Original Research Paper

New Families of Distributions for Modeling Bivariate Data, with Applications

[1]Haroon Barakat, [2]Osama Mohareb Khaled and [3]Nourhan Khalil

[1]*Department of Math, Faculty of Science, Zagazig University, Zagazig, Egypt*
[2]*Department of Math, Faculty of Science, Port said University, Port Said, Egypt*
[3]*Department of Physics and Engineering Math, Faculty of Engineering, Port Said University, Port Said, Egypt*

Abstract: In this study we introduce a new method of adding two shape parameters to any baseline bivariate distribution function (df) to get a more flexible family of bivariate df's. Through the additional parameters we can fully control the type of the resulting family. This method is applied to yield a new two-parameter extension of the bivariate standard normal distribution, denoted by BSSN. The statistical properties of the BSSN family are studied. Moreover, via a mixture of the BSSN family and the standard bivariate logistic df, we get a more capable family, denoted by FBSSN. Theoretically, each of the marginals of the FBSSN contains all the possible types of df's with respect to the signs of skewness and excess kurtosis. In addition, each possesses very wide range of the indices of skewness and kurtosis. Finally, we compare the families BSSN and FBSSN with some important competitors (i.e., some generalized families of bivariate df's) via real data examples. AMS 2010 Subject Classification: 62-07; 62E10; 62F99.

Keywords: Bivariate Non-Normal Distribution, Parametric Family, Mixture Distributions, Bivariate Data Modeling

Introduction

Multivariate data is obtained when we observe more than one statistical outcome variable at a time. Therefore, it consists of two or more observed random variables (rv's) that are usually correlated. The most appropriate and commonplace multivariate distribution that can describe the multivariate data is the multivariate normal distribution. Most likely, the popularity of normal distribution is due to the ease of simulation and the possibility of deriving closed-form theoretical results. It is rare, however, that the multivariate normal distribution can describe the given multivariate data. Virtually, most of the obtainable multivariate data are skewed and heavy tailed. So we are in bad need to non-normal multivariate df's to describe such data. Examples of random phenomenona that are governed by non normal multivariate df's would be lengths of time among malfunctions in machinery, waiting times and growth data such as bacterial growth.

Actually, the non-normal multivariate df's are needed in several cases. First, when the marginal df's are not normal. Second, when the contour of the constant density is not ellipse. Third, the conditional expectations are not linear regressions. Last, the variances and covariances of conditional df's are dependent of the values of the conditioning variables. Although the roots of multivariate distribution theory lie in univariate distribution theory, the extension to the multivariate domain introduces additional concepts and issues of particular relevance.

Despite the paucity of studies tackling the generalized families of multivariate df's in comparison to the studies, that tackle the generalized families of univariate df's, many authors have considered and studied the multivariate df's that describe non-normal multivariate data. Among those authors are Cook and Johnson (1981; 1986). They found that non-normal df's with normal margins would be appropriate for some bivariate geological data. Early work goes back to Edgeworth (1896), where families of bivariate distributions are derived as series involving the bivariate probability density function (pdf) and all its partial derivatives as well as the cumulants of the distribution (cf. Mardia, 1970). In our study, we mainly focus on studying the bivariate models. Generally, a

bivariate df is better suited for practical use when it has the following properties:

- The bivariate df is mathematically tractable, i.e., we can derive its distributional characterizations, e.g., moments, cross moments, median vector, mode vector, marginal skewness and kurtosis. Moreover, via the parameters of the bivariate df, we can control some of its characterizations
- The bivariate df is flexible to model various types of data by possessing a wide range of the indices of the marginal skewness, the marginal kurtosis and the correlation coefficient
- The bivariate df is motivated by important applications

In this study we suggest a method of adding two parameters to any base bivariate df to obtain a new generalized family of bivariate df's, which is more flexible to fit bivariate data. This new method is based on the mixture of the df's of some random vectors (X_1, X_2) and (-X_1, -X_2). We apply this method on the bivariate standard normal distribution, denoted by BN, to yield an extension of the bivariate normal distribution, which can serve as a strong competitor of many known non-normal bivariate distributions. For example, the new extension of the bivariate normal distribution contains, beside the normal df itself, positively-negatively asymmetric and leptokurtic-platykurtic marginals. Moreover, through the added two parameters, one can fully control the skewness, kurtosis of the marginals of this family, as well as its correlation coefficient. Briey, the suggested new family of bivariate df's satisfies the above mentioned preferable properties 1-3.

The rest of the paper unfolds as follows. Section 2 briey outlines two important nonnormal bivariate df's. In section 3, we introduce and study a new suggested family of bivariate normal df, denoted by BSSN. Moreover, we show that the BSSN family is capable of describing many types of statistical bivariate data. Finally, we suggest a more competent family, which is an extension of the BSSN family, denoted by FBSSN. Theoretically, each of the marginals of the FBSSN family contains all the possible types of df's with respect to the signs of skewness and excess kurtosis and possesses very wide range of the indices of skewness and kurtosis. In section 4, the AIC and BIC criteria are used to compare the families BSSN, FBSSN, Bivariate Tukey *g-h* (denoted by BT-*g-h*), Bivariate Skew Normal (denoted by BSN), Bivariate Mixture Normal (denoted by BMN) and the usual Bivariate Normal df (denoted by BN) by means of three real bivariate data sets, concerning the air pollution. Section 5, is devoted to a brief discussion.

Two Important Non-Normal Bivariate df's

In this section we review the bivariate g-and-h df, which is introduced by Kowalchuk and Headrick (2010) and the bivariate skew normal df (Azzalini, 2005).

