Yet another skew-elliptical family but of a different kind: return to Lemma 1

Adelchi Azzalini  
Dipartimento di Scienze Statistiche  
Università di Padova  
Italia

Giuliana Regoli  
Dipartimento di Matematica e Informatica  
Università di Perugia  
Italia

28th September 2018

Abstract

In the context of modulated-symmetry distributions, there exist various forms of skew-elliptical families. We present yet another one, but with an unusual feature: the modulation factor of the baseline elliptical density is represented by a distribution function with an argument which is not an odd function, as it occurs instead with the overwhelming majority of similar formulations, not only with other skew-elliptical families. The proposal is obtained by going back to the use of Lemma 1 of Azzalini and Capitanio (1999), which can be seen as the general frame for a vast number of existing formulations, and use it on a different route. The broader target is to show that this ‘mother lemma’ can still generate novel progeny.

Some key-words: elliptical distributions, skew-elliptical distributions, symmetry-modulated distributions.
1 Background and aims

In the last fifteen years or so, there has been a formidable development on the theme of ‘symmetry-modulated distributions’, also called ‘skew-symmetric distributions’. In the context of continuous random variables, the key tool for building a symmetry-modulated distribution starting from a centrally symmetric density \( f_0 \) on \( \mathbb{R}^d \), that is, such that \( f_0(x) = f_0(-x) \), is represented by the expression

\[
f(x) = 2 f_0(x) G_0(w(x)), \quad x \in \mathbb{R}^d,
\]

where \( G_0 \) is a univariate continuous distribution function such that \( G_0(-x) = 1 - G_0(x) \), that is, it establishes to a symmetric distribution about 0, and \( w(\cdot) \) is a real-valued function such that \( w(-x) = -w(x) \). Under the stated conditions, for any 'baseline' density \( f_0 \) and any choice of the ingredients \( G_0 \) and \( w \) which enter the final 'modulating term' of \( f(x) \), this is a proper density.

The above passage has reproduced Proposition 1 of Azzalini & Capitanio (2003). A slightly different formulation, essentially equivalent to the one above, has been presented by Wang et alii (2004). Even a brief summary of the developments stemming from the symmetry-modulated distributions in (1) would take an enormous space. For such objective, the reader is referred to the recent account by Azzalini & Capitanio (2014). Here we limit ourselves to recall the facts directly related to our plan of work.

The fact that (1) constitutes a density function is actually a corollary of a more general result, namely Lemma 1 of Azzalini & Capitanio (1999); its exact statement will be recalled in the next section. Right after this lemma, Azzalini and Capitanio formulated an immediate corollary by taking \( f_0 \) of elliptical type and \( w \) of linear type. That corollary prompted a stream of literature focusing on successive layers of generalizations, where \( f_0 \) fulfils some symmetry condition and \( w \) is odd, eventually leading to symmetry-modulated distributions.

There exist, however, other constructions encompassed by the above-quoted Lemma 1, falling outside the domain of symmetry-modulated distributions in (1). This territory has been barely explored. As far as we known, the only two publications of this sort are those of Azzalini (2012) and Jupp et alii (2016). The broad target of the present note is to investigate further this area. In spite of the common domain of interest, operationally there is little overlap between our development and the publications just quoted.

The present specific contribution, developed along the indicated direction, is constituted by a new proposal of skew-elliptical or, equivalently, skew-elliptically contoured (SEC) density. A number of SEC constructions in the form of symmetry modulated densities, obtained by some form of perturbation of an elliptically contoured (EC) density, have already been examined in the literature. Besides the above-mentioned 'linear-type' SEC, the more commonly adopted form of SEC is the one introduced by Branco & Dey (2001) via a selection mechanism applied to an elliptical distribution. In their exposition, it was not evident that their SEC distribution had a structure like (1), but this was shown in some important special cases by Azzalini & Capitanio (2003) and later in general by Azzalini & Regoli (2012, p.872). An additional form of SEC distribution has been presented by Sahu et alii (2003).
There is structural difference between the existing SEC distributions and the one to be presented here. Because of its genesis, the new SEC is not constrained by the conditions pertaining to [1]. It obviously retains the fact that \( f_0 \) is of elliptical class, otherwise the term SEC would not apply to it, but its \( w(\cdot) \) does not satisfy the condition \( w(-x) = -w(x) \).

2 A ‘mother’ lemma

The wording of the next statement is slightly different from the one in the original source, but completely equivalent to it. The expression ‘symmetric about 0’ is an abbreviated term for ‘symmetrically distributed about 0’, in case of a univariate random variable. We also use the term ‘symmetric about 0’ for a univariate distribution function \( H \) such that \( H(-x) = 1 - H(x) \).

**Lemma 1 (Azzalini & Capitanio, 1999)** Denote by \( G_0 \) the continuous distribution function of a univariate random variable symmetric about 0, by \( w(\cdot) \) a real-valued function on \( \mathbb{R}^d \) and by \( Z_0 \) a d-dimensional variable with density function \( f_0 \), such that \( W = w(Z_0) \) is symmetric about 0. Then

\[
f(x) = 2f_0(x) G_0(w(x)), \quad x \in \mathbb{R}^d,
\]

is a density function.

At first sight, the distribution in (2) coincides with (1). While the mathematical expressions look the same, the meaning of their symbols, and consequently their mathematical meaning, are largely different:

- in (2), \( f_0 \) is not required to satisfy the condition \( f_0(-x) = f_0(x) \);
- in (2), \( w \) is not required to satisfy the condition \( w(-x) = -w(x) \);
- the only common assumption is that \( G_0(-x) = 1 - G_0(x) \).

The key point for applying Lemma 1 is that \( W = w(Z_0) \) is symmetric about 0. Once this holds, any \( G_0 \) symmetric about 0 can be adopted to build a valid density. This scheme was particularly simple in the first application of the lemma, namely Corollary 2 of Azzalini & Capitanio (1999): if \( Z_0 \) is elliptical centred at 0, then the linear transform \( w(Z_0) = a^\top Z_0 \) is symmetric about 0 for any choice of the vector \( a \) of coefficients and this suffices for applying the lemma, irrespective of the choice of \( G_0 \). It only takes a little more effort along the same line of argument to show that \( f(x) \) in (1) is a proper density.

We also recall the stochastic representation for distribution (2) stated by Azzalini & Capitanio (1999, p. 599). If \( X \) is a d-dimensional variable with density \( f_0 \) and \( T \) is an independent variable with distribution function \( G_0 \), satisfying the conditions of Lemma 1, then

\[
Z = \begin{cases} 
X & \text{if } T \leq w(X) \\
-X & \text{otherwise}
\end{cases}
\]
has distribution [2]

Besides symmetry-modulated distributions, Lemma 1 embraces other constructions, although so far these have little been studied, as already mentioned. This overarching capability of Lemma 1 explains the term ‘mother lemma’ which we have employed. The aim of the present note is to examine a case of this sort, specifically when $f_0$ is an elliptical density.

3 Functions of orthogonal elliptical components

We want to examine a construction where the density function $f_0$ of $Z_0$ is of EC type and $W = w(Z_0)$ is symmetric about 0.

Start by considering the case where $f_0$ is a bivariate elliptically contoured distribution with ‘standardized’ marginals. Specifically, assume that $f_0$ is the density function of

$$(X, Y) \sim \text{EC}_2 \left(0, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}, \psi \right)$$

following the notation adopted by Fang et alii (1990), up to a minor change in the symbol of the ‘generator’, $\psi$ for us.

**Proposition 2** Given the continuous random variable $(X, Y)$ with distribution as in (4), the transformed variable

$$W = (Y - \rho X) h(X)$$

has is symmetric about 0 for any choice of the real-valued function $h$.

**Proof** From Theorem 2.18 of Fang et alii (2012), the conditional distribution of $W$ given $X = x$ is

$$(W|X = x) \sim \text{EC}_1(0, (1 - \rho^2) h(x)^2, \psi_x)$$

where $\psi_x$ is another generator, which in general depends on $x$, but its explicit expression is irrelevant to us. All we need is that the corresponding density $f_{W|X}(w)$ is symmetric about 0, so that $f_{W|X}(w) = f_{W|X}(-w)$. Therefore the unconditional density of $W$ is

$$f_W(w) = \int_{\mathbb{R}} f_{W|X}(w) f_{0X}(x|x) dx$$

$$= \int_{\mathbb{R}} f_{W|X}(-w) f_{0X}(x|x) dx$$

$$= f_W(-w).$$

by integration with respect to the distribution $f_{0X}(x)$ of the X component. QED

This result can immediately be extended to a general $d$-dimensional continuous random variable with EC distribution ($d > 1$), whose components are partitioned as follows:

$$(X, Y) \sim \text{EC}_d (\mu, \Sigma, \psi), \quad \mu = \begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}$$

(6)
where now $X$ is $(d - 1)$-dimensional; $W$ takes now the more general form

$$W = (Y - m_Y(X)) h(X) \quad \text{or} \quad W = \frac{Y - m_Y(X)}{\sigma_{22.1}^{1/2}} h(X)$$  \hspace{1cm} (7)

where

$$m_Y(X) = \mu_y + \sigma_{21} \Sigma_{11}^{-1}(X - \mu_x) = \beta_0 + \beta^\top X, \text{ say,} \hspace{1cm} (8)$$
$$\sigma_{22.1} = \sigma_{22} - \sigma_{21} \Sigma_{11} \sigma_{12} \hspace{1cm} (9)$$

are as in equation (2.42) of Fang et alii (2012). In case the mean value of $Y$ conditionally on $X$, it coincides with (8). The second form of (7) is just a minor variant of the first one, but it is conceptually convenient to have the first factor ‘standardized’, in the sense of being free from scale factors, besides the location parameter. Finally, note that symmetry about 0 persists if $Y - m_Y(X)$ is transformed to $w_0(Y - m_Y(X))$ by an odd function $w_0$.

We can summarize the combination of the above discussion and Lemma I into the following conclusion.

Proposition 3 Denote by $f_0(x, y)$ the $d$-dimensional elliptically contoured density of the random variable $(X, Y)$ in (6), by $G_0$ a continuous distribution function such that $G_0(-x) = 1 - G_0(x)$ and by $w(x, y)$ either of the two forms

$$w(x, y) = w_0(y - m_Y(x)) h(x), \quad w(x, y) = w_0\left(\frac{y - m_Y(x)}{\sigma_{22.1}^{1/2}}\right) h(x)$$  \hspace{1cm} (10)

for any function $h(x)$ from $\mathbb{R}^{d-1}$ to $\mathbb{R}$ and any odd function $w_0$ on the real line; here $m_Y(\cdot)$ and $\sigma_{22.1}$ are given by (8) and (9). Then

$$f(x, y) = 2 f_0(x, y) G_0[w(x, y)]$$  \hspace{1cm} (11)

is a proper density function.

Some example bivariate densities $f(x, y)$ of the type introduced in Proposition 3 are displayed in Figure 1 in the form of contour level plots. The second form of (10) has been used and the other ingredients are as follows: $f_0(x) = \varphi_2(x; \Sigma)$ where $\Sigma$ is the same matrix occurring in (4). $G_0$ is the standard Cauchy distribution function and

$$h(x) = \frac{h_N(x)}{h_0(x)} = \frac{1 + a_1 x + a_2 x^2}{1 + b_1 x + b_2 x^2}, \quad w_0(x) = \frac{c_1 x + c_3 x^3}{1 + c_2 x^2}$$  \hspace{1cm} (12)

for various choices of the coefficients $a, b, c$ and the correlation $\rho$ in $\Sigma$. The specific choices of the coefficients $a, b, c$ are indicated at the top of each pane of Figure 1. The plots indicate a wide flexibility of the family of distribution, even employing a relatively limited number of coefficients in the ratios (12).

The general stochastic representation (3) applies directly to distribution (11).
Figure 1: Example plots of densities defined in Proposition 3 using functions in (12).
4 The unperturbed component

Consider the \(d\)-dimensional density

\[
f(x, y) = 2f_0(x, y)G_0((y - \beta_0 - \beta^\top x) h(x)), \quad x \in \mathbb{R}^{d-1}, \ y \in \mathbb{R}, \tag{13}
\]

where \(f_0\) is as in (6) and we use the expression in (8). If \((\tilde{X}, \tilde{Y})\) is a random variable with density (13), the marginal distribution of \(\tilde{X}\) is

\[
f_{\tilde{X}}(x) = \int_{\mathbb{R}} f(x, y) dy
\]

\[
= \int_{\mathbb{R}} 2f_{0,X}(x) f_{0,Y\mid X}(y) G_0((y - \beta_0 - \beta^\top x) h(x)) dy
\]

\[
= f_{0,X}(x) \int_{\mathbb{R}} 2f_{0,Y\mid X}(z) G_0(z \sigma_{22} h(x)) dz
\]

on recalling (i) the expressions of the conditional mean (8) and the scale factor (9), (ii) the fact that \(f_{0,Y\mid X}(\cdot)\) is symmetric about its mean (conditional) value, (iii) Lemma 1 of Azzalini (1985).

Therefore the marginal distribution \(f_{\tilde{X}}\) of the first \(d-1\) component variables after the perturbation operation is equal to the original unperturbed density \(f_{0,X}\).

5 Moment generating function in a simple case

We want to compute the moment generating function (MGF) for the basic case (4) of normal type, \(W\) as in (5) and \(G_0 = \Phi\). In other words, consider the density

\[
f(x, y) = 2\varphi(x, y; \rho)\Phi((y - \rho x) h(x)), \quad (x, y) \in \mathbb{R}^2. \tag{14}
\]

For notational convenience, we modify slightly the notation and, from now on, we use \((X, Y)\) to denote a random variable with distribution (14). The MGF of \((X, Y)\) is

\[
M(t_1, t_2) = \int_{\mathbb{R}} \int_{\mathbb{R}} \exp(t_1 x + t_2 y) f(x, y) x y
\]

\[
= 2 \int_{\mathbb{R}} e^{t_1 x} M_c(t_2, x) \varphi(x) x
\]

where

\[
M_c(t_2, x) = \int_{\mathbb{R}} e^{t_2 y} f_{0,Y\mid X}(y) \Phi((y - \rho x) h(x)) y
\]

\[
= \int_{\mathbb{R}} e^{t_2 y} \varphi \left( \frac{y - \rho x}{\sqrt{1 - \rho^2}} \right) \frac{1}{\sqrt{1 - \rho^2}} \Phi((y - \rho x) h(x)) y
\]

\[
= \exp(t_2 \rho x) \int_{\mathbb{R}} \exp \left( t_2 z \sqrt{1 - \rho^2} \right) \varphi(z) \Phi \left( z \sqrt{1 - \rho^2} h(x) \right) z
\]
\[
\begin{align*}
= \exp(t_2 \rho x) \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2} \left[ z^2 - 2 t_2 z \sqrt{1 - \rho^2} \pm t_2^2 (1 - \rho^2) \right] \right\} \Phi \left( z \sqrt{1 - \rho^2} h(x) \right) z \\
= \exp \left[ t_2 \rho x + \frac{1}{2} t_2^2 (1 - \rho^2) \right] \int_{\mathbb{R}} \varphi(u) \Phi \left\{ (u + t_2 \sqrt{1 - \rho^2}) h(x) \right\} u \\
= \exp \left[ t_2 \rho x + \frac{1}{2} t_2^2 (1 - \rho^2) \right] \Phi \left( \frac{t_2 (1 - \rho^2) h(x)}{\sqrt{1 + (1 - \rho^2) h(x)^2}} \right)
\end{align*}
\]

where the last equality follows from Corollary 1 of Ellison’s (1964) Theorem 2. Therefore

\[
M(t_1, t_2) = 2 \exp \left\{ \frac{1}{2} t_2^2 (1 - \rho^2) \right\} \int_{\mathbb{R}} \exp(x(t_1 + t_2 \rho)) \Phi \left( \frac{t_2 (1 - \rho^2) h(x)}{\sqrt{1 + (1 - \rho^2) h(x)^2}} \right) \varphi(x) x.
\]

The above integral does not lend itself to explicit solution. To compute the moment of the \(X\) and \(Y\) components of (14), we exchange the integration and differentiation steps; hence consider

\[
\left. \frac{\partial M}{\partial t_1} \right|_{t_1=t_2=0} \quad \text{and} \quad \left. \frac{\partial M}{\partial t_2} \right|_{t_1=t_2=0}
\]

with differentiation under the integration sign. The first expression leads to \(E(X) = 0\), as expected, since the marginal density of \(X\) is \(\varphi(x)\), from §4. The second expression leads to

\[
E(Y) = \sqrt{\frac{2}{\pi}} \frac{1 - \rho^2}{\sqrt{2 - \rho^2}} \int_{\mathbb{R}} \frac{h(x)}{\sqrt{1 + (1 - \rho^2) h(x)^2}} \varphi(x) x.
\]

\[
(15)
\]

**A very special case** As a simple check of the above result, consider the elementary case \(h(x) \equiv 1\), for which \((15)\) lends

\[
\sqrt{\frac{2}{\pi}} \frac{(1 - \rho^2)}{\sqrt{2 - \rho^2}}.
\]

In this case, \((14)\) is of type

\[
\text{SN}_2 \begin{pmatrix} \xi = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Omega = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}, \alpha = \begin{pmatrix} -\rho \\ 1 \end{pmatrix} \end{pmatrix}
\]

and known formulae such as (5.31) and (5.11) of the Azzalini & Capitanio (2014) say that

\[
E\left\{ \begin{pmatrix} X \\ Y \end{pmatrix} \right\} = \sqrt{\frac{2}{\pi}} \left( \frac{0}{\sqrt{2 - \rho^2}} \right).
\]

Addition numerical checks with other choices of \(h\) confirm expression \((15)\) by use of numerical integration.

**Other feasible solutions** The explicit expression of the integral in \((15)\) is feasible only in favourable cases.

A relatively common situation occurs when \(h(x)\) is odd, so that the integrand of \((15)\) is odd and

\[
E(Y) = 0.
\]
Note that, when $h$ is odd, $w(x, y) = (y - \rho x) h(x)$ is an even function of $(x, y)$. The special case $h(x) = \alpha x$ reduces to the distribution in the first expression of equation (17) of Azzalini (2012).

Another class of functions for which explicit integration is feasible is as follows. Suppose $s(x)$ is a function for which $\mathbb{E}\{s(Z)\}$ is known to be $S$, say, when $Z \sim \mathcal{N}(0, 1)$. Then solve

$$
\frac{h(x)}{\sqrt{1 + (1 - \rho^2) h(x)^2}} = s(x)
$$

and obtain

$$
h(x) = \frac{s(x)}{\sqrt{1 - (1 - \rho^2) s(x)^2}}.
$$

For this choice of $h(x)$, we can say that, by construction,

$$
\mathbb{E}\{Y\} = \sqrt{\frac{2}{\pi}} (1 - \rho^2) S.
$$

An especially simple example is $s(x) = \cos x$, such that $S = e^{-1/2}$, leading to

$$
h(x) = \frac{\cos x}{\sqrt{1 - (1 - \rho^2) \cos x^2}}, \quad \mathbb{E}\{Y\} = \sqrt{\frac{2}{\pi}} e (1 - \rho^2).
$$

Another special case leading to a solution, with help from the Maxima symbolic manipulation system, is $h(x) = \alpha |x|$ for which we obtain

$$
\mathbb{E}\{Y\} = \frac{2}{\sqrt{1 - \rho^2}} \left[ 1 - \Phi \left( \frac{1}{\alpha \sqrt{1 - \rho^2}} \right) \right] \exp \left( \frac{1}{2 \alpha^2 (1 - \rho^2)} \right) .
$$

Although in many other cases [15] does not lend itself to an explicit expression, it still simplifies computation with respect to direct two-dimensional numerical integration of $y f(x, y)$, since [15] requires only one-dimensional integration.

### 6 Extension to multivariate $Y$

Consider the more general case where $Y$ is $m$-dimensional, with $1 \leq m < d$. Hence write the joint distribution of $(X, Y)$ as

$$
\left( \begin{array}{c} X \\ Y \end{array} \right) \sim \mathcal{E}C_d \left( \mu, \Sigma, \psi \right), \quad \mu = \left( \begin{array}{c} \mu_x \\ \mu_y \end{array} \right), \quad \Sigma = \left( \begin{array}{cc} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{array} \right)
$$

where $X$ is $(d - m)$-dimensional. We want to extend the expressions in [5] and [7] to this new case.

**Proposition 4** For a $d$-dimensional variable $(X, Y)$ distributed as in (16), consider the transform

$$
W = w_0 \left( Y - m_Y(X) \right) h(X)
$$

where $h$ is an arbitrary function function and $w_0$ is an odd function, that is, $w_0(u) \in \mathbb{R}$ and $w_0(-u) = -w_0(u)$ for $u \in \mathbb{R}^m$. Then $W$ is distributed symmetrically around 0.
Proof. Write \( U = Y - m_Y(X) \) and \( W_0 = w_0(U) \). First we show that the distribution of \( W_0 \) conditionally on \( X = x \) is symmetric about 0, for any given \( x \); in fact
\[
\mathbb{P}(W_0 \leq t | X = x) = \mathbb{P}(-W_0 \geq -t | X = x) = \mathbb{P}(w_0(-U) \geq -t | X = x) = \mathbb{P}(w_0(U) \geq -t | X = x) = \mathbb{P}(W_0 \geq -t | X = x)
\]
for any real \( t \), and the last equality implies symmetry of \( (W_0 | X = x) \). Then argue like in Proposition 2 to conclude that \( W \) is symmetric about 0. \( \) QED

Corollary 5 Under the assumption of Proposition 3, denote by \( f_0 \) the density of (16) and let \( \beta_0 + \beta^\top x = m_Y(x) \) as given in equation (2.42) of Fang et alii (2012). Then expression (11) with
\[
w(x, y) = w_0(y - \beta_0 - \beta^\top x) h(x)
\]
is a proper density function.

References

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

Azzalini, A. (2012). Selection models under generalized symmetry settings. Annals of the Institute of Statistical Mathematics, 64, 737–750.

Azzalini, A., and Capitanio, A. (1999). Statistical applications of the multivariate skew normal distribution. J. R. Statist. Soc., ser. B, 61, 579–602. Full version of the paper at arXiv.org:0911.2093.

Azzalini, A., and Capitanio, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t distribution. J. R. Statist. Soc., ser. B, 65, 367–389. Full version of the paper at arXiv.org:0911.2342.

A. Azzalini, with the collaboration of A. Capitanio (2014). The Skew-Normal and Related Families. Cambridge University Press, Cambridge.

Azzalini, A., and Regoli, G. (2012). Some properties of skew-symmetric distributions. Ann. Inst. Statist. Math., 64, 857–879. Available online 09 Sept 2011.

Ellison, B. E. (1964). Two theorems for inferences about the normal distribution with applications in acceptance sampling. J. Amer. Statist. Assoc., 59, 89–95.

Fang, K.-T., Kotz, S., and Ng, K. W. (1990). Symmetric multivariate and related distributions. London: Chapman & Hall.
Jupp, P.E., Regoli, G. and Azzalini, A. (2016) A general setting for symmetric distributions and their relationship to general distributions. *Journal of Multivariate Analysis* 148, 107–119.

Sahu, K., Dey, D. K., and Branco, M. D. 2003. A new class of multivariate skew distributions with applications to Bayesian regression models. *Canad. J. Statist.*, 31, 129–150. Corrigendum: vol. 37 (2009), 301–302.

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