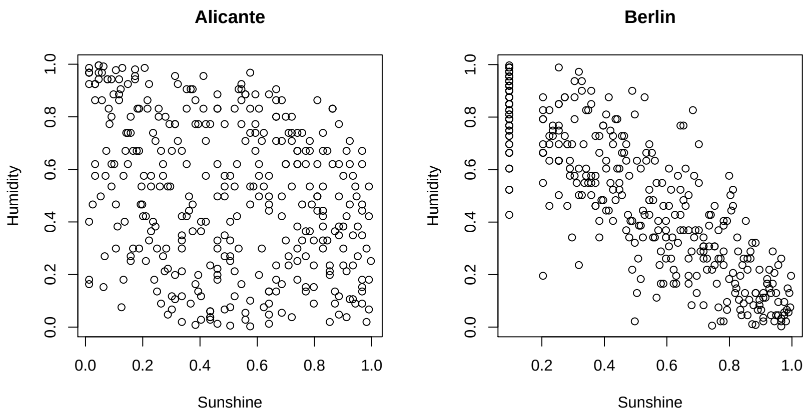

Figure 7: Empirical copulas for $\mathbf{X}_1$ and $\mathbf{X}_2$.

Figure 8 depicts the QQ-plots and the conditional Gini and skewness coefficients for both random vectors. Note that although the QQ-plots in (a) are not clearly convex, the two indexes are ordered.

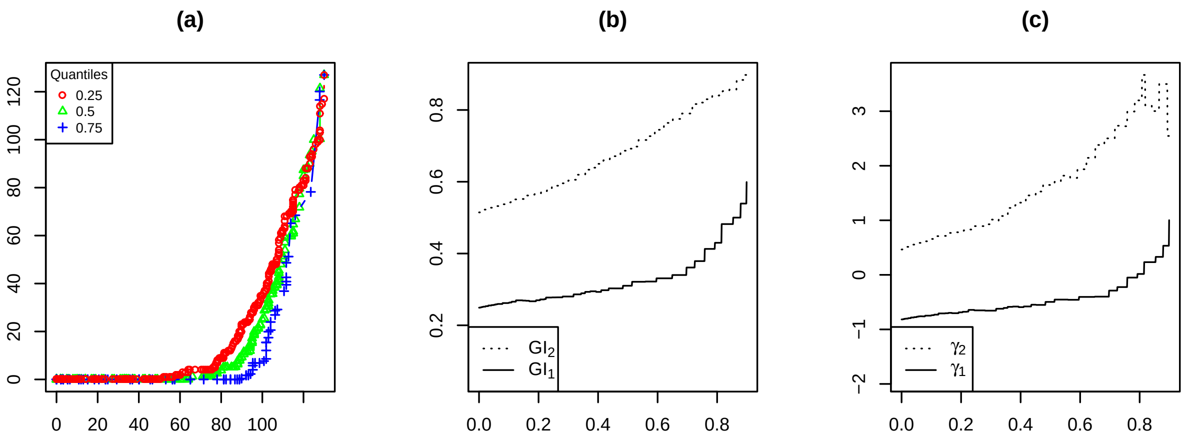

Figure 8: *QQ-plots of the conditional random variables for* $p = 0.25, 0.5, 0.75$ *(a), conditional Gini indexes (b) and conditional skewness coefficients for* $\mathbf{X}_1$ *and* $\mathbf{X}_2$ *(c).*

4 Conclusions and further remarks

Skewness can be studied through comparisons of many single measures, but stochastic orderings and, in particular, the univariate convex transform ordering proposed by van Zwet

(1964), provide a more complete comparison between skewness for two random variables. In situations where bivariate random vectors are considered, interest focused on the behaviour of a certain type of conditional random variable constructed with the margins of such random vectors. For example, given $\mathbf{X} = (X_1, Y_1)$ and $\mathbf{Y} = (X_2, Y_2)$, we may be interested in comparing $[X_1|Y_1 > F_{Y_1}^{-1}(p)]$ and $[X_2|Y_2 > F_{Y_2}^{-1}(p)]$, and hence we may want to decide which of those vectors is more right-skewed.

As was demonstrated above, the dependency structure of the random vectors plays an important role in an analysis of skewness through the cs ordering. When we are dealing with two random vectors with the same copula, it is merely a matter of comparing the first margins in the univariate convex transform order. The situation is more complicated for random vectors with two different copulas. Analytical comparison can be made for some Archimedean copulas. For the rest of cases, two graphical tools were described.

