A survey on continuous elliptical vector distributions *

Eusebio Gómez [(1)], Miguel A. Gómez-Villegas [(1)]

and J. Miguel Marín [(2)]

(1) Departamento de Estadística e Investigación Operativa I
Universidad Complutense de Madrid
28040 Madrid, Spain
eusebio_gomez@mat.ucm.es
ma.gv@mat.ucm.es
(2) ESCET, Universidad Rey Juan Carlos
28933 Móstoles (Madrid), Spain
j.m.marin@escet.urjc.es

Recibido: 14 de Marzo de 2001
Aceptado: 8 de Mayo de 2002

ABSTRACT

In this paper it is taken up a revision and characterization of the class of absolutely continuous elliptical distributions upon a parameterization based on the density function. Their properties (probabilistic characteristics, affine transformations, marginal and conditional distributions and regression) are shown in a practical and easy to interpret way. Two examples are fully undertaken: the multivariate double exponential distribution and the multivariate uniform distribution.

2000 Mathematics Subject Classification: 60E05, 62H05.
Key words: Elliptical distribution; ellipticaly contoured distribution; stochastic representation;
conditional and marginal distributions; regression.

*Supported by DGICYT, grant PB98-0797.

1. Introduction

The purpose of this paper is to serve as a practical and quick survey on the class of continuous elliptical vector distributions, including specialized results for the continuous case which are meant to be precise, clear, easy to interpret and closer to applications.

Elliptical continuous distributions are those whose density functions are constant over ellipsoids as it is the case of normal distributions. This fact involves that many of the properties of normal distributions are maintained; for example, the regression functions are linear as in the normal case.

On the other hand, the whole class of continuous elliptical distributions is a very rich family, which can be used in many statistical procedures instead of the gaussian distributions, which are usually assumed.

General elliptical distributions, including the non continuous ones, have been fully studied: see [7], [2], [1], [4], [3] and [6]; actually, many of the results shown here can be viewed as specializations of some ones from these references. Nevertheless, these studies has been made in a general way upon a parameterization based on the characteristic function, and may not enlighten a first introduction to continuous elliptical distributions for applied statisticians. We make an approach focusing on the continuous case and using a parametrization closer to the density function. A wider development and revision of this subject may be found in [9].

In Section 2 the family of absolutely continuous elliptical distributions is defined and its stochastic representation is shown. In Section 3 the probabilistic characteristics and affine transformations of an elliptical vector are studied. In Section 4 it is studied the marginal and conditional distributions and Section 5 studies the regression. Finally, in section 6 two particular cases are developed, as examples of application: the multivariate double exponential distribution and the multivariate uniform distribution.

2. Definition and stochastic representation of the family of absolutely continuous elliptical distributions

In this Section the definition of the family of absolutely continuous elliptical distributions is shown, its parameters are considered and some important subfamilies are shown. As an alternative characterization of elliptical distributions it is shown the stochastic representation.

The concept of elliptical distribution was firstly exposed in [7] and developed in the rest of the referenced bibliography, mainly from the point of view of characteristic functions. We show the definition based upon the density function.

The stochastic representation was proposed in [1] as an alternative description of elliptical distributions.

Definition 1. *An absolutely continuous random vector $X = (X_1, \ldots, X_n)'$ has a n-*

dimensional elliptical distribution with parameters μ, Σ and g, where $\mu \in \mathbb{R}^n$, Σ is a positive definite $(n \times n)$ matrix and g is a non negative Lebesgue measurable function on $[0, \infty)$ such that

$$\int_0^\infty t^{\frac{n}{2}-1} g(t) dt < \infty, \tag{1}$$

if its density function is

$$f(x; \mu, \Sigma, g) = c_n \left|\Sigma\right|^{-\frac{1}{2}} g\left((x-\mu)' \Sigma^{-1} (x-\mu)\right), \tag{2}$$

where $c_n = \frac{\Gamma\left(\frac{n}{2}\right)}{\pi^{\frac{n}{2}} \int_0^\infty t^{\frac{n}{2}-1} g(t) dt}$. *The notation $X \sim E_n(\mu, \Sigma, g)$ will be used.*

As it follows immediately from next lemma, the function (2) is really a density function.

Lemma 2. *If $\mu \in \mathbb{R}^n$, Σ is a positive definite $(n \times n)$ matrix and g is a non negative Lebesgue measurable function on $[0, \infty)$, then*

$$\int_{\mathbb{R}^n} g\left((x-\mu)' \Sigma^{-1} (x-\mu)\right) dx_1 \ldots dx_n = \frac{\pi^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)} \left|\Sigma\right|^{\frac{1}{2}} \int_0^\infty t^{\frac{n}{2}-1} g(t) dt.$$

Proof. The result is obtained by changing $y = \Sigma^{-\frac{1}{2}}(x-\mu)$ and applying the lemmata 2.4.3 and 2.4.4 from [4], with $\alpha_i = \frac{1}{2}$, for $i = 1, \ldots, n$. ■

Although the parametric representation (μ, Σ, g) is essentially unique, it admits a minor variability in the parameters Σ and g. In fact, if $X \sim E_n(\mu, \Sigma, g)$, then $X \sim E_n(\mu^*, \Sigma^*, g^*)$ if and only if there exist two positive numbers a and b such that $\mu^* = \mu$, $\Sigma^* = a\Sigma$, $g^*(t) = bg(at)$, for almost all $t \in \mathbb{R}$. The proof of this statement can be found in [9].

The parameters μ and Σ are location and scale parameters. The *functional parameter* g is a non-normality parameter.

