

# A MATRIX VARIATE GENERALIZATION OF THE POWER EXPONENTIAL FAMILY OF DISTRIBUTIONS

Key Words: vector distribution; elliptically contoured distribution; stochastic representation.

## ABSTRACT

This paper proposes a matrix variate generalization of the power exponential distribution family, which can be useful in generalizing statistical procedures in multivariate analysis and in designing robust alternatives to them. An example is added to show an application of the generalization.

## 1. INTRODUCTION

In this paper, we make a matrix variate generalization of the power exponential distribution and study its properties. The one-dimensional power exponential distribution was established in [1] and has been used in many studies about robust inference (see [2], [3]). A multivariate generalization was proposed in [4].

The power exponential distribution has proved useful to model random phenomena whose distributions have tails that are thicker or thinner than those of the normal distribution, and so to supply robust alternatives to many statistical procedures.

The location parameter of these distributions is the mean, so linear and nonlinear models can be easily constructed. The covariance matrix permits a structure that embodies the uniform and serial dependence (see [5]).

The power exponential multivariate distribution has been applied in several fields. An application to repeated measurements can be seen in [5]. It has also been applied to obtain robust models for nonlinear repeated measurements, in order to model dependencies among responses, as an alternative to models where the multivariate  $t$  distribution is used (see [6]).

In Bayesian network applications, these distributions have been used as an alternative to the mixture of normal distributions; some references in the field of speech recognition and image processing are [7], [8], [9], [10] and [11].

The matrix variate generalization defines an absolutely continuous subfamily of the matrix elliptical distribution; this subfamily includes the normal family.

Power exponential matrices can be used to generalize the sampling processes to the case where observations are not independent. Each row of a  $(p \times n)$  power exponential matrix  $X$  can be considered as an observation of an  $n$ -dimensional variable. The  $p$  rows have the same distribution but are not independent from each other. If needed, the parameters of the distribution of  $X$  can be chosen so that rows turn out to be not correlated.

In Section 2 elliptical and power exponential matrix distributions are defined. Section 3 studies the distribution of the transpose matrix and the quadratic form  $XX'$  as well as the characteristics of the power exponential matrices. Section 4 deals with affine transformations. Section 5 studies the marginal and conditional distributions and Section 6 shows an example.

## 2. DEFINITIONS

In this section we begin by defining the absolutely continuous elliptical matrix distribution. Then, we present the definition of the matrix variate power exponential distribution, and we show that it is a particular case of the matrix elliptical distribution.

Some words about the notation are in order. We use lower case letters for scalars and upper case letters for matrices and vectors. We use expressions like  $p \times n$ , mainly as subscripts, to refer to a  $(p \times n)$  matrix, while subscripts like  $pn$  refers to vectors whose dimension is  $p$  times  $n$ . For any  $(p \times n)$  matrix  $X$  we shall denote  $\text{Vec}(X) = (X'_1, \dots, X'_n)'$ , where  $X_1, \dots, X_n$  are the column vectors of  $X$ . Properties of function  $\text{Vec}$  may be found in [12].

For the general matrix variate elliptical distributions, several definitions have been given in the literature (see [12], [13]). We concentrate on the absolutely continuous ones and establish the following definition upon a parameter based on density functions:

**Definition 1. (Absolutely continuous matrix variate elliptical distribution).** *An absolutely continuous random  $(p \times n)$  matrix  $X$ , has a  $(p \times n)$ -matrix variate elliptical distribution with parameters  $M$ , a  $(p \times n)$  matrix;  $\Sigma$ , a  $(p \times p)$  definite positive matrix;  $\Phi$  a  $(n \times n)$  definite positive matrix; and  $g$  a non negative Lebesgue measurable function on  $[0, \infty)$  such*

that

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

if

$$\text{Vec}(X') \sim E_{pn}(\text{Vec}(M'), \Sigma \otimes \Phi, g), \quad (1)$$

where  $E_{pn}$  is the vector variate  $pn$ -dimensional elliptical distribution established in [14]. This is equivalent to saying that the density function of  $X$  is

$$f(X; M, \Sigma, \Phi, g) = \frac{\Gamma\left(\frac{pn}{2}\right)}{\pi^{\frac{pn}{2}} \int_0^\infty t^{\frac{pn}{2}-1} g(t) dt} |\Sigma|^{-\frac{1}{2}n} |\Phi|^{-\frac{1}{2}p} g\left[\text{tr}\left((X - M)' \Sigma^{-1} (X - M) \Phi^{-1}\right)\right]. \quad (2)$$