The study of real phenomena usually involves random vectors with more than two random variables. Note that the cs ordering ordering can easily be extended to the general multivariate case. Thus, given two random vectors $\mathbf{X} = (X_1, \ldots, X_n)$ and $\mathbf{Y} = (Y_1, \ldots, Y_n)$, $\mathbf{X}$ is said to be smaller than $\mathbf{Y}$ in the conditionally skewed to the right order, denoted by $\mathbf{X} \leq_{cs} \mathbf{Y}$, if

$$[X_1|X_2 > F_{X_2}^{-1}(p), \ldots, X_n > F_{X_n}^{-1}(p)] \leq_c [Y_1|Y_2 > F_{Y_2}^{-1}(p), \ldots, Y_n > F_{Y_n}^{-1}(p)], \text{ for all } p \in (0, 1).$$

Note that all properties and relationships of the cs ordering can be generalized to this general case and that the cs ordering can also be generalized to the conditional distribution of X_1 given $F_{Y_1}(Y_1) \in [p_1, p_2]$ for $0 \leq p_1 < p_2 \leq 1$. In that case, $P(X_1 \leq x|F_{Y_1}(Y_1) \in [p_1, p_2]) = l_{p_1,p_2,C}(F_{X_1}(x))$, where C_1 is the copula of $\mathbf{X}_1 = (X_1, Y_1)$ and

$$l_{p_1,p_2,C_1}(u) = \frac{C_1(u, p_2) - C_1(u, p_1)}{p_2 - p_1}.$$

Since the comparison is based on a function ψ_{p_1,p_2}, many of the properties for ψ_p can be easily translated to ψ_{p_1,p_2}. Special cases are, for instance, $p_1 = p$, $p_2 = 1$ (which is the focus of the present work), and also $p_1 = 0$, $p_2 = p$ or $p_2 \to p_1 = p$, to be considered in future research.

Acknowledgments

The authors wish to thank Profs. Shiau and Modarres for the drought dataset and two anonymous referees and the journal editor for comments that have significantly improved this paper. The research of Félix Belzunce, Julio Mulero and José María Ruíz is partially funded by the Ministerio de Economía y Competitividad (Spain) under Grant MTM2012-34023-FEDER. Alfonso Suárez-Llorens acknowledges support received from the Ministerio de Economía y Competitividad (Spain) under Grant MTM2014-57559-P.

References

[1] Arnold B.C., Groeneveld R.A. 1995. Measuring skewness with respect to the mode. *The American Statistician* **49**, 34–38.

[2] Balakrishnan N., Belzunce F., Sordo M.A., Suárez-Llorens A. 2011. Increasing directionally convex orderings of random vectors having the same copula, and their use in comparing ordered data. *Journal of Multivariate Analysis* **105**, 45-54.

[3] Belzunce F., Castaño A., Olvera-Cervantes A., Suárez-Llorens A. 2007. Quantile curves and dependence structure for bivariate distributions. *Computational Statistics & Data Analysis* **51**, 5112–5129.

[4] Belzunce F., Mulero J., Ruíz J.M., Suárez-Llorens A. 2012. New multivariate orderings based on conditional distributions. *Applied Stochastic Models in Business and Industry* **28**, 467–484.

[5] Belzunce, F., Mulero, J., Ruíz J.M., Suárez-Llorens, A., 2015. An R script to plot the conditioned Gini (or skewness) coefficient in the bivariate case. Available at http://hdl.handle.net/10045/52083.

[6] Belzunce F., Ruíz J.M., Suárez-Llorens A. 2008. On multivariate dispersion orderings based on the standard construcion. *Statistics and Probability Letters* **78**, 271–281.

[7] Chandler K.N. 1952. The distribution and frequency of record values. *Journal of the Royal Statistical Society, Series B* **14**, 220–228.

[8] Chebana F., Ouarda T.B.M.J. 2011. Multivariate quantiles in hydrological frequency analysis. *Environmetrics* **22**, 63–78.

[9] Chen B., Gu G., Wang H., Tao J., Wu Y., Liu Q., Xu H. 2012. Research on Gini coefficient application in the distribution of water pollution. *Journal of Food, Agriculture & Environment* **10**, 833–838.

[10] Clapp L.J., Jenkin M.E. 2001. Analysis of the relationship between ambient levels of O_3, NO_2 and NO as a function of NO_X in the UK. *Atmospheric Environment* **35**, 6391-6405.

[11] Crutzen P.J. 1970. The influence of nitrogen oxides on the atmospheric ozone content. *Quarterly Journal of the Royal Meteorological Society* **96**, 320-325.

[12] Cullis J., Van Koppen B. 2007. Applying the Gini coefficient to measure inequality of water use in the olifants river water management area. *International Water Management Institute (IWMI): Colombo, Sri Lanka*. Research Report 113.

[13] De Michele C., Salvadori G. 2003. A generalized Pareto intensity-duration model of storm rainfall exploiting 2-copulas. *Journal of Geophysical Research: Atmospheres* **108**, Art. No. 4067.

[14] Favre A.-C., El-Adlouni S., Perreault L., Thiémonge N., Bobée B. 2004. Multivariate hydrological frequency analysis using copulas. *Water Resources Research* **40**, W01101.

[15] Genest C., Favre A.C. 2007. Everything you always wanted to know about copula modelling but were afraid to ask. *Journal of Hydrologic Engineering* **12**, 347-368.

[16] Groeneveld R.A., Meeden G. 1984. Measuring skewness and kurtosis. *The Statistician* **33**, 391–399.

[17] Joanes D.N., Gill C.A. 1998. Comparing measures of sample skewness and kurtosis. *The Statistician* **47**, 183-189.

[18] Joe H. 1997. *Multivariate Models and Dependence Concepts*. Chapman and Hall, London,

[19] Johnson N.L., Kotz S., Balakrishnan N. 1994. *Continuous Univariate Distributions*. Wiley, New York.

[20] Kamps U. 1995a. A concept of generalized order statistics. *Journal of Statistical Planning and Inference* **48**, 1-23.

[21] Kamps U. 1995b. *A Concept of Generalized Order Statistics*, B.G. Taubner, Stuttgart.

[22] Khaledi B.E., Kochar S.C. 2005. Dependence, dispersiveness, and multivariate hazard rate ordering. *Probability in the Engineering and Informational Sciences* **19**, 427-446.

[23] Klein-Tank A.M.G. and Coauthors. 2002. Daily dataset of $20th$-century surface air temperature and precipitation series for the European Climate Assessment. *International Journal of Climatology* **22**, 1441-1453.

[24] MacGillivray H.L. 1986. Skewness and asymmetry: measures and orderings. *The Annals of Statistics* 14, 994-1011.

[25] Malkovich J.F., Afifi A.A. 1973. On tests for multivariate normality. *Journal of the American Statistical Association* **68**, 176-179.

[26] Mardia K.V. 1970. Measures of multivariate skewness and kurtosis with applications. *Biometrika 57*, 519-530.

[27] Marshall A.W., Olkin I. 2007. *Life Distributions. Structure of Nonparametric, Semiparametric, and Parametric Families*. Springer Series in Statistics. Springer, New York.

[28] McKee T.B., Doesken J., Kleist J.A.M.S. 1993. The relationship of drought frequency and duration to time scales. *Proceedings of the 8th Conference on Applied Climatology*, 179-184.

[29] Mori T.F., Rohatgi V.K., Székely G.J. 1993. On multivariate skewness and kurtosis. *Theory of Probability and its Applications* **38**, 547-551.

[30] Müller A., Scarsini M. 2001. Stochastic comparison of random vectors with a common copula. *Mathematics of Operations Research* **26**, 723-740.

[31] Müller A., Stoyan D. 2002. *Comparison Methods for Stochastic Models and Risks*. John Wiley & Sons, Chichester, England.

[32] Nelsen R.B. 1999. *An Introduction to Copulas*. Lectures Notes in Statistics **139**. Springer-Verlag, New York.

[33] Oja H. 1983. Descriptive statistics for multivariate distributions. *Statistics and Probability Letters* **1**, 327-332.

[34] Pareto V. 1897. *Cours d'Économie Politique Vol. 2*. F. Pichou, Paris.

[35] Patel N.R., Chopra P., Dadhwal V.K. 2007. Analyzing spatial patterns of meteorological drought using standardized precipitation index. *Meteorological Applications* **14**, 329-336.

[36] Pearson K. 1895. Contributions to the Mathematical Theory of Evolution, II: Skew Variation in Homogeneous Material. *Transactions of the Royal Philosophical Society, Series A* **186**, 343-414.

[37] Roy D. 2002. Classification of multivariate life distributions based on partial ordering. *Probability in Engineering and Information Sciences* **16**, 129–137.

[38] Salvadori G., De Michele C., Kottegoda N. T., Rosso R. 2007. *Extremes in Nature. An Approach Using Copulas* **56** of Water Science and Technology Library. Springer, Dordrecht.

[39] Shaked M., Shanthikumar J.G. 2007. *Stochastic Orders*. Springer Series in Statistics. Springer, New York.

[40] Shiau J.T. 2003. Return period of bivariate distributed hydrological events. *Stochastic Environmental Research and Risk Assessment* **17**, 42–57.

[41] Shiau J.T. 2006. Fitting drought duration and Severity with two-dimensional copulas. *Water Resources Management* **20**, 795–815.

[42] Shiau J.T., Modarres R. 2009. Copula-based drought severity-duration-frequency analysis in Iran. *Meteorological Aplications* **16**, 481-489.

[43] Shiau J.T., Shen H.W. 2001. Recurence analysis of hydrologic droughts of difering severity. *Journal of Water Resources Planning and Management* **127**, 30-40.

[44] Sklar A. 1959. Fonctions de répartitions à n dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Université de Paris* **8**, 229-231.

[45] Sordo M.A., Suárez-Llorens A., Bello A.J. 2015. Comparison of conditional distributions in portfolios of dependent risks. *Insurance: Mathematics and Economics* **61**, 62-69.

[46] Van Zwet W.R. 1964. *Convex Transformations of Random Variables*. Mathematical Centre Tracts, 7 Mathematisch Centrum, Amsterdam.

[47] Vicente-Serrano S.M. 2006. Differences in spatial patterns of drought on different time scales: an analysis of the Iberian Peninsula. *Water Resources Management* **20**, 37-60.

[48] Wu Y., Li Y., Liu W. 2006. Study on Gini coefficient method of total pollutant load allocation for water bodies. *Research of Environmental Sciences* **19**, 66-70.

[49] Zhang L., Singh V.P. 2007. Bivariate rainfall frequency distributions using Archimedean copulas. *Journal of Hydrology* **332**, 93-109.