Bivariate g-and-h Distribution

The *g*-and-*h* univariate distribution was suggested by Tukey (1977) and discussed by Hoaglin and Peters (1979) and Hoaglin (1983). This distribution is defined by transforming the standard normal variable Z to:

$$X = T_{g,h}(Z) = \frac{\exp(gZ) - 1}{g} \exp\left(\frac{h}{2}Z^2\right), \tag{2.1}$$

where, g is a real constant controlling the skewness and h is a nonnegative real constant controlling the kurtosis, or elongation (for the definition of *g*-and-*h* distribution when $h \in \mathbb{R}$, see Martinez and Iglewicz, 1984). It can be shown that $T_{0,0}(Z) = Z$ is the normal one, the distribution of $T_{0,h}(Z)$ has heavier tails than normal (increasingly with *h*) and $T_{g,0}(Z) = \frac{\exp(gZ) - 1}{g}$, coincides with the location-scale log-normal distribution. The sign of *g* controls the direction of skewness but not the amount. Positive values of *g* skew the distribution to right tail while negative values of *g* skew the distribution to the left tail. Therefore, this family of df's encompasses a remarkable variety of distribution types. Moreover, since $T_{g,h}(z)$ increases monotonically in z, the df of *X* can be written as $F_X(x) = \Phi\left(T_{g,h}^{-1}(x)\right)$, where $\Phi(.)$ denotes the standard normal df. Clearly, this distribution is not easily mathematically tractable, because it does not have simple expression for density. However, Martinez (1981) derives the nth moment about the origin, when $g \neq 0$ and $0 \leq h < \frac{1}{n}$, as:

$$E(X^n) = \frac{1}{g^n \sqrt{1-nh}} \sum_{i=0}^{n} (-1)^i \binom{n}{i} \exp\left(\frac{\left[(n-i)g\right]^2}{2(1-nh)}\right).$$

A common approach for generating correlated *g*-and-*h* distributions uses the Equation 2.1 with bivariate normal data with a specified correlation structure (see Wilcox, 1995; Wilcox and Keselman, 2001; 2003; Wilcox *et al.*, 2000). Namely, a random vector $\underline{X} \in \mathbb{R}^2$ is said to have a standard bivariate *g*-and-*h* distribution, where $\underline{g} = (g_1, g_2)^T \in \mathbb{R}^2$ controls the skewness and h = $(h_1, h_2)^T \in \mathbb{R}_+^2$ controls the kurtosis, if it can be represented as:

$$\underline{X} = \left(T_{g_1,h_1}(Z_1), T_{g_2,h_2}(Z_2)\right)^T, \tag{2.2}$$

where, $(Z_1, Z_2)^T$ has a standard bivariate normal $\Phi_\rho(.,.)$ with correlation coefficient ρ and $T_{g_i, h_i}(Z_i)$, $i = 1, 2$, is defined in (2.1). It is worth mentioning that the original (normal) correlation structure is not maintained because this structure is not invariant to the transformation (2.1). For example, Wilcox (1995) used the transformation (2.1), with $h = 0$, in the context of a bivariate case with a correlation of $\rho = 0.5$ where both distributions have $g = 1$ and $h = 0$. Wilcox (1995) showed that $X_1 = T_{1,0}(Z_1)$ and $X_2 = T_{1,0}(Z_2)$ will have the correlation coefficient of 0.378. In general the correlation between $X_1 = T_{g_1, h_1}(Z_1)$ and $X_2 = T_{g_2, h_2}(Z_2)$ can be expressed as:

$$\rho_{\underline{x}} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} T_{g_1, h_1}(z_1) T_{g_2, h_2}(z_2) \phi_\rho(z_1, z_2) dz_1 dz_2,$$

where, $\phi_\rho(z_1, z_2)$ is the standard bivariate normal density, with correlation ρ and in general we have $\rho_X \neq \rho$.

Bivariate Skew Normal Distribution

The term Skew-Normal (SN) df refers to a parametric class of df's which includes the standard normal as a special case. A rv Y is said to be skew-normal with parameter λ, written $Y \sim SN(\lambda)$, if its density function (cf. Azzalini, 1985; Gupta *et al.*, 2002) is:

$$A_Y(z; \lambda) = 2\phi(z)\Phi(\lambda z), z \in \mathbb{R} \qquad (2.3)$$

where, $\phi(.)$ denotes the standard normal density function. The parameter λ regulates the skewness and $\lambda = 0$ corresponds to standard normal df. It is worth mentioning that the rv Y itself may be interpreted as the following transformation of normal rv's: $Y = \frac{\lambda}{\sqrt{1 + \lambda^2}} |Z_1| + \frac{1}{\sqrt{1 + \lambda^2}} Z_2$, where Z_1 and Z_2 are two independent rv's following the standard normal distribution. The advantage of this distribution family is that it persists many statistical properties of the normal distribution, in particular, $Y^2 \sim \chi_1^2$. The SN df enables us getting rid of the assumption that the underlying population distribution is symmetric. Such an extension is necessary because in practice, the underlying distribution may be well skew rather than being symmetric.

When applying the SN df in statistical inference, we frequently need to study the joint df of a random sample from the population. Therefore, the study of multivariate SN df is of a considerable interest. Azzalini (1985) introduced the following bivariate extension of the density in (2.3). A random vector $\underline{Y} \in \mathbb{R}^2$ follows a bivariate SN, denoted BSN$_1$, if the joint density of $\underline{Y}$ has the following form:

$$f_{\underline{Y}}^{(1)}(\underline{Y}; \underline{\lambda}) = c\phi(\underline{y}; \Omega)\Phi(\lambda_1 y_1)\Phi(\lambda_2 y_2), \qquad (2.4)$$

where, $\lambda_1, \lambda_2 \in \mathbb{R}$, c is the normalizing constant and $\phi(\underline{y}; \Omega)$ is the density of a bivariate normal distribution with correlation matrix $\Omega = (\rho_{ij})$, $i, j = 1, 2$, $\rho_{ij} = \rho$, if $i = j$, while $\rho_{ij} = \frac{\lambda_i}{\sqrt{1 + \lambda_i}} \frac{\lambda_j}{\sqrt{1 + \lambda_j}} \rho$, if $i \neq j$. Note that Ω is generally not the correlation matrix of $\underline{Y}$. Clearly, the df defined in (2.4) is just a direct and formal extension of the univariate SN distribution defined in (2.3). In Azzalini and Dalla Valle (1996) pointed out the disadvantage of the model defined in (2.4) and suggested the following version for bivariate SN distribution, denoted by BSN$_2$. A random vector $\underline{V} \in \mathbb{R}^2$ follows BSN$_2$ distribution if the joint density of $\underline{V}$ has the following form:

$$f_{\underline{V}}^{(2)}(\underline{V}; \underline{\lambda}) = 2\phi(\underline{v}; \Omega)\Phi(\lambda_1 v_1 + \lambda_2 v_2). \qquad (2.5)$$

Azzalini and Dalla Valle (1996) and Azzalini and Capitanio (1999) showed that the model (2.5) has some interesting statistical properties such as the quadratic form of a skew normal random vector is χ_2^2, along with a probabilistic interpretation on the basis of the stochastic representation. However, in the definition is Ω not the correlation matrix of $\underline{V}$. Moreover, $\Omega = I$ (where I is the identity matrix of order 2×2) does not imply that all components of $\underline{V}$ are independent. In a series of articles, Gupta and Chen (2001; 2003; 2004) pointed that neither of the bivariate skew-normal models (2.4, 2.5) cohere with the joint distribution of a random sample from a univariate skew-normal distribution. They introduced an alternative bivariate skew-normal model which overcomes this problem.

Bivariate SS-Normal Distribution and the Full Family

It is known that the mixture model, which is a convex combination of two pdf's is a powerful and flexible tool of modeling complex data because it combines the properties of the individual densities. In Barakat (2015) suggested a univariate generalized distribution family, which is simpler than the most well-known families. This is because it is a mixture of a baseline distribution and its reverse (i.e., the distribution of the negative rv), after adding a location parameter to the baseline distribution and adding the same location parameter with a different sign to the reverse of the baseline distribution. By this way the utilized location parameter turns to a shape parameter and we get a two-parameter generalized distribution family. This family has an individual trait, on which it has been built, that if this family contained

any distribution, it should contain also its reverse distribution. This trait makes us name this family as the stable. Barakat (2015) suggested a new univariate generalized family, denoted by SSN, by taking the standard normal distribution as the baseline distribution of the stable symmetric family. Barakat (2015) showed that the SSN family contains all the possible types of df's (there are nine possible types, see Barakat, 2015. Moreover, any family that contains these nine types is called full family, cf. Barakat and Khaled, 2017), except the type "symmetric and leptokurtic", denoted by 0+; i.e., positive excess kurtosis or the df has a more acute peak around the mean and fatter tails than normal df. Moreover, the SSN family has a very remarkable wide range of the indices of skewness and kurtosis. Therefore, it is capable of describing more types of statistical data than many other known families.

Bivariate SS-normal distribution

Barakat's method (2015) can be applied on any baseline bivariate df $F_{\underline{X}}(\underline{x}) = F_{X_1,X_2}(x_1,x_2)$ to get the following bivariate stable symmetric family:

$$
\begin{aligned}
F_{\underline{\xi}}(\underline{x};\alpha,c) &= F_{\xi_1,\xi_2}(\underline{x};\alpha,c) = \alpha F_{\underline{X}}(x_1+c,x_2+c) \\
&+ \bar{\alpha} F_{-\underline{X}}(x_1-c,x_2-c) \\
&= \alpha F_{\underline{X}}(x_1+c,x_2+c) + \bar{\alpha}\tilde{F}_{\underline{X}}(-x_1+c,-x_2+c),
\end{aligned}
\tag{3.1}
$$

where, $0 \leq \bar{\alpha} = 1-\alpha \leq 1$, $c \in \mathbb{R}$, $-\underline{X} = (-X_1,-X_2)$ and $\tilde{F}_{\underline{X}}(\underline{x}) = 1 - F_{X_1}(x_1) - F_{X_2}(x_2) + F_{\underline{X}}(\underline{x})$.

The survivor function of the bivariate df (3.1) is given by:

$$
\begin{aligned}
\tilde{F}_{\underline{\xi}}(\underline{x};\alpha,c) &= 1 - F_{\xi_1}(x_1;\alpha,c) - F_{\xi_2}(x_2;\alpha,c) + F_{\underline{\xi}}(\underline{x};\alpha,c) \\
&= 1 - (\alpha+\bar{\alpha}) \\
&+ \alpha\left(1 - F_{X_1}(x_1+c) - F_{X_2}(x_2+c) + F_{\underline{X}}(x_1+c,x_2+c)\right) \\
&+ \bar{\alpha}\left(1 - \bar{F}_{X_1}(-x_1+c) - \bar{F}_{X_2}(-x_2+c) + \tilde{F}_{\underline{X}}(-x_1+c,-x_2+c)\right) \\
&= \alpha\tilde{F}_{\underline{X}}(x_1+c,x_2+c) + \bar{\alpha}F_{\underline{X}}(-x_1+c,-x_2+c),
\end{aligned}
$$

where, $\bar{F}_{X_i}(x_i) = 1 - F_{X_i}(x_i)$, $i = 1, 2$. Moreover, the pdf of the df (3.1) is given by:

$$
\begin{aligned}
f_{\underline{\xi}}(\underline{x};\alpha,c) &= \alpha f_{\underline{X}}(x_1+c,x_2+c) + \bar{\alpha} f_{-\underline{X}}(-x_1+c,-x_2+c) \\
&= \alpha f_{\underline{X}}(x_1+c,x_2+c) + \bar{\alpha} f_{-\underline{X}}(x_1-c,x_2-c).
\end{aligned}
$$

If the baseline df $F_{\underline{X}}$ is symmetric, i.e., $F_{\underline{X}}(\underline{x}) = F_{-\underline{X}}(\underline{x}) = \tilde{F}_{\underline{X}}(-\underline{x})$, the definition (3.1) becomes:

$$
F_{\underline{\xi}}(\underline{x};\alpha,c) = \alpha F_{\underline{X}}(x_1+c,x_2+c) + \bar{\alpha}F_{\underline{X}}(x_1-c,x_2-c). \tag{3.2}
$$

Let us now apply (3.2) on the standard bivariate normal distribution $\Phi_\rho(\underline{x})$, with correlation coefficient ρ. Then, a new skew bivariate normal distribution, denoted by BSSN, or $\underline{W} = (W_1, W_2) \sim$ BSSN, is defined by:

$$
F_{\underline{W}}(\underline{x};\alpha,c,\rho) = \alpha\Phi_\rho(x_1+c,x_2+c) + \bar{\alpha}\Phi_\rho(x_1-c,x_2-c). \tag{3.3}
$$

The moment generating function can be easily derived by using (3.3) in the form:

$$
M_{\underline{W}}^{(0,0)}(t,s) = E\left(e^{tW_1+sW_2}\right) = L_{\alpha,c}^{(0,0)}(t,s) + L_{\bar{\alpha},-c}^{(0,0)}(t,s), \tag{3.4}
$$

where:

$$
L_{\alpha,c}^{(0,0)}(t,s) = \alpha\exp\left[-(t+s)c + \frac{1}{2}\left(t^2 + 2\rho ts + s^2\right)\right].
$$

The following lemma, which is directly obtained by using (3.4), gives the marginal means, marginal variances, marginal coefficients of skewness and kurtosis of the BSSN family (3.3)

Lemma 3.1

For each of the marginals of the BSSN df, we have:

- The mean is $\mu_{W_i} = -c(2\alpha-1)$, $i = 1, 2$
- The variance is $\sigma_{W_i}^2 = 1 + 4c^2\alpha\bar{\alpha}$, $i = 1, 2$
- The coefficient of skewness is:

$$
\gamma_{W_i}^{[1]} = \frac{8c^3\alpha\bar{\alpha}(2\alpha-1)}{\left(1+4c^2\alpha\bar{\alpha}\right)^{\frac{3}{2}}}
\begin{cases}
= 0, & if\ \alpha = \frac{1}{2}\ or\ c = 0, \\
> 0, & if\ \alpha > \frac{1}{2}, c > 0\ or\ \alpha < \frac{1}{2}, c < 0, \\
< 0, & if\ \alpha > \frac{1}{2}, c < 0\ or\ \alpha < \frac{1}{2}, c > 0; i=1,2.
\end{cases}
$$

- The coefficient of kurtosis is:

$$
\gamma_{W_i}^{[2]} = \frac{3 + 4c^2\alpha\bar{\alpha}\left[c^2\left(1+3(2\alpha-1)^2+6\right)\right]}{\left(1+4c^2\alpha\bar{\alpha}\right)^2}
$$

$$
\begin{cases}
= 3, & if\ \alpha = \frac{3\pm\sqrt{3}}{6}, \\
< 3, & if\ \alpha \in \left(\frac{3-\sqrt{3}}{6},\frac{3+\sqrt{3}}{6}\right), \\
> 3, & if\ \alpha \notin \left(\frac{3-\sqrt{3}}{6},\frac{3+\sqrt{3}}{6}\right); i=1,2.
\end{cases}
$$

Clearly, if $c = 0$ the BSSN family becomes the standard bivariate normal df for any α. Thus, apart from the value $c = 0$, Lemma 3.1 shows that the parameter α solo controls the marginal kurtosis of the BSSN family. On the other hand, both the parameters α and c control the marginal skewness of the BSSN family in the

following manner: Each of the marginals of BSSN family has the type:

$$
\begin{cases}
symmetric\ and\ mesokurtuc,\ i\ e.,00, & if\ c = 0, \\[4pt]
summetric\ and\ platykurtic,\ i\ e.,0-, & if\ \alpha = \dfrac{1}{2}, \\[4pt]
positive\ skewness\ and\ mesokurtic,\ i\ e.,+0, & if\ \alpha = \dfrac{3+\sqrt{3}}{6}, c < 0, \\[4pt]
positive\ skewness\ and\ leptokurtic,\ i\ e.,++, & if\ \alpha > \dfrac{3+\sqrt{3}}{6}, c > 0, \\[4pt]
positive\ skewness\ and\ platykurtic,\ i\ e.,+-,& if\ \dfrac{1}{2} < \alpha < \dfrac{3+\sqrt{3}}{6}, c > 0, \\[4pt]
negative\ skewnedd\ and\ mesokurtic,\ i\ e.,-0, & if\ \alpha = \dfrac{3-\sqrt{3}}{6}, c > 0, \\[4pt]
negative\ skewness\ and\ leptokuric,\ i\ e.,-+, & if\ \alpha < \dfrac{3-\sqrt{3}}{6}, c < 0, \\[4pt]
negative\ skewness\ and\ platykurtic,\ i\ e.,--,& if\ \dfrac{3-\sqrt{3}}{6} < \alpha < \dfrac{1}{2}, c < 0.
\end{cases}
$$

It is worth noting that, each of the marginals of BSSN df has eight types out of the nine possible types of df's. Namely, the only excluded type from the BSSN family is the type 0+. Moreover, in view of the results of Barakat (2015) the BSSN family possesses a very wide range of the indices of marginal skewness and kurtosis. The following interesting theorem shows that the BSSN family is not only tractable but we can easily control its correlation coefficient.

Theorem 3.1

The correlation coefficient, denoted by $\rho_{\underline{W}}$, of the BSSN family is given by:

$$
\rho_{\underline{W}} = \frac{4c^2\bar{\alpha}\alpha + \rho}{4c^2\bar{\alpha}\alpha + 1} \geq \rho,
$$

for all $-1 \leq \rho \leq 1$, $c \in \mathbb{R}$ and $0 \leq \alpha \leq 1$, Moreover:

- $\rho_{\underline{W}} = \rho$, if and only if $\rho = 1$, or $\bar{\alpha}\alpha c^2 = 0$
- $\rho_{\underline{W}} = 1$, if and only if $\rho = 1$
- $\rho_{\underline{W}} = -1$, if and only if $\rho = -1$ and $\bar{\alpha}\alpha c^2 = 0$
- $\rho_{\underline{W}} = 0$, if and only if $-1 \leq \rho = -4\bar{\alpha}\alpha c^2 \leq 0$

Proof

First, let us adopt the conventions $\dfrac{\partial L_{\alpha,c}^{(0,0)}(t,s)}{\partial t} = L_{\alpha,c}^{(1,0)}(t,s)$, $\dfrac{\partial L_{\alpha,c}^{(0,0)}(t,s)}{\partial s} = L_{\alpha,c}^{(0,1)}(t,s)$ and $\dfrac{\partial^2 L_{\alpha,c}^{(0,0)}(t,s)}{\partial t \partial s} = L_{\alpha,c}^{(1,1)}(t,s)$. Clearly:

$$
L_{\alpha,c}^{(1,0)}(t,s) = L_{\alpha,c}^{(0,0)}(t,s)[t + s\rho - c]
$$

and:

$$
L_{\alpha,c}^{(1,1)}(t,s) = L_{\alpha,c}^{(0,0)}(t,s)\Big[\big(t + s\rho - c\big)\big(s + t\rho - c\big) + \rho\Big]
$$

Thus $L_{\alpha,c}^{(1,1)}(0,0) = \alpha(\rho + c^2)$ and $L_{\bar{\alpha},-c}^{(1,1)}(0,0) = \bar{\alpha}\big(c^2 + \rho\big)$. Therefore:

$$
E\big(W_1 W_2\big) = M_{\underline{W}}^{(1,1)}(0,0) = L_{\alpha,c}^{(1,1)}(0,0) + L_{\bar{\alpha},-c}^{(1,1)}(0,0) = c^2 + \rho. \qquad (3.5)
$$

On combining (3.5) with the first and second results of Lemma 3.1, we get:

$$
\rho_{\underline{W}} = \frac{COV\big(W_1, W_2\big)}{\sigma_{W_1}\sigma_{W_2}} = \frac{\rho + 4\bar{\alpha}\alpha c^2}{1 + 4\alpha\bar{\alpha}c^2}.
$$

On the other hand, since $4\bar{\alpha}\alpha c^2 \geq 0$, the inequality $\rho_{\underline{W}} \geq \rho$ is obvious, as well as the remaining implications concerning $\rho_{\underline{W}}$. This completes the proof of the theorem. For any fitting data purpose, it is more convenient to consider the location-scale BSSN df:

$$
\begin{aligned}
& F_{\underline{W}}\big(\underline{x}; \alpha, c, \rho, \mu_1, \mu_2, \sigma_1, \sigma_2\big) \\
&= \alpha\Phi_\rho\left(\frac{x_1 - \mu_1}{\sigma_1} + c, \frac{x_2 - \mu_2}{\sigma_2} + c\right) \\
&\quad + \bar{\alpha}\Phi_\rho\left(\frac{x_1 - \mu_1}{\sigma_1} - c, \frac{x_2 - \mu_2}{\sigma_2} - c\right) \\
&= \alpha\Phi_\rho\left(\frac{x_1 - c_1}{\sigma_1}, \frac{x_2 - c_2}{\sigma_2}\right) + \bar{\alpha}\Phi_\rho\left(\frac{x_1 - c_1^\star}{\sigma_1}, \frac{x_2 - c_2^\star}{\sigma_2}\right),
\end{aligned} \qquad (3.6)
$$

where, $c_1 = \mu_1 - c\sigma_1$, $c_1^\star = \mu_1 + c\sigma_1$, $c_2 = \mu_2 - c\sigma_2$ and $c_2^\star = \mu_2 + c\sigma_2$. Moreover, in any fitting problem we should get the estimates $\hat{\alpha}, \hat{\rho}, \hat{c}_1, \hat{c}_2, \hat{c}_1^\star, \hat{c}_2^\star, \hat{\sigma}_1$ and $\hat{\sigma}_2$. Clearly, the estimation problem for these parameters is strongly relevant to the problem of estimating mixtures of bivariate normal distributions. Actually, many efficient methods were discussed by many authors, among them we mention Quandt and Ramsey (1978) and Lin *et al.* (2007).

Further Extension for BSSN Family

Barakat and Khaled (2017) called any generalized family that contains all the possible types of df's (nine types), a full family. As we have seen before the BSSN family has eight marginal types out of the nine possible types of df's. Namely, the only excluded marginal type from the BSSN family is the type 0+. Since the mixture operation is linear with respect to all the moments, we can suggest a full bivariate family of df's by mixing the BSSN family with any tractable symmetric leptokurtic df, namely the standard bivariate logistic df suggested by Gumbel (1961). The standard bivariate logistic df is defined by:

$$L_U(\underline{u}) = P(U_1 \le u_1, U_2 \le u_2)$$

$$= \frac{1}{1 + e^{-u_1} + e^{-u_2}}, u_i \in (-\infty, \infty), i = 1, 2. \qquad (3.7)$$

Therefore, given three parameters $0 \le \alpha, \beta \le 1$ and $c \in \mathbb{R}$, the full family of df's is given by:

$$F_{\underline{\eta}}(\underline{x}; \alpha, \beta, c) = \beta F_U(\underline{x}; \alpha, c) + \bar{\beta} L_U(\underline{x})$$

$$= \beta[\alpha \Phi_\rho(x_1 + c, x_2 + c) + \bar{\alpha} \Phi_\rho(x_1 - c, x_2 - c)] \qquad (3.8)$$

$$+ \bar{\beta} L_U(\underline{x}).$$

The moment generating function can be easily derived by using (3.8) in the form:

$$M_{\underline{\eta}}^{(0,0)}(t, s) = E(e^{t\eta_1 + s\eta_2}) = \beta M_{\underline{W}}^{(0,0)}(t, s) + \underline{\beta} M_{\underline{U}}^{(0,0)}(t, s), \quad (3.9)$$

where, $M_{\underline{W}}^{(0,0)}(t, s)$ is defined by (3.4) and $M_{\underline{U}}^{(0,0)}(t, s) = \Gamma(1-t)\Gamma(1+t+s)\Gamma(1-s)$ is the moment generating function of the df (3.7) (cf. Gumbel, 1961). By using the moment generating function (3.9), we can compute all the marginal and the cross moments of the family (3.8). For examples:

$$E(\eta_i^k) = \begin{cases} \beta E(W_i^k), & \text{if } k \text{ is odd,} \\ \beta E(W_i^k) + \bar{\beta} E(U_i^k) & \text{if } k \text{ is even}, i = 1, 2, \end{cases}$$

$\sigma_{\eta_i}^2 = \beta[1 + c^2(1 - \beta(\alpha - \bar{\alpha})^2)] + \bar{\beta}\frac{\pi^2}{3}$, $i = 1, 2$, (note that $\sigma_{U_i} = \frac{\pi}{\sqrt{3}}$) and:

$$E(\eta_1 \eta_2) = \beta E(W_1 W_2) + \bar{\beta} E(U_1 U_2) = \beta(c^2 + \rho) + \bar{\beta}\frac{\pi^2}{6}.$$

Thus $COV(\eta_1, \eta_2) = \beta[\rho + c^2(1 - \beta(\alpha - \bar{\alpha})^2)] + \bar{\beta}\frac{\pi^2}{6}$, which implies:

$$\rho_{\underline{\eta}} = \frac{\beta\left[\rho + c^2\left(1 - \beta(\alpha - \bar{\alpha})^2\right)\right] + \bar{\beta}\frac{\pi^2}{6}}{\beta\left[1 + c^2\left(1 - \beta(\alpha - \bar{\alpha})^2\right)\right] + \bar{\beta}\frac{\pi^2}{3}}. \qquad (3.10)$$

Note that, in (3.10) if $\beta = 1$, then $\rho_{\underline{\eta}} = \rho_{\underline{W}}$. Moreover, if $\beta = 0$, then $\rho_{\underline{\eta}} = \rho_{\underline{U}}\frac{1}{2}$.

Clearly, each marginal of the FBSSN family possesses a very wide range of the indices of skewness and kurtosis (at least not less than the BSSN family). Moreover, each marginal of the FBSSN family contains all the possible types of df's. Therefore, the FBSSN family provides a general and flexible mechanism for fitting a wide spectrum of real world bivariate data set.

The location-scale FBSSN df is given by:

$$F_{\underline{\eta}}(\underline{x}; \alpha, c, \rho, \mu_1, \mu_2, \sigma_1, \sigma_2)$$

$$= \alpha\beta\Phi_\rho\left(\frac{x_1 - \mu_1}{\sigma_1} + c, \frac{x_2 - \mu_2}{\sigma_2} + c\right)$$

$$+ \bar{\alpha}\beta\Phi_\rho\left(\frac{x_1 - \mu_1}{\sigma_1} - c, \frac{x_2 - \mu_2}{\sigma_2} - c\right)$$

$$+ \bar{\beta} L_U\left(\frac{x_1 - \mu_1}{\sigma_1}, \frac{x_2 - \mu_2}{\sigma_2}\right) \qquad (2.11)$$

$$= \alpha_1 \Phi_\rho\left(\frac{x_1 - c_1}{\sigma_1}, \frac{x_2 - c_2}{\sigma_2}\right) + \alpha_2 \Phi_\rho\left(\frac{x_1 - c_1^\star}{\sigma_1}, \frac{x_2 - c_2^\star}{\sigma_2}\right)$$

$$+ \alpha_3 L_U\left(\frac{x_1 - \mu_1}{\sigma_1}, \frac{x_2 - \mu_2}{\sigma_2}\right),$$

where, $c_1 = \mu_1 - c\sigma_1$, $c_1^\star = \mu_1 + c\sigma_1$, $c_2 = \mu_2 - c\sigma_2$, $c_1^\star = \mu_2 + c\sigma_2$, $\alpha_1 = \alpha\beta$, $\alpha_2 = \bar{\alpha}\beta$ and $\alpha_3 = \bar{\beta}$.

Modeling of Air Pollution

Air pollution is a very important environmental and social issue. In addition, it is a complex problem posing multiple challenges in terms of management and reduction of harmful pollutants. Air pollutants are released from anthropogenic and natural sources.

In this section we consider three daily data sets for air pollution from the London Air Quality Network (LAQN), from 01/01/2010 to 31/12/2016. This network is a united resource for air pollution measurements, which are fundamental to support air quality administration. The LAQN was framed in 1993 to arrange and enhance air pollution checking in London. People living in the city of London can check the updates of the level of air pollution data through the web site "www:londonair:org:uk". We consider the following three pollutants: The Nitric Oxide (NO), the Nitrogen Dioxide (NO_2) and the Sulphur Dioxide (SO_2). These data sets can be downloaded by any researcher in the form of a report every half hour, every hour or every day according to the type of study from the following site "www.londonair.org.uk/london/asp/datadownload.asp". The summary statistics for these data sets are given in Table 1.

In this study, we compare the location-scale BSSN and FBSSN families (3.6) and (3.11) with the location-scale BT-*g-h* family $F_{\underline{x}}\left(\frac{x_1 - \mu}{\sigma_1}, \frac{x_2 - \mu_2}{\sigma_2}; \underline{h}, \underline{g}\right)$ (where $\sigma > 0$ and

the rv $\underline{X}$ is defined by (2.2)), the location-scale BSN2 family $F_{\underline{V}}^{\langle 2\rangle}\left(\dfrac{x_1 - \mu_1}{\sigma_1}, \dfrac{x_2 - \mu_2}{\sigma_2}; \underline{\lambda}\right)$, (where its pdf is defined by (2.5)), the location-scale bivariate mixture normal df $BMN(\underline{x}; \ \alpha, \ c, \ \rho, \ \underline{\mu}, \ \underline{\sigma}) = \alpha \Phi_\rho\left(\dfrac{x_1 - \mu_1}{\sigma_1}, \dfrac{x_2 - \mu_2}{\sigma_2}\right) + \bar{\alpha}\Phi_\rho\left(\dfrac{x_1 - \mu_3}{\sigma_3}, \dfrac{x_2 - \mu_4}{\sigma_4}\right)$ (the correlation coefficient of MBN is denoted by ρ_{MBN}) and the location-scale bivariate normal df $BN(\underline{z}; \rho, \underline{\mu}, \underline{\sigma}) = \Phi_\rho\left(\dfrac{z_1 - \mu_1}{\sigma_2}, \dfrac{z_1 - \mu_1}{\sigma_2}\right)$ (with correlation coefficient $\rho_{BN} = \rho$). The estimates of the parameters of these families are calculated using the R package as

follows: For BT-*g-h* we use ghyp package, for BSN2 we use sn package, for BN families we use stats package, for BSSN, BMN families we use maxLik package and for FBSSN we use maxLik and VGAM packages. These estimates are summarized in Tables 2-4. To compare the fittings of these families we used Akaike Information Criterion (AIC) (Akaike, 1973) and the Bayesian Information Criterion (BIC) (Schwarz, 1978). These criteria are based on the likelihood value of the model, the number of observations and the number of parameters thereof. The comparison results are summarized in Table 5-7. These tables show that the FBSSN family has the smallest values of the AIC and BIC criteria in the three tables, followed by the BSSN family. On the contrary, the families BT-*g-h* and BSN$_2$ have the largest values of those criteria.

Table 1: Summary statistics

Pollution	n	Descriptive statistics for air pollutants						
		Minimum	Maximum	Median	Mean	SD	Skewness	Kurtosis
Nitric oxide	2138	0	279.00	3.600	8.704	16.30	6.24	66.84
Nitrogen Dioxide	2138	1	105.40	21.900	24.630	13.15	1.15	5.02
Sulfur Dioxide	2138	0	28.10	2.400	2.635	1.52	5.46	75.89

Table 2: Parameter estimation for (NO, NO$_2$)

ML parameters estimation

BT-*g-h* family	h_1	h_2	g_1	g_2				μ_1	μ_2					σ_1	σ_2	$\rho_{\underline{X}}$
	0.05	4.09	6.98	4.74				2.14	20.18					11.34	12.91	0.65
BSN$_2$ family	λ_1		λ_2					μ_1	μ_2					σ_1	σ_2	$\rho_{\underline{V}}$
	3786893.35		-35996.82					0.01	19.87					18.47	13.98	0.73
BSSN family		α_1	α_2		c_1	$c_1^\star$	c_2	$c_2^\star$						σ_1	σ_2	$\rho_{\underline{W}}$
		0.30	0.70		15.02	15.98	7.93	10.06						20.00	11.98	0.69
FBSSN family	α_1	α_2	α_3		c_1	$c_1^\star$	c_2	$c_2^\star$	μ_1	μ_2				σ_1	σ_2	$\rho_{\underline{n}}$
	0.002	1.8e-02	0.98		40.27	15.80	7.85	10.07	14.6	19.9				50.03	12.1	0.49
BMN family	α_1		α_2			μ_1	μ_2	μ_3	μ_4		σ_1	σ_2	σ_3	σ_4		ρ_{MBN}
	0.001		0.999			8.36	4.75	5.95	4.28		1.72	14.49	11.02	10.65		0.20
BN df								μ_1	μ_2					σ_1	σ_2	ρ_{BN}
								8.70	24.63					16.30	13.15	0.68

Table 3: Parameter estimation for (NO, SO$_2$)

ML parameters estimation

BT-*g-h* family	h_1	h_2	g_1	g_2				μ_1	μ_2					σ_1	σ_2	$\rho_{\underline{X}}$
	0.03	4.05	6.77	0.39				2.66	2.29					11.48	1.33	0.26
BSN$_2$ family	λ_1		λ_2			μ_1		μ_2			σ_1		σ_2			$\rho_{\underline{V}}$
	13253651.9		14891.3			-0.007752721		2.373941			18.48162		1.53843			0.358536
BSSN family		α_1	α_2			c_1	$c_1^\star$	c_2	$c_2^\star$					σ_1	σ_2	$\rho_{\underline{W}}$
		0.69	0.31			10.96	2.60	9.85	2.48					16.98	2.19	0.48
FBSSN family	α_1	α_2	α_3		c_1	$c_1^\star$	c_2	$c_2^\star$	μ_1	μ_2				σ_1	σ_2	$\rho_{\underline{n}}$
	0.93	0.05	0.015		14.93	10.05	18.16	17.09	16.74	20.25				22.99	5.94	0.45
BMN family		α_1	α_2			μ_1	μ_2	μ_3	μ_4		σ_1	σ_2	σ_3	σ_4		ρ_{BMN}
		0.997	0.003			9.88	10.72	10.45	17.04		14.63	15.72	7.88	12.47		0.37
BN df								μ_1	μ_2					σ_1	σ_2	ρ_{BN}
								8.70	2.63					16.30	1.51	0.32

Table 4: Parameter estimation for (NO_2, SO_2)

ML parameters estimation

BT-g-h family	h_1	h_2	g_1	g_2				μ_1	μ_2					σ_1	σ_2	ρ_X
	0.18	4.30	3.53	0.31				21.23	2.34					13.31	1.31	0.30
BSN$_2$ family	λ_1	λ_2						μ_1	μ_2					σ_1	σ_2	$\rho_{\underline{V}}$
	0.38	3.84						18.83	1.05					14.37	2.19	0.49
BSSN family	α_1	α_2			c_1	$c_1^\star$	c_2	$c_2^\star$						σ_1	σ_2	$\rho_{\underline{W}}$
	0.05	0.95			8.47	9.61	11.55	15.531						15.12	11.83	0.47

FBSSN family	α_1	α_2		α_3		c_1	$c_1^\star$	c_2	$c_2^\star$	μ_1	μ_2			σ_1	σ_2	$\rho_{\text{ŋ}}$
	0.007	2.03e-01		0.79		43.45	15.17	8.06	10.59	15.39	20.24			29.18	10.39	0.42

BMN family			α_1	α_2		μ_1	μ_2	μ_3	μ_4			σ_1	σ_2	σ_3	σ_4	ρ_{BMN}
			0.99	0.01		8.42	11.35	9.34	15.46			15.37	10.47	8.92	13.59	0.18
BN df						μ_1		μ_2						σ_1	σ_2	ρ_{BN}
						24.63		2.63						13.15	1.52	0.31

Table 5: AIC and BIC for (NO, NO_2)

	BT-*g-h* family	BSN$_2$ family	BSSN family	FSSBN family	BMN family	BN df
AIC	31252.1	31316.18	15627.55	13976.85	22788.66	25884.06
BIC	31297.44	31350.19	15672.89	14039.2	22845.34	25935.4

Table 6: AIC and BIC for (NO, SO_2)

	BT-*g-h* family	BSN$_2$ family	BSSN family	FBSSN family	BMN family	BN df
AIC	22531.32	23202.76	11093.45	8066.54	11201.42	17768.64
BIC	22576.66	23236.77	11138.79	8128.884	11258.09	17819.98

Table 7: AIC and BIC for (NO_2, SO_2)

	BT-*g-h* family	BSN$_2$ family	BSSN family	FSSN family	BMN family	BN df
AIC	23625.52	24050.74	13377.9	11692.85	14145.68	16866.95
BIC	23670.86	24084.75	13423.24	11755.19	14202.35	16918.29

Therefore, the best model that fits these data sets is the FBSSN family and the second model is the BSSN family. On the other hand, the worst two models that fit these data sets are BT-*g-h* and BSN$_2$ families.

Concluding Remark

In this study, we adhered to the most agreed upon theoretical conditions that make any family of bivariate df's to be capable of fitting many different types of bivariate data. According to these theoretical conditions we constructed the BSSN family and extended it to the FBSSN family, which is more a capable family to describe many bivariate data sets of different types. We applied the constructed families with other popular families on three bivariate data sets concerning three air pollutants. We used the AIC and BIC criteria to compare the fittings of these families. The result of the comparisons shows that the FBSSN family is the best followed by the BSSN family.

Acknowledgement

Author's Contributions

Haroon M. Barakat and Osama Mohareb Khaled: With the consultation of each other carried out this work and drafted the manuscript together.

Norhan Khalil: Carried out all the computations of the paper and made the code of calculations. All authors read and approved the final manuscript.

Ethics

The author declares that there is no conflict of interests regarding the publication of this article.

References

Akaike, H., 1973. Information theory and the Maximum Likelihood Principle. Proceedings of the 2nd International Symposium on Information Theory, (SIT' 73), Akademiai Kiado, Budapest, pp: 267-281.

Azzalini, A., 1985. A class of distributions which includes the normal ones. Scandinavian J. of Stat., 12: 171-178.

Azzalini, A., 2005. The skew-normal distribution and related multivariate families. Scandinavian J. Stat., 32: 159-188. DOI: 10.1111/j.1467-9469.2005.00426.x

Azzalini, A. and A. Capitanio, 1999. Statistical applications of the multivariate skew normal distribution. J. Royal Statist. Soc., 3: 579-602. DOI: 10.1111/1467-9868.00194

Azzalini, A. and A. Dalla Valle, 1996. The multivariate skew-normal distribution. Biometrika, 83: 715-726. DOI: 10.1093/biomet/83.4.715

Barakat, H.M., 2015. A new method for adding two parameters to a family of distributions with application to the normal and exponential families. Stat. Meth. Applic. 24: 359-372. DOI: 10.1007/s10260-014-0265-8

Barakat, H.M. and O.M. Khaled, 2017. Toward the establishment of a family of distributions that may fit any dataset. Commun. Stat. Simul. Comput., 46: 6129-6143. DOI: 10.1080/03610918.2016.1197245

Cook, D. and M.E. Johnson, 1981. A family of distributions for modelling non-elliptically symmetric multivariate data. J. Royal Stat. Soc., 43: 210-218.

Cook, D. and M.E. Johnson, 1986. Generalized Burr-Pareto-logistic distributions with applications to a uranium exploration data set. Technometrics, 28: 123-131. DOI: 10.1080/00401706.1986.10488113

Edgeworth, F.Y., 1896. The compound law of error. Philosophical Magazine, 41: 207-215. DOI: 10.1080/14786449608620839

Hoaglin, D.C., 1983. Summarizing shape numerically: The g-and-h distributions.

Hoaglin, D.C. and S.C. Peter, 1979. Software for sxploring distributional shapes. Proceedings of the Computer Science and Statistics: 12th Annual Symposium on the Interface, Iniverty of Waterloo, Ontario, Canada, pp: 418-443.

Gumbel, E.J., 1961. Bivariate logistic distributions. J. Am. Stat. Assoc., 56: 335-349. DOI: 10.1080/01621459.1961.10482117

Gupta, A.K. and J.T. Chen, 2001. Goodness-of-fit tests for the skew-normal distribution. Commun. Stat., Simul. Comput., 30: 907-930. DOI: 10.1081/SAC-100107788

Gupta, A.K., F.C. Chang and W.J. Huang, 2002. Some skew-symmetric models. Random Oper. Stoch. Eq., 10: 133-140. DOI: 10.1515/rose.2002.10.2.133

Gupta, A.K. and J.T. Chen, 2003. On the sample characterization criterion for normal distributions. J. Stat. Comput. Simul., 73: 188-163. DOI: 10.1080/00949650215867

Gupta, A.K. and J.T. Chen, 2004. A class of multivariate skew-normal models. Annals Ins. Stat. Math., 56: 305-315. DOI: 10.1007/BF02530547

Kowalchuk, R.K. and T.C. Headrick, 2010. Simulating multivariate *g*-and-*h* distributions. British J. Math. Stat. Psychol., 63: 63-74. DOI: 10.1348/000711009X423067

Lin, T.I., J.C. Lee and S.Y. Yen, 2007. Finite mixture modelling using the skew normal distribution. Stat. Sinica, 17: 909-927.

Mardia, K.V., 1970. Families of Bivariate Distributions. 1st Edn., Griffin, London, pp: 109.

Martinez, J., 1981. Some properties of robust scale estimators. Ph.D. dissertation, Temple University, Philadelphia, PA.

Martinez, J. and B. Iglewicz, 1984. Some properties of the Tukey g and h family of distributions. Commun. Stat.: Theory Meth., 13: 353-369. DOI: 10.1080/03610928408828687

Quandt, R.E. and J.B. Ramsey, 1978. Estimating mixtures of normal distributions and switching regressions. J. Amer. Stat. Assoc., 73: 730-738. DOI: 10.1080/01621459.1978.10480085

Schwarz, G., 1978. Estimating the dimension of a model. Ann. Stat., 6: 461-464. DOI: 10.1214/aos/1176344136

Tukey, J.W., 1977. Modern techniques in data analysis. Proceedings of the NSF Sponsored Regional Research Conference at Southern Massachusetts University, (SMU' 77), North Dartmouth, MA.

Wilcox, R.R., 1995. Simulation results on solutions to the multivariate behrens-fisher problem via trimmed means. Statistician, 44: 213-225. DOI: 10.2307/2348445

Wilcox, R.R. and H.J. Keselman, 2001. Using trimmed means to compare k measures corresponding to two independent groups. Multivariate Behav. Res., 36: 421-444. DOI: 10.1207/S15327906421-444

Wilcox, R.R. and H.J. Keselman, 2003. Repeated measures one-way ANOVA based on a modified one-step M-estimator. British J. Math. Stat. Psychol., 56: 15-25. DOI: 10.1348/000711003321645313

Wilcox, R.R., H.J. Keselman, J. Muska and R. Cribbie, 2000. Repeated measures ANOVA: Some new results on comparing trimmed means and means. British J. Math. Stat. Psychol., 53: 69-82. DOI: 10.1348/000711000159187