We will use the following notation  $X \sim \text{ME}_{p \times n}(M, \Sigma, \Phi, g)$ .

We will define the matrix variate power exponential distribution as a generalization of the multivariate (vector) power exponential distribution. The vector distribution (see [4]) is defined as follows: an absolutely continuous random vector  $X = (x_1, \dots, x_p)'$  is said to have a  $p$ -variate power exponential distribution with parameters  $\mu \in \mathfrak{R}^p$ ;  $\Sigma$ , a  $(p \times p)$  definite positive symmetric matrix; and  $\beta \in (0, \infty)$ , if its density function is

$$f(x; \mu, \Sigma, \beta) = \frac{p \Gamma\left(\frac{p}{2}\right)}{\pi^{\frac{p}{2}} \Gamma\left(1 + \frac{p}{2\beta}\right) 2^{1 + \frac{p}{2\beta}}} |\Sigma|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2} \left((x - \mu)' \Sigma^{-1} (x - \mu)\right)^\beta\right\}. \quad (3)$$

We use the following notation  $X \sim \text{PE}_p(\mu, \Sigma, \beta)$ .

Parameters  $\mu$  and  $\Sigma$  in (3) are location (mean) and scale parameters; whereas parameter  $\beta$  is related to kurtosis, which depends only on it. When  $\beta = 1$  the corresponding distribution is normal; thus, the parameter  $\beta$  shows the disparity from the normal distribution.

The next definition generalizes this distribution to random matrices.

**Definition 2. (Matrix variate power exponential distribution).** A random  $(p \times n)$  matrix  $X$  has a  $(p \times n)$ -variate power exponential distribution with parameters  $M$ , a  $(p \times n)$  matrix;  $\Sigma$ , a  $(p \times p)$  definite positive matrix;  $\Phi$  a  $(n \times n)$  definite positive matrix; and  $\beta \in (0, \infty)$ , if

$$\text{Vec}(X') \sim \text{PE}_{pn}(\text{Vec}(M'), \Sigma \otimes \Phi, \beta). \quad (4)$$

We will use the following notation  $X \sim \text{MPE}_{p \times n}(M, \Sigma, \Phi, \beta)$ .

The density function of  $X$  turns out to be

$$f(X; M, \Sigma, \Phi, \beta) = k |\Sigma|^{-\frac{1}{2}n} |\Phi|^{-\frac{1}{2}p} \exp \left\{ -\frac{1}{2} (\text{tr} ((X - M)' \Sigma^{-1} (X - M) \Phi^{-1}))^\beta \right\}, \quad (5)$$

where  $k = \frac{pn\Gamma(\frac{pn}{2})}{\pi^{\frac{pn}{2}} \Gamma(1+\frac{pn}{2\beta}) 2^{1+\frac{pn}{2\beta}}}$ .

The matrix variate power exponential distribution generalizes the multivariate one since for  $n = 1$  (5) is equal to (3).

The matrix variate power exponential distribution is a particular case of matrix elliptical distributions, obtained by taking  $g(t) = \exp \{-\frac{1}{2}t^\beta\}$  in definition 1; in fact, it belongs to the family of matrix variate symmetric Kotz type distributions (see [13]).

If we set  $\beta = 1$  in (5), we obtain the matrix variate normal distribution; thus the parameter  $\beta$  shows the disparity from the normal distribution as before. For  $\beta = \frac{1}{2}$  we have a matrix generalization of the double exponential distribution.