The elliptical family contains the normal family, which corresponds to the functional parameter $g(t) = \exp\left\{-\frac{1}{2}t\right\}$. It also contains a number of important subfamilies, which can be useful for robustness purposes. It has the scale mixtures of normal distributions; the reciprocal is partially true: it can be shown (see [2]) that any elliptical density f can be written in the form

$$f(x) = \int_0^\infty \frac{1}{(2\pi)^{\frac{n}{2}}} \left|t^{-1}\Sigma\right|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(x-\mu)' \left(t^{-1}\Sigma\right)^{-1} (x-\mu)\right\} dH(t),$$

but the function H does not need to be a non increasing one.

As an intermediate family between the normal and the elliptical ones there is the power exponential family (see [5] and [9]), which corresponds to $g(t) = \exp\left\{-\frac{1}{2}t^\beta\right\}$ for $\beta \in (0, \infty)$. The normal family is a subfamily of this one, corresponding to the value $\beta = 1$. Other outstanding subclasses of the power exponential family are the double exponential and the uniform ones, which correspond to $\beta = 1/2$ and to the limit value $\beta = \infty$.

The following theorem shows the *stochastic representation* of an absolutely continuous elliptical vector X. A general matrix variate version of this result, applicable to any elliptical matrix and based on the characteristic function, can be found in [6] (theorem 2.5.2).

Theorem 3. *A vector X verifies that $X \sim E_n(\mu, \Sigma, g)$ if and only if*

$$X \stackrel{d}{=} \mu + A' R U^{(n)}, \tag{3}$$

where A is any square matrix such that $A'A = \Sigma$, $U^{(n)}$ is a random vector uniformly distributed on the unit sphere of $\mathbb{R}^n$ and R is an absolutely continuous non negative random variable independent from $U^{(n)}$ whose density function is

$$h(r) = \frac{2}{\int_0^\infty t^{\frac{n}{2}-1} g(t) dt} r^{n-1} g(r^2) I_{(0,\infty)}(r). \tag{4}$$

Proof. To proof the "only if" part, by making the change of variable $Y = A'^{-1}(X - \mu)$, where A is a square matrix such that $A'A = \Sigma$, it is obtained that Y has a centered spherical distribution. Then, from theorem 2.5.3 from [4], it follows that $Y \stackrel{d}{=} RU^{(n)}$ and $X \stackrel{d}{=} \mu + A'RU^{(n)}$. The density function (4) for R is obtained by applying theorem 2.5.5 from [4].

The"if" part is obtained by making the change of variable $Y = A'^{-1}(X - \mu)$ and applying theorem 2.5.5 from [4]. ∎

Now, in addition to the parametric representation (μ, Σ, g), there is the stochastic representation (μ, A, h), where $A'A = \Sigma$ and the relationship between functions g and h is given by (4) and the equation $g(t) = t^{\frac{1-n}{2}} h\left(t^{\frac{1}{2}}\right)$. The density (2) can be rewritten in terms of this new parametrization as $f(x) = \frac{\Gamma\left(\frac{n}{2}\right)}{2\pi^{\frac{n}{2}}} |A|^{-1} (q(x))^{\frac{1-n}{2}} h\left((q(x))^{\frac{1}{2}}\right)$, where $q(x) = (x-\mu)'(A'A)^{-1}(x-\mu)$.

The module of $Y = \Sigma^{-\frac{1}{2}}(X - \mu)$ is distributed as R since, from (3), $|Y| \stackrel{d}{=} \left|RU^{(n)}\right| = R$; so, R will be named the *modular variable* and h the *modular density.* The direction $\frac{Y}{|Y|}$ of Y is uniformly distributed on the unit sphere, as $U^{(n)}$. The module and the direction are independent.

The moments $E[R^s]$ of the modular variable do exist if and only if

$$\int_0^\infty t^{\frac{n+s}{2}-1} g(t) dt < \infty;$$

in this case

$$E[R^s] = \frac{\int_0^\infty t^{\frac{n+s}{2}-1} g(t)dt}{\int_0^\infty t^{\frac{n}{2}-1} g(t)dt}. \tag{5}$$

$Q = (X-\mu)'\Sigma^{-1}(X-\mu)$, the quadratic form taken from (2) and applied to vector X, has the same distribution that R^2, since $Q = Y'Y$.

The representation (3) suggests that an algorithm to generate an observation of a vector $X \sim E_n(\mu, \Sigma, g)$ would consist of: *(1)* finding a square matrix A such that $A'A = \Sigma$; *(2)* generating an observation of the modular variable R according to its own distribution; *(3)* generating an observation of the uniform vector $U^{(n)}$; *(4)* obtaining X according to (3).

3. Probabilistic characteristics and affine transformations

We study probabilistic characteristics of elliptical distributions as well as their affine transformations.

Some aspects of the mean vector and covariance matrix may be found in [4] and [6]. The other characteristics, such as the kurtosis coefficient, are easily derived.

The basis references for affine transformations are also [4] and [6].

Next theorem 4 shows the main moments and the skewness and kurtosis coefficients (in the general sense of [8]) of elliptical vectors.

Theorem 4. *Let $X \sim E_n(\mu, \Sigma, g)$ and consider the stochastic representation $X \stackrel{d}{=} \mu + A'RU^{(n)}$ given by theorem 3.*
(i) The mean vector $E[X]$ exists if and only if $E[R]$ exists. In this case, $E[X] = \mu$.
(ii) The covariance matrix $Var[X]$ exists if and only if $E[R^2]$ exists. In this case, $Var[X] = \frac{1}{n}E\left[R^2\right]\Sigma$.
(iii) If $E[R^3]$ exists, then the skewness coefficient of X exists. In this case, $\gamma_1[X] = 0$.
(iv) If $E[R^4]$ exists, then the kurtosis coefficient of X exists. In this case,

$$\gamma_2[X] = n^2 \frac{E[R^4]}{\left(E\left[R^2\right]\right)^2}. \tag{6}$$

Proof. (i) and (ii) follow from theorem 2.6.4 from [4].
(iii) The vectors X and Y that appear in the definition

$$\gamma_1[X] = E[((X - E[X])' \left(Var[X]\right)^{-1} \left(Y - E[Y]\right))^3]$$

have the same distribution as $\mu + A'RU^{(n)}$ and $\mu + A'SV^{(n)}$, respectively, in the sense of theorem 3, where R, S, $U^{(n)}$ and $V^{(n)}$ are independent variables. So, denoting

$a = \frac{1}{n} E\left[R^2\right]$,

$$\begin{aligned}
\gamma_1[X] &= E\left[\left((X-\mu)'(Var[X])^{-1}(Y-\mu)\right)^3\right] \\
&= E\left[\left(\left(A'RU^{(n)}\right)'(a\Sigma)^{-1}\left(A'SV^{(n)}\right)\right)^3\right] \\
&= a^{-3} E\left[\left(RU^{(n)\prime}A\Sigma^{-1}A'V^{(n)}S\right)^3\right] \\
&= a^{-3} E\left[(RS)^3\left(U^{(n)\prime}V^{(n)}\right)^3\right] \\
&= a^{-3} E\left[(RS)^3\right] E\left[\left(U^{(n)\prime}V^{(n)}\right)^3\right] = 0,
\end{aligned}$$

since $E\left[\left(U^{(n)\prime}V^{(n)}\right)^3\right] = 0$. To prove this last statement, notice that for every $u_0^{(n)}$

$$E\left[\left(U^{(n)\prime}V^{(n)}\right)^3 \middle| U^{(n)} = u_0^{(n)}\right] = E\left[\left(u_0^{(n)\prime}V^{(n)}\right)^3\right] = 0, \tag{7}$$

since $u_0^{(n)\prime}V^{(n)}$ is symmetrical around the origin. Since all the conditional expectations (7) are null, so is the absolute expectation.
(iv) Since $(X-\mu)'\Sigma^{-1}(X-\mu) \stackrel{d}{=} R^2$, then

$$\begin{aligned}
\gamma_2[X] &= E\left[\left((X-\mu)'(Var[X])^{-1}(X-\mu)\right)^2\right] = (Var[X])^{-1} E[R^4] \\
&= n^2 \frac{E[R^4]}{\left(E\left[R^2\right]\right)^2}.
\end{aligned}$$

■

It can be observed that the mean vector is the location parameter μ, the covariance matrix is proportional to the scale parameter Σ and the kurtosis coefficient depends only on the non-normality parameter g. The factor $\frac{E[R^4]}{(E[R^2])^2}$, that appears in (6), is equivalent to the kurtosis coefficient of a symmetrical random variable Z with mean 0 such that $|Z| \stackrel{d}{=} R$, whose density function, when $\Sigma = I_n$, is $f_Z(z) = \frac{1}{2}h(|z|)$ where h is defined in (4).

Now, affine transformations of the form $Y = CX + b$ of a vector $X \sim E_n(\mu_X, \Sigma_X, g_X)$ are considered. If C is a nonsingular $(n \times n)$ matrix it can be easily verified that $Y \sim E_n(C\mu_X + b, C\Sigma_X C', g_X)$. The following theorem studies the case in which C is a $(p \times n)$ matrix with $p < n$ and $rank(C) = p$.

Theorem 5. *Let $X \sim E_n(\mu_X, \Sigma_X, g_X)$, with stochastic representation $X \stackrel{d}{=} \mu_X + A'_X R_X U^{(n)}$. Let $Y = CX + b$, where C is a $(p \times n)$ matrix such that $p < n$ and $rank(C) = p$ and $b \in \mathbb{R}^p$. Then* *(i) $Y \sim E_p(\mu_Y, \Sigma_Y, g_Y)$ with $\mu_Y = C\mu_X + b$, $\Sigma_Y = C\Sigma_X C'$, $g_Y(t) = \int_0^\infty w^{\frac{n-p}{2}-1} g_X(t+w)dw$.*

(ii) Vector Y has the stochastic representation $Y \stackrel{d}{=} \mu_Y + A'_Y R_Y U^{(p)}$, where $R_Y \stackrel{d}{=} R_X\sqrt{B}$ and $\sqrt{B}$ is the non negative square root of a random variable B independent of R_X with distribution Beta $\left(\frac{p}{2}, \frac{n-p}{2}\right)$.

Proof. By substitution it is obtained that $Y \stackrel{d}{=} \mu_Y + CA'_X R_X U^{(n)}$. Then, by corollary 1 of theorem 2.6.1 from [4], Y is elliptically distributed and has the following stochastic representation

$$Y \stackrel{d}{=} \mu + A'_Y R_Y U^{(p)}. \tag{8}$$

Thus, there are two distribution equalities for Y, and from theorem 2.6.2(ii) from [4], making $c = 1$, it is obtained that $\mu = \mu_Y,$ $A'_Y A_Y = CA'_X A_X C' = \Sigma_Y$ and $R_Y \stackrel{d}{=} R_X\sqrt{B}$, where B is the random variable in the statement; the density of R_Y is

$$h_Y(r) = \frac{2r^{p-1}}{B\left(\frac{p}{2}, \frac{n-p}{2}\right)\int_0^\infty t^{\frac{n}{2}-1} g_X(t)dt} \left(\int_0^\infty w^{\frac{n-p}{2}-1} g_X(r^2+w)dw\right) I_{(0,\infty)}(r).$$

This proves part (ii). Now, since Y has the stochastic representation (8), it follows that Y has the elliptical distribution obtained in part (i). ■

4. Marginal and conditional distributions

In this section we show the features of marginal and conditional distributions in an explicit, intuitive and easy-to-implement way. A more abstract study of this subject may be found in [1].

Theorems 6 and 7 show the distribution and probabilistic characteristics of a subvector $X_{(1)}$ of a vector $X \sim E_n(\mu, \Sigma, g)$.

Theorem 6. *Let $X \sim E_n(\mu, \Sigma, g)$, with $n \geq 2$ and stochastic representation $X \stackrel{d}{=} \mu + A'RU^{(n)}$. Let us consider the partition $X = (X'_{(1)}, X'_{(2)})'$, where $X_{(1)} = (X_1, \ldots, X_p)'$ and $X_{(2)} = (X_{p+1}, \ldots, X_n)'$, $p < n$; as well as the respective partitions of μ and Σ: $\mu = (\mu'_{(1)}, \mu'_{(2)})'$, with $\mu_{(1)} = (\mu_1, \ldots, \mu_p)'$, $\mu_{(2)} = (\mu_{p+1}, \ldots, \mu_n)'$; and $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$, where Σ_{11} is the upper left $(p \times p)$ submatrix of Σ.*

(i) The marginal distribution of $X_{(1)}$ is $X_{(1)} \sim E_p(\mu_{(1)}, \Sigma_{11}, g_{(1)})$, with

$$g_{(1)}(t) = \int_0^\infty w^{\frac{n-p}{2}-1} g(t+w)dw.$$

(ii) Vector $X_{(1)}$ has a stochastic representation of the form $X_{(1)} \stackrel{d}{=} \mu_{(1)} + A'_{(1)} R_{(1)} U^{(p)}$, where $R_{(1)} \stackrel{d}{=} R\sqrt{B}$ and $\sqrt{B}$ are as in theorem 5 (ii).

Proof. It is clear that $X_{(1)} = CX$, where $C = \begin{pmatrix} I_p & 0_{p\times(n-p)} \end{pmatrix}$ is a $(p \times n)$ matrix, where the box I_p is the $(p \times p)$ identity matrix and $0_{p\times(n-p)}$ is the $(p \times (n-p))$ null matrix. Thus, $X_{(1)}$ is an affine transformation of X, and the theorem is immediately obtained by applying theorem 5. ■

Theorem 7. *With the same hypotheses and notations of theorem 6,*
(i) Vector $E\left[X_{(1)}\right]$ exists if and only if $E[X]$ exists, or equivalently if $E[R]$ exists. In this case $E\left[X_{(1)}\right] = \mu_{(1)}$.
(ii) Covariance matrix $Var\left[X_{(1)}\right]$ exists if and only if $Var[X]$ exists, or equivalently if $E\left[R^2\right]$ exists. In this case $Var\left[X_{(1)}\right] = \frac{1}{n}E\left[R^2\right]\Sigma_{11} = Var[X]\Sigma^{-1}\Sigma_{11}$.
(iii) If $E[R^3]$ exists, then the skewness coefficient of $X_{(1)}$ exists and it is $\gamma_1\left[X_{(1)}\right] = 0$.
(iv) If $E[R^4]$ exists, then the kurtosis coefficient of $X_{(1)}$ exists and it is

$$\gamma_2\left[X_{(1)}\right] = \frac{np(p+2)}{n+2}\frac{E[R^4]}{\left(E\left[R^2\right]\right)^2} = \frac{p(p+2)}{n(n+2)}\gamma_2[X].$$

Proof. The theorem is obtained by applying theorem 4 to vector $X_{(1)}$. Here, the moment $E\left[R^s_{(1)}\right]$ exists if and only if $E\left[R^s\right]$ exists, since $R_{(1)} \stackrel{d}{=} R\sqrt{B}$; in this case $E\left[R^s_{(1)}\right] = E\left[R^s\right]\frac{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{p+s}{2}\right)}{\Gamma\left(\frac{p}{2}\right)\Gamma\left(\frac{n+s}{2}\right)}$. ■

Theorem 8 and 10 study the conditional distribution and properties of the subvector $X_{(2)}$.

Theorem 8. *With the same hypotheses as in theorem 6, for each $x_{(1)} \in \mathbb{R}^p$ such that $0 < f_{(1)}\left(x_{(1)}\right) < \infty$, the distribution of $X_{(2)}$ conditional to $X_{(1)} = x_{(1)}$ is*

$$\left(X_{(2)} \left| X_{(1)} = x_{(1)}\right.\right) \sim E_{n-p}\left(\mu_{(2.1)}, \Sigma_{22.1}, g_{(2.1)}\right), \tag{9}$$

with $\mu_{(2.1)} = \mu_{(2)} + \Sigma_{21}\Sigma_{11}^{-1}\left(x_{(1)} - \mu_{(1)}\right)$, $\Sigma_{22.1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$ and $g_{(2.1)}(t) = g\left(t + q_{(1)}\right)$, where $q_{(1)} = \left(x_{(1)} - \mu_{(1)}\right)'\Sigma_{11}^{-1}\left(x_{(1)} - \mu_{(1)}\right)$.

Proof. The conditional density function is proportional to the joint density function of $X = (X'_{(1)}, X'_{(2)})'$; then

$$\begin{aligned} f_{(2.1)}(x_{(2)} \left| X_{(1)} = x_{(1)}\right.) &\propto g\left((x-\mu)'\Sigma^{-1}(x-\mu)\right) \\ &= g\left(q_{(1)} + \left(x_{(2)} - \mu_{(2.1)}\right)'\Sigma_{22.1}^{-1}\left(x_{(2)} - \mu_{(2.1)}\right)\right), \end{aligned} \tag{10}$$

since for each $x \in \mathbb{R}^n$

$$\begin{aligned} &x'\Sigma^{-1}x = x'_{(1)}\Sigma_{11}^{-1}x_{(1)} + \\ &+ \left(x_{(2)} - \Sigma_{21}\Sigma_{11}^{-1}x_{(1)}\right)'\left(\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\right)^{-1}\left(x_{(2)} - \Sigma_{21}\Sigma_{11}^{-1}x_{(1)}\right). \end{aligned}$$

The density (10) corresponds to distribution (9). ■

Next lemma will be used to proof theorem 10.

Lemma 9. *If n is an integer greater than 1 and g is a non negative Lebesgue measurable function on $[0,\infty)$ such that $\int_0^\infty t^{\frac{n}{2}-1}g(t)dt < \infty$, then $\int_0^\infty t^{\frac{r}{2}-1}g(t+q)dt < \infty$, for each $r = 1, ..., n-1$ and almost every $q \geq 0$.*

Proof. Since function g verifies condition (1), let $(V_1,..., V_n)'$ be a random vector such that $V \sim E_n(0, I_n, g)$. For each $r = 1, ..., n-1$, the marginal density of $V_{(1)} = (V_1, ..., V_{n-r})'$ in any point $v_{(1)} = (v_1, ..., v_{n-r})'$ is proportional to $\int_0^\infty w^{\frac{r}{2}-1} g\left(q\left(v_{(1)}\right) + w\right) dw$, where $q\left(v_{(1)}\right) = \left|v_{(1)}\right|^2$. Therefore, the previous integral is finite for almost each $v_{(1)} \in \mathbb{R}^{n-r}$. Thus, $\int_0^\infty w^{\frac{r}{2}-1} g\left(q+w\right) dw < \infty$ for almost every $q \geq 0$. ■

Theorem 10. *With the same hypotheses and notations as in theorem 8 we have*
(i) If $p \geq 2$, or else if $p = 1$ and $E[R]$ exists, then the mean vector of $X_{(2)}$ conditional to $X_{(1)} = x_{(1)}$ exists and it is $E\left[X_{(2)} \left| X_{(1)} = x_{(1)}\right.\right] = \mu_{(2.1)}$.
(ii) If $p \geq 3$, or else if $p \in \{1, 2\}$ and $E[R^{3-p}]$ exists, then the covariance matrix of $X_{(2)}$ conditional to $X_{(1)} = x_{(1)}$ exists and it follows next expression

$$Var\left[X_{(2)} \left| X_{(1)} = x_{(1)}\right.\right] = \frac{1}{n-p} \frac{\int_0^\infty t^{\frac{n-p}{2}} g(t+q_{(1)})dt}{\int_0^\infty t^{\frac{n-p}{2}-1} g(t+q_{(1)})dt} \Sigma_{22.1}.$$

(iii) If $p \geq 4$, or else if $p \in \{1, 2, 3\}$ and $E[R^{4-p}]$ exists, then the skewness coefficient of $X_{(2)}$ conditional to $X_{(1)} = x_{(1)}$ exists and it is

$$\gamma_1\left[X_{(2)} \left| X_{(1)} = x_{(1)}\right.\right] = 0.$$

(iv) If $p \geq 5$, or else if $p \in \{1, ..., 4\}$ and $E[R^{5-p}]$ exists, then the kurtosis coefficient of $X_{(2)}$ conditional to $X_{(1)} = x_{(1)}$ exists and it follows next expression

$$\gamma_2\left[X_{(2)} \left| X_{(1)} = x_{(1)}\right.\right] = (n-p)^2 \frac{\left(\int_0^\infty t^{\frac{n-p}{2}+1} g(t+q_{(1)})dt\right)\left(\int_0^\infty t^{\frac{n-p}{2}-1} g(t+q_{(1)})dt\right)}{\left(\int_0^\infty t^{\frac{n-p}{2}} g(t+q_{(1)})dt\right)^2}.$$

Proof. The theorem is obtained by applying theorem 4 to the conditional distribution (9) of $X_{(2)}$. Here, under the conditions of each one of these parts, the moments $E\left[R_{(2.1)}^s\right]$ of the modular variable do exist (for $s = 1, 2, 3, 4$), as can easily be derived from lemma 9. These moments are $E[R_{(2.1)}^s] = \frac{\int_0^\infty t^{\frac{n-p+s}{2}-1} g(t+q_{(1)})dt}{\int_0^\infty t^{\frac{n-p}{2}-1} g(t+q_{(1)})dt}$. ■

Since $E\left[X_{(2)} \left| X_{(1)} = x_{(1)}\right.\right] = \mu_{(2)} + \Sigma_{21}\Sigma_{11}^{-1}\left(x_{(1)} - \mu_{(1)}\right)$, the regression function of $X_{(2)}$ over $X_{(1)}$ is linear, as in the normal case.

5. Regression

As a natural development of the former concepts, we undertake in this section the study of regression. We show that the elliptical distributions behavior as the normal distribution in that their general function of regression is affine. We consider the characteristics of forecasts and errors as well as the residual variance and the correlation ratio.

With the same notations as in section 4, we define:

- *forecast function* of $X_{(2)}$ from $X_{(1)}$ is any function φ defined over the set $\left\{x_{(1)} \in \mathbb{R}^p \;\middle|\; 0 < f_{(1)}\left(x_{(1)}\right) < \infty\right\}$ taking values in $\mathbb{R}^{n-p}$;
- for each forecast function φ, the *functional of error* is
$$\triangle(\varphi) = E\left[\left|X_{(2)} - \varphi\left(X_{(1)}\right)\right|^2\right];$$
- the *general regression function* is the best forecast function, which minimizes the functional $\triangle$;
- the *residual variance* is the value of the functional $\triangle$ corresponding to the general regression function.

As a reference, we consider the best *constant* forecast function. It can be shown that this function is, in any case,

$$\varphi\left(x_{(1)}\right) = E\left[X_{(2)}\right],$$

and its functional of error $\triangle$ is

$$\triangle(\varphi) = \sum_{i=p+1}^{n} Var\left[X_i\right] = tr(Var\left[X_{(2)}\right]),$$

where *tr* means the trace.

It can also be shown that the general regression function is, in any case,

$$\varphi\left(x_{(1)}\right) = E\left[X_{(2)} \left| X_{(1)} = x_{(1)}\right.\right], \tag{11}$$

and the residual variance is

$$Res.Var. = \sum_{i=p+1}^{n} E\left[Var\left[X_i \left| X_{(1)}\right.\right]\right]. \tag{12}$$

We define the *correlation ratio* as

$$\eta^2_{(2.1)} = \frac{\sum_{i=p+1}^{n} Var\left[X_i\right] - Res.Var.}{\sum_{i=p+1}^{n} Var\left[X_i\right]} = \frac{\sum_{i=p+1}^{n} Var\left[E\left[X_i \left| X_{(1)}\right.\right]\right]}{\sum_{i=p+1}^{n} Var\left[X_i\right]}. \tag{13}$$

In the next theorem the former concepts are applied to elliptical distributions.

Theorem 11. *With the same hypotheses and notations of Theorem 8, if*

$$X \stackrel{d}{=} \mu + A' R U^{(n)}$$

is the stochastic representation of X in the sense of Theorem 3 (i), and supposing that moments exist:

(i) The general regression function of $X_{(2)}$ over $X_{(1)}$ is the affine function φ defined as

$$\begin{aligned} \varphi\left(x_{(1)}\right) &= \mu_{(2.1)} \\ &= \mu_{(2)} + \Sigma_{21}\Sigma_{11}^{-1}\left(x_{(1)} - \mu_{(1)}\right), \end{aligned}$$

for all $x_{(1)}$ which $0 < f_{(1)}\left(x_{(1)}\right) < \infty$.

(ii) For the random vector of forecasts,

$$\begin{aligned} \widehat{X}_{(2)} &= \varphi\left(X_{(1)}\right) \\ &= \mu_{(2)} + \Sigma_{21}\Sigma_{11}^{-1}\left(X_{(1)} - \mu_{(1)}\right), \end{aligned}$$

we have

$$\begin{aligned} E\left[\widehat{X}_{(2)}\right] &= E\left[X_{(2)}\right] = \mu_{(2)}, \\ Var\left[\widehat{X}_{(2)}\right] &= Var\left[X_{(2)}\right] - \frac{1}{n}E\left[R^2\right]\Sigma_{22.1} \\ &= \frac{1}{n}E\left[R^2\right]\left(\Sigma_{22} - \Sigma_{22.1}\right) \\ &= \frac{1}{n}E\left[R^2\right]\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}. \end{aligned}$$

(iii) If we denote as Z the random vector of errors,

$$Z = X_{(2)} - \widehat{X}_{(2)},$$

and H the joint vector

$$H = (X'_{(1)}, Z')',$$

then the distribution of H is elliptical:

$$H \sim E_n\left(\mu_H, \Sigma_H, g_H\right),$$

with

$$\begin{aligned}
\mu_H &= (\mu'_{(1)}, 0'_{(n-p)\times 1})', \\
\Sigma_H &= \begin{pmatrix} \Sigma_{11} & 0_{p\times(n-p)} \\ 0_{(n-p)\times p} & \Sigma_{22.1} \end{pmatrix}, \\
g_H &= g.
\end{aligned}$$

(iv) The (marginal) distribution of the vector of errors Z is $E_{n-p}(0_{(n-p)\times 1}, \Sigma_{22.1}, g_Z)$, with

$$g_Z(t) = \int_0^\infty w^{\frac{p}{2}-1} g(t+w)dw.$$

The characteristics of Z are

$$\begin{aligned}
E[Z] &= 0, \\
Var[Z] &= Var\left[X_{(2)}\right] - Var\left[\widehat{X}_{(2)}\right], \\
&= \frac{1}{n} E\left[R^2\right] \Sigma_{22.1}, \\
\gamma_1[Z] &= 0, \\
\gamma_2[Z] &= \gamma_2\left[X_{(2)}\right] \\
&= \frac{(n-p)(n-p+2)}{n(n+2)} \gamma_2[X].
\end{aligned}$$

(v) The residual variance is

$$\begin{aligned}
Res.Var. &= tr(Var[Z]) \\
&= \frac{1}{n} E\left[R^2\right] \cdot tr\left(\Sigma_{22.1}\right).
\end{aligned}$$

(vi) The correlation ratio is

$$\eta^2_{(2.1)} = \frac{tr(Var\left[\widehat{X}_{(2)}\right])}{tr(Var\left[X_{(2)}\right])} = \frac{tr(\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})}{tr(\Sigma_{22})}.$$

Proof. (i) By replacing in (11) the value of the conditional mean, given by Theorem 10, we have

$$\begin{aligned}\varphi\left(x_{(1)}\right) &= E\left[X_{(2)}\left|X_{(1)}=x_{(1)}\right.\right]\\ &= \mu_{(2.1)}\\ &= \mu_{(2)}+\Sigma_{21}\Sigma_{11}^{-1}\left(x_{(1)}-\mu_{(1)}\right).\end{aligned}$$

(ii) The mean of $\widehat{X}_{(2)}$ is

$$\begin{aligned}E\left[\widehat{X}_{(2)}\right] &= \mu_{(2)}+\Sigma_{21}\Sigma_{11}^{-1}E[X_{(1)}-\mu_{(1)}]\\ &= \mu_{(2)}\\ &= E\left[X_{(2)}\right];\end{aligned}$$

its covariance matrix is

$$\begin{aligned}Var\left[\widehat{X}_{(2)}\right] &= \Sigma_{21}\Sigma_{11}^{-1}\left(Var\left[X_{(1)}\right]\right)\Sigma_{11}^{-1}\Sigma_{12}\\ &= \Sigma_{21}\Sigma_{11}^{-1}\left(\frac{1}{n}E\left[R^2\right]\Sigma_{11}\right)\Sigma_{11}^{-1}\Sigma_{12}\\ &= \frac{1}{n}E\left[R^2\right]\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\\ &= \frac{1}{n}E\left[R^2\right]\left(\Sigma_{22}-\Sigma_{22.1}\right)\\ &= Var\left[X_{(2)}\right]-\frac{1}{n}E\left[R^2\right]\Sigma_{22.1}.\end{aligned}$$

(iii) We have that

$$\begin{aligned}H &= \begin{pmatrix} X_{(1)}\\ Z\end{pmatrix}\\ &= \begin{pmatrix} I_p & 0_{p\times(n-p)}\\ -\Sigma_{21}\Sigma_{11}^{-1} & I_{n-p}\end{pmatrix}\begin{pmatrix} X_{(1)}\\ X_{(2)}\end{pmatrix}+\begin{pmatrix} 0\\ -\mu_{(2)}+\Sigma_{21}\Sigma_{11}^{-1}\mu_{(1)}\end{pmatrix}.\end{aligned}$$

The rank of the matrix $\begin{pmatrix} I_p & 0_{p\times(n-p)}\\ -\Sigma_{21}\Sigma_{11}^{-1} & I_{n-p}\end{pmatrix}$ is n; now the result is obtained by applying Theorem 5.

(iv) The distribution of Z is obtained as the marginal of its joint distribution with $X_{(1)}$ by applying Theorem 6. Moments are obtained by applying Theorem 7 and considering that the variable R_H, from the stochastic representation of H, has the same distribution as the corresponding of X.

(v) For the residual variance:

$$\begin{aligned} Res.Var. &= \triangle(\varphi) \\ &= E\left[\left|X_{(2)} - \varphi\left(X_{(1)}\right)\right|^2\right] \\ &= E\left[|Z|^2\right] \\ &= E\left[\sum_{i=p+1}^{n} Z_i^2\right] \\ &= \sum_{i=p+1}^{n} Var\left[Z_i\right] \\ &= tr(Var[Z]) \\ &= \frac{1}{n} E\left[R^2\right] \cdot tr\left(\Sigma_{22.1}\right). \end{aligned}$$

(vi) For the correlation ratio:

$$\begin{aligned} \eta^2_{(2.1)} &= \frac{tr(Var\left[X_{(2)}\right]) - Res.Var.}{tr(Var\left[X_{(2)}\right])} \\ &= \frac{tr(Var\left[\widehat{X}_{(2)}\right])}{tr(Var\left[X_{(2)}\right])} \\ &= \frac{tr(\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})}{tr(\Sigma_{22})}. \end{aligned}$$

■

These equalities relate parallel concepts:

$$\begin{aligned} Var\left[X_{(2)}\right] &= Var\left[\widehat{X}_{(2)}\right] + Var[Z]; \\ \sum_{i=p+1}^{n} Var\left[X_i\right] &= \sum_{i=p+1}^{n} Var\left[E\left[X_i \left| X_{(1)}\right.\right]\right] + \sum_{i=p+1}^{n} E\left[Var\left[X_i \left| X_{(1)}\right.\right]\right]; \\ \Sigma_{22} &= \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} + \Sigma_{22.1} \end{aligned}$$

We can observe that the correlation ratio $\eta^2_{(2.1)}$ depends only on the scale parameter Σ.

6. Examples

The most important and best studied example of absolutely continuous elliptical distributions, of course, the normal distribution; we only point out that its main properties could be derived from the previous Sections by making $g(t) = \exp\left\{-\frac{1}{2}t\right\}$. Here we show two other examples: the multivariate double exponential distribution and the multivariate uniform distribution.

The multivariate double exponential distribution is the elliptical distribution $E_n(\mu, \Sigma, g)$ corresponding to the functional parameter $g(t) = \exp\left\{-\frac{1}{2}t^{\frac{1}{2}}\right\}$, which satisfies condition (1) for each $n \geq 1$. Its density function is

$$f\left(x; \mu, \Sigma, \frac{1}{2}\right) = \frac{n\Gamma\left(\frac{n}{2}\right)}{\pi^{\frac{n}{2}}\Gamma(1+n)2^{1+n}} |\Sigma|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left((x-\mu)'\Sigma^{-1}(x-\mu)\right)^{\frac{1}{2}}\right\}. \tag{14}$$

As an example, figure 1 shows the graph of the density function (14), for $n = 2$, $\mu = (6,4)'$ and $\Sigma = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$.

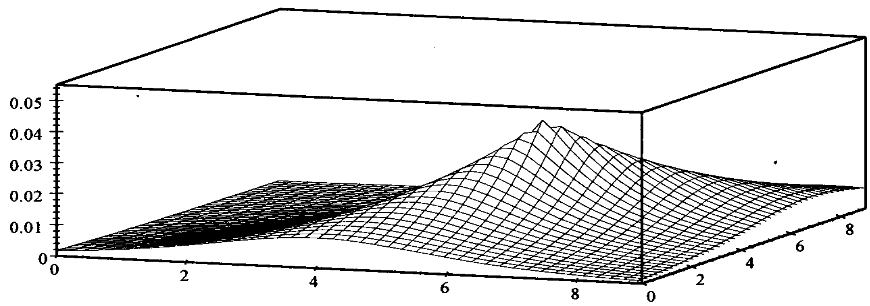

Figure 1. A multivariate double exponential density.

The multivariate uniform distribution corresponds to $g(t) = I_{[0,1]}(t)$, which also satisfies condition (1) for each $n \geq 1$. Its density function is

$$f(x; \mu, \Sigma) = \frac{n\Gamma\left(\frac{n}{2}\right)}{2\pi^{\frac{n}{2}}} |\Sigma|^{-\frac{1}{2}} I_S(x), \tag{15}$$

where I_S is the indicator function of ellipsoid S with equation $(x-\mu)'\Sigma^{-1}(x-\mu) \leq 1$. Figure 2 shows the graph of density function (15), for $n = 2$, $\mu = (6,4)'$ and $\Sigma = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$.

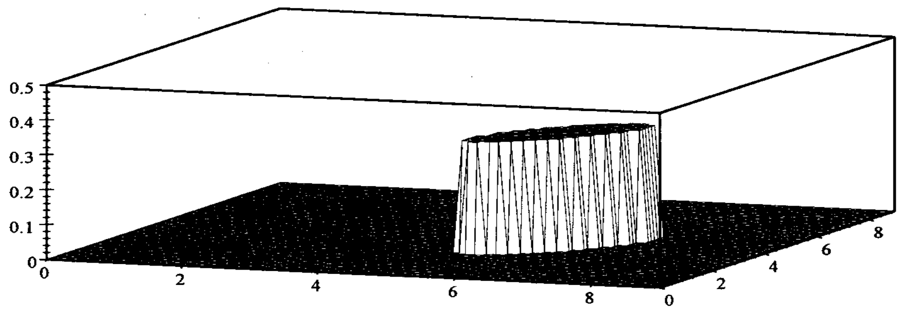

Figure 2. A multivariate uniform density.

The modular density h, the moments of the modular variable R and the probabilistic characteristics of a vector X, for these two distributions, are shown in table 1:

	Double Exponential	Uniform
$h(r)$	$\frac{1}{2^n (n-1)!} r^{n-1} \exp\left\{-\frac{1}{2}r\right\} I_{[0,1]}(r)$	$nr^{n-1} I_{[0,1]}(r)$
$E[R^s]$	$2^s \frac{\Gamma(n+s)}{\Gamma(n)}$	$\frac{n}{s+n}$
$E[X]$	μ	μ
$Var[X]$	$4(n+1)\Sigma$	$\frac{1}{n+2}\Sigma$
$\gamma_1[X]$	0	0
$\gamma_2[X]$	$\frac{n(n+2)(n+3)}{n+1}$	$\frac{n(n+2)^2}{n+4}$

Table 1. Probabilistic characteristics.

The functional parameter $g_{(1)}$ and the (marginal) probabilistic characteristics of a subvector $X_{(1)}$ are shown in table 2:

	Double Exponential	Uniform
$g_{(1)}(t)$	$\int_0^\infty w^{\frac{n-p}{2}-1} \exp\left\{-\frac{1}{2}(t+w)^{\frac{1}{2}}\right\} dw$	$(1-t)^{\frac{n-p}{2}} I_{[0,1]}(t)$
$E[X_{(1)}]$	$\mu_{(1)}$	$\mu_{(1)}$
$Var[X_{(1)}]$	$4(n+1)\Sigma_{11}$	$\frac{1}{n+2}\Sigma_{11}$
$\gamma_1[X_{(1)}]$	0	0
$\gamma_2[X_{(1)}]$	$\frac{p(p+2)(n+3)}{n+1}$	$\frac{p(p+2)(n+2)}{n+4}$

Table 2. Marginal characteristics.

Distribution of $X_{(2)}$ conditional to $X_{(1)} = x_{(1)}$ is defined for all $x_{(1)} \in \mathbb{R}^p$ in the case of the double exponential distribution, and for all $x_{(1)}$ such that $q_{(1)} < 1$ in the case of the uniform distribution. The functional parameter $g_{(2.1)}$ and the conditional covariance matrix are shown in table 3:

	Double Exponential	Uniform
$g_{(2.1)}(t)$	$\exp\left\{-\frac{1}{2}\left(t+q_{(1)}\right)^{\frac{1}{2}}\right\}$	$I_{\left[0,1-q_{(1)}\right]}(t)$
$Var\left[X_{(2)}\left\|X_{(1)}=x_{(1)}\right.\right]$	$\frac{1}{n-p}\frac{\int_0^\infty t^{\frac{n-p}{2}}\exp\left\{-\frac{1}{2}\left(t+q_{(1)}\right)^{\frac{1}{2}}\right\}dt}{\int_0^\infty t^{\frac{n-p}{2}-1}\exp\left\{-\frac{1}{2}\left(t+q_{(1)}\right)^{\frac{1}{2}}\right\}dt}\Sigma_{22.1}$	$\frac{1-q_{(1)}}{n-p+2}\Sigma_{22.1}$

Table 3. Conditional characteristics.

References

[1] S. Cambanis, S. Huang and G. Simons (1981). On the Theory of Elliptically Contoured Distributions. *J. of Multivariate Analysis*, **11**, 365–385.

[2] K. Chu, (1973). Estimation and Decision for Linear Systems with Elliptical Random Processes. *IEEE Trans. on Aut. Control*, **18**, 499–505.

[3] K. Fang, S. Kotz, and K. Ng (1990). *Symmetric Multivariate and Related Distributions.* Chapman and Hall, London.

[4] K. Fang, and Y. Zhang (1990). *Generalized Multivariate Analysis.* Springer-Verlag, Beijing.

[5] E. Gómez, M. A. Gómez-Villegas and J. M. Marín (1998). A Multivariate Generalization of the Power Exponential Family of Distributions. *Comm. in Statist. Th. and M.*, **27**, 589–600.

[6] A. K. Gupta, and T. Varga, (1993). *Elliptically Contoured Models in Statistics.* Kluwer Ac. Press. London.

[7] D. Kelker, (1970). Distribution Theory of Spherical Distributions and a Location-Scale Parameter Generalization. *Sankhya, A*, **32**, 419–430.

[8] K. V. Mardia, J. T. Kent and J. M. Bibby (1979). *Multivariate Analysis.* Academic Press, London.

[9] J. M. Marín, (1998). *Continuous elliptical and power exponential distributions with applications to linear dynamic models.* Ph. D. Thesis, Universidad Complutense de Madrid.