### 3. DISTRIBUTION OF THE QUADRATIC FORM AND PROBABILISTIC CHARACTERISTICS

In this section we study the most important characteristics of the matrix variate power exponential distribution: the distribution of the transpose matrix and the quadratic form  $XX'$ ; its probabilistic characteristics and stochastic representation. We also deal with the simulation of the distribution. Several of these results are a direct generalization of those from the multivariate (vector) power exponential distribution, shown in [4].

The next theorem, concerning the distribution of the transpose matrix, follows immediately from definition 2.

**Theorem 1. (Distribution of the transpose matrix).** *If  $X \sim \text{MPE}_{p \times n}(M, \Sigma, \Phi, \beta)$  then  $X' \sim \text{MPE}_{n \times p}(M', \Phi, \Sigma, \beta)$ .*

The next theorem shows the distribution of the quadratic form  $XX'$ , which may be considered as an extension of the Wishart distribution (see, for example, [16]).

**Proposition 1. (Distribution of the quadratic form).** *If  $X \sim \text{MPE}_{p \times n}(0, I_p \otimes I_n, \beta)$ ,*

then density function of the quadratic form  $W = XX'$  is

$$\frac{pn\Gamma\left(\frac{pn}{2}\right)}{\Gamma_p\left(\frac{n}{2}\right)\Gamma\left(1+\frac{pn}{2\beta}\right)}2^{1+\frac{pn}{2\beta}}|W|^{\frac{1}{2}(n-p-1)}\exp\left\{-\frac{1}{2}[\text{tr}(W)]^\beta\right\}. \quad (6)$$

*Proof.* It follows from Lemma 5.1.1 of [13] by substituting the function  $f$  by the corresponding one of the matrix variate power exponential distribution. ■

If we make  $\beta = 1$  in (6), we obtain the standard Wishart distribution, whose scale parameter is  $I_n$ .

The next theorem shows the probabilistic characteristics of the matrix  $X$ . It is understood that  $\text{Var}[X]$  means  $\text{Var}[\text{Vec}(X')]$ , the covariance matrix of the vector  $\text{Vec}(X')$ ; it is also understood that the asymmetry and kurtosis coefficients  $\gamma_1$  and  $\gamma_2$  of a random matrix  $X$  are defined as

$$\begin{aligned} \gamma_1[X] &= \text{E}\left[\left((\text{Vec}(X') - \text{Vec}(M'))' \text{Var}[X]^{-1} (\text{Vec}(Y') - \text{Vec}(M'))\right)^3\right], \\ \gamma_2[X] &= \text{E}\left[\left((\text{Vec}(X') - \text{Vec}(M'))' \text{Var}[X]^{-1} (\text{Vec}(X') - \text{Vec}(M'))\right)^2\right], \end{aligned}$$

where  $Y$  is a random matrix independent of  $X$  and having the same distribution than  $X$ . The theorem follows directly from (4) and proposition 3.2 from [4].

**Theorem 2. (Probabilistic characteristics).** *If  $X \sim \text{MPE}_{p \times n}(M, \Sigma, \Phi, \beta)$  then,*

$$\begin{aligned} \text{E}[X] &= M, \\ \text{Var}[X] &= \frac{2^{\frac{1}{\beta}}\Gamma\left(\frac{pn+2}{2\beta}\right)}{pn\Gamma\left(\frac{pn}{2\beta}\right)}(\Sigma \otimes \Phi), \\ \gamma_1[X] &= 0, \\ \gamma_2[X] &= \frac{p^2n^2\Gamma\left(\frac{pn}{2\beta}\right)\Gamma\left(\frac{pn+4}{2\beta}\right)}{\Gamma^2\left(\frac{pn+2}{2\beta}\right)}. \end{aligned}$$

The covariance between any two elements  $x_{ij}$  and  $x_{kl}$  of  $X$  is  $\frac{2^{\frac{1}{\beta}}\Gamma\left(\frac{pn+2}{2\beta}\right)}{pn\Gamma\left(\frac{pn}{2\beta}\right)}\sigma_{ik}\phi_{jl}$ . If  $\Sigma$  is a diagonal matrix then the covariance between any two elements  $i$  and  $j$  belonging to different rows is 0. In this case we may say that the rows of  $X$  are not correlated with each other.

The stochastic representation of matrix  $X$  is shown in the next theorem.

**Theorem 3. (Stochastic representation).** *If  $X \sim \text{MPE}_{p \times n}(M, \Sigma, \Phi, \beta)$  then  $X$  has the same distribution as*

$$M + rA'U^{(p \times n)}B', \quad (7)$$

where  $A$  is any  $(p \times p)$  matrix such that  $A'A = \Sigma$ ,  $B$  is any  $(n \times n)$  matrix such that  $BB' = \Phi$ ,  $r$  is an absolutely continuous non negative random variable, whose density function is

$$h(r) = \frac{pn}{\Gamma\left(1 + \frac{pn}{2\beta}\right) 2^{\frac{pn}{2\beta}}} r^{pn-1} \exp\left\{-\frac{1}{2}r^{2\beta}\right\} I_{(0,\infty)}(r), \quad (8)$$

and  $U^{(p \times n)}$  is a random  $(p \times n)$  matrix, independent of  $r$ , such that  $\text{Vec}(U^{(p \times n)})$  is uniformly distributed on the unit sphere in  $\Re^{pn}$ .

*Proof.* From proposition 3.1 of [4], the vector  $\text{Vec}(X')$  has the same distribution as  $\text{Vec}(M') + r(A' \otimes B)U^{(pn)}$ , where  $r$ ,  $A$  and  $B$  are as in (7) and  $U^{(pn)} = \text{Vec}(U^{(p \times n)})$ . The theorem then follows from properties of operator  $\text{Vec}$ . ■

The moments of  $r$ , obtained by applying corollary 3.1(i) from [4], are

$$\text{E}[r^s] = \frac{2^{\frac{s}{2\beta}} \Gamma\left(\frac{pn+s}{2\beta}\right)}{\Gamma\left(\frac{pn}{2\beta}\right)};$$

also, the distribution of  $t = r^{2\beta}$  is *Gamma*  $\left(\frac{1}{2}, \frac{pn}{2\beta}\right)$  and the distribution of  $z = \text{tr}((X - M)' \Sigma^{-1} (X - M))$  is the same as that of  $r^2$ .

$X$  can be simulated by means of its stochastic representation (7) as follows: (a) determine two matrices  $A$  and  $B$  such that  $A'A = \Sigma$  and  $BB' = \Phi$ ; (b) simulate a vector  $U^{(pn)}$  uniformly distributed on the unit sphere of  $\Re^{pn}$  and set a matrix  $U^{(p \times n)}$  such that  $\text{Vec}(U^{(p \times n)}) = U^{(pn)}$ ; (c) generate an observation of a random variable  $r$  with density (8); (d) make  $X = M + rA'U^{(p \times n)}B'$ . We have written a program in FORTRAN 90 implementing this process. This is available upon request.

#### 4. AFFINE TRANSFORMATIONS

In this section we study bijective and dimension-reducing affine transformations of a power exponential matrix distribution. These will be useful to derive the marginal distributions in the next section.

**Theorem 4. (Bijective transformations).** *If  $X \sim \text{MPE}_{p \times n}(M, \Sigma, \Phi, \beta)$  and  $Y = AXB + C$  where  $A$  and  $B$  are non singular ( $p \times p$ ) and ( $n \times n$ ) matrices and  $C$  is a ( $p \times n$ ) matrix, then  $Y \sim \text{MPE}_{p \times n}(AMB + C, A\Sigma A', B'\Phi B, \beta)$ .*

*Proof.*  $\text{Vec}(Y')$  is obtained from  $\text{Vec}(X')$  as  $\text{Vec}(Y') = \text{Vec}(C') + (A \otimes B')$ ; on the other hand,  $\text{Vec}(X') \sim \text{PE}_{pn}(\text{Vec}(M'), (\Sigma \otimes \Phi), \beta)$  and  $(A \otimes B')$  is a non singular matrix; then by proposition 4.1 from [4], we have that  $\text{Vec}(Y') \sim \text{PE}_{pn}(\text{Vec}(C') + (A \otimes B') \text{Vec}(M'), (A\Sigma A') \otimes (B'\Sigma B), \beta)$  and the theorem follows from definition 2. ■

**Theorem 5. (Dimension-reducing affine transformations).** *If  $X \sim \text{MPE}_{p \times n}(M, \Sigma, \Phi, \beta)$  and  $Y = AXB + C$  where  $A$  is ( $q \times p$ ) matrix,  $q < p$ ,  $\text{rk}(A) = q$ ,  $B$  is ( $n \times m$ ) matrix,  $m < n$ ,  $\text{rk}(B) = m$  and  $C$  is a ( $q \times m$ ) matrix, then  $Y \sim \text{ME}_{q \times m}(AMB + C, A\Sigma A', B'\Phi B, g_Y)$  with*

$$g_Y(t) = \int_0^\infty w^{\frac{pn-qm}{2}-1} \exp\left\{-\frac{1}{2}(t+w)^\beta\right\} dw. \quad (9)$$

*Proof.* Since  $\text{Vec}(X') \sim \text{PE}_{pn}(\text{Vec}(M'), (\Sigma \otimes \Phi), \beta)$  and

$$\text{Vec}(Y') = \text{Vec}(C') + (A \otimes B') \text{Vec}(X'), \quad (10)$$

then by proposition 4.2 from [4], we have that  $\text{Vec}(Y') \sim \text{E}_{qm}(\text{Vec}(C') + (A \otimes B') \text{Vec}(M'), (A\Sigma A') \otimes (B'\Sigma B), g)$  where  $g$  is as in (9). Then, the theorem follows from definition 1. ■

The next theorem shows the probabilistic characteristics of matrix  $Y$ .

**Theorem 6. (Probabilistic characteristics).** *Under the conditions of theorem 5,*

$$\begin{aligned} E[Y] &= AMB + C, \\ \text{Var}[Y] &= \frac{2^{\frac{1}{\beta}} \Gamma\left(\frac{pn+2}{2\beta}\right)}{pn \Gamma\left(\frac{pn}{2\beta}\right)} ((A\Sigma A') \otimes (B'\Phi B)), \\ \gamma_1[Y] &= 0, \\ \gamma_2[Y] &= \frac{pnqm(qm+2)}{pn+2} \frac{\Gamma\left(\frac{pn}{2\beta}\right) \Gamma\left(\frac{pn+4}{2\beta}\right)}{\Gamma^2\left(\frac{pn+2}{2\beta}\right)}. \end{aligned}$$

*Proof.* It follows from (10) and proposition 4.4 from [4]. ■

## 5. MARGINAL AND CONDITIONAL DISTRIBUTIONS

In this section we study the marginal distributions and characteristics of rows, columns and single elements, as well as the conditional distributions. These results are a direct generalization of those from the multivariate (vector) power exponential distribution, shown in [4].

A note about the notation: to denote the dimension of a submatrix we use the same letters as for the original matrix with appropriate subscripts. In theorems 11 and 12 we use the letter  $q$  to denote a quadratic form.

**Theorem 7. (Row marginal distribution).** *If  $X \sim \text{MPE}_{p \times n}(M, \Sigma, \Phi, \beta)$  and*

$$X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}, \quad M = \begin{pmatrix} M_1 \\ M_2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix},$$

where  $X_1$  and  $M_1$  are  $(p_1 \times n)$  matrices and  $\Sigma_{11}$  is a  $(p_1 \times p_1)$  matrix, then  $X_1 \sim \text{ME}_{p_1 \times n}(M_1, \Sigma_{11}, \Phi, g_1)$  where

$$g_1(t) = \int_0^\infty w^{\frac{(p-p_1)n}{2}-1} \exp\left\{-\frac{1}{2}(t+w)^\beta\right\} dw. \quad (11)$$

*Proof.* If we consider the block matrix  $A = \begin{pmatrix} I_{p_1} & 0_{p_1 \times (p-p_1)} \end{pmatrix}$ , then  $X_1 = AX$  and the statement is immediately obtained by applying theorem 5. ■



**Theorem 8. (Column marginal distribution).** *If  $X \sim \text{MPE}_{p \times n}(M, \Sigma, \Phi, \beta)$  and*

$$X = \begin{pmatrix} X_1 & X_2 \end{pmatrix}, \quad M = \begin{pmatrix} M_1 & M_2 \end{pmatrix}, \quad \Phi = \begin{pmatrix} \Phi_{11} & \Phi_{12} \\ \Phi_{21} & \Phi_{22} \end{pmatrix},$$

where  $X_1$  and  $M_1$  are  $(p \times n_1)$  matrices and  $\Phi_{11}$  is a  $(n_1 \times n_1)$  matrix, then  $X_1 \sim \text{ME}_{p \times n_1}(M_1, \Sigma, \Phi_{11}, g_1)$  where

$$g_1(t) = \int_0^\infty w^{\frac{p(n-n_1)}{2}-1} \exp\left\{-\frac{1}{2}(t+w)^\beta\right\} dw. \quad (12)$$

*Proof.* Since  $X' \sim \text{MPE}_{n \times p}(M', \Phi, \Sigma, \beta)$ , the result is immediate from theorem 7. ■

**Theorem 9. (Single element marginal distribution).** *If  $X \sim \text{MPE}_{p \times n}(M, \Sigma, \Phi, \beta)$  then  $x_{ij} \sim \text{MPE}_{1,1}(m_{ij}, \sigma_{ii}, \phi_{jj}, g_{ij})$  where*

$$g_{ij}(t) = \int_0^\infty w^{\frac{pn-1}{2}-1} \exp\left\{-\frac{1}{2}(t+w)^\beta\right\} dw. \quad (13)$$

*Proof.* It follows by applying theorem 7 and theorem 8 in turn. ■

**Theorem 10. (Marginal probabilistic characteristics).** *Let  $X \sim \text{MPE}_{p \times n}(M, \Sigma, \Phi, \beta)$ .*

(i) *If  $X_1$  is as in theorem 7 then,*

$$\begin{aligned} \mathbb{E}[X_1] &= M_1, \\ \text{Var}[X_1] &= \frac{2^{\frac{1}{\beta}} \Gamma\left(\frac{pn+2}{2\beta}\right)}{pn \Gamma\left(\frac{pn}{2\beta}\right)} \Theta, \\ \gamma_1[X] &= 0, \\ \gamma_2[X] &= \frac{pn \Gamma\left(\frac{pn}{2\beta}\right) \Gamma\left(\frac{pn+4}{2\beta}\right)}{(pn+2) \Gamma^2\left(\frac{pn+2}{2\beta}\right)} d, \end{aligned}$$

where  $M_1$  is as in theorem 7,  $\Theta = \Sigma_{11} \otimes \Phi$  and  $d = p_1 n (p_1 n + 2)$ .

(ii) *If  $X_1$  is as in theorem 8 then the formulae in (i) hold with  $M_1$  as in theorem 8,  $\Theta = \Sigma \otimes \Phi_{11}$  and  $d = pn_1 (pn_1 + 2)$ .*

(iii) *If  $x_{ij}$  is as in theorem 9 then the formulae in (i) hold with  $M_1 = m_{ij}$ ,  $\Theta = \sigma_{ii} \phi_{jj}$  and  $d = 3$ .*

*Proof.* Since  $X_1 = AX$  with  $A = \begin{pmatrix} I_{p_1} & 0_{p_1 \times (p-p_1)} \end{pmatrix}$  part (i) is immediately obtained from theorem 6. Since  $X' \sim \text{MPE}_{n \times p}(M', \Phi, \Sigma, \beta)$ , part (ii) follows immediately from part (i). Part (iii) is obtained by applying part (i) and part (ii). ■

The  $i^{\text{th}}$  row of the matrix has the covariance matrix  $\frac{2^{\frac{1}{\beta}} \Gamma\left(\frac{pn+2}{2\beta}\right)}{pn\Gamma\left(\frac{pn}{2\beta}\right)} \sigma_{ii} \Phi$  and the  $j^{\text{th}}$  column has the covariance matrix  $\frac{2^{\frac{1}{\beta}} \Gamma\left(\frac{pn+2}{2\beta}\right)}{pn\Gamma\left(\frac{pn}{2\beta}\right)} \phi_{jj} \Sigma$ .

According to theorem 7, every row of  $X$  has the same distribution. We have also seen that if  $\Sigma$  is a diagonal matrix, then the rows are not correlated. They turn out to be independent if (and only if)  $\beta = 1$ , that is, if the distribution of  $X$  is normal.

We see that the power exponential distribution is not closed under marginalization, since the marginal distribution of each row is elliptical but not power exponential; nevertheless, in some cases, this elliptical distribution may be replaced, as an approximation, by the power exponential distribution with the same moments.

The following theorems, about conditional distributions, show that the regression function is linear, as it happens in the normal case.

**Theorem 11. (Conditional distributions).** *Let  $X \sim \text{MPE}_{p \times n}(M, \Sigma, \Phi, \beta)$  and  $X, M$  and  $\Sigma$  be partitioned as in theorem 7. The distribution of  $X_2$  conditional by  $X_1 = \tilde{X}_1$ , is  $\left(X_2 \mid X_1 = \tilde{X}_1\right) \sim \text{ME}_{(p-p_1) \times n}(M_{2.1}, \Sigma_{22.1}, \Phi, g_{2.1})$  where  $M_{2.1} = M_2 + \Sigma_{21} \Sigma_{11}^{-1} (\tilde{X}_1 - M_1)$ ,  $\Sigma_{22.1} = \Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12}$ ,*

$$g_{2.1}(t) = \exp \left\{ -\frac{1}{2} (t + q)^\beta \right\}, \quad (14)$$

and  $q = \left( \text{Vec} \left( \tilde{X}_1 - M_1 \right) \right)' (\Sigma_{11} \otimes \Phi)^{-1} \text{Vec} \left( \tilde{X}_1 - M_1 \right)$ .

*Proof.* Since  $\text{Vec}(X') \sim \text{PE}_{pn}(\text{Vec}(M'), (\Sigma \otimes \Phi), \beta)$ , the statement is obtained by applying proposition 5.1(ii) from [4]. ■

**Theorem 12. (Conditional probabilistic characteristics).** *Under the same hypotheses*

and notations as in theorem 11 we have

$$\begin{aligned} \mathbb{E} \left[ X_2 \mid X_1 = \tilde{X}_1 \right] &= M_{2,1}, \\ \text{Var} \left[ X_2 \mid X_1 = \tilde{X}_1 \right] &= \frac{1}{(p - p_1) n} \frac{\psi(0)}{\psi(-1)} \Sigma_{22,1}, \\ \gamma_1 \left[ X_2 \mid X_1 = \tilde{X}_1 \right] &= 0, \\ \gamma_2 \left[ X_2 \mid X_1 = \tilde{X}_1 \right] &= (p - p_1)^2 n^2 \frac{\psi(1) \psi(-1)}{(\psi(0))^2}, \end{aligned}$$

where we use the notation  $\psi(r) = \int_0^\infty t^{\frac{(p-p_1)n}{2}+r} \exp \left\{ (t+q)^\beta \right\} dt$ .

*Proof.* By theorem 11,  $\text{Vec}(X_2' \mid X_1 = \tilde{X}_1)$  has an elliptical distribution. The results follow then directly from the probabilistic characteristics of elliptical distributions (see, for example, theorem 9 in [15]). ■

## 6. EXAMPLE

We make use of the set of data about air pollution in 80 U.S. cities, collected in 1960 by [17] and shown in [18], to test the normality of their distribution. We only consider a subsample of 12 cities (this subsample was used by [18] for other purposes) and the variables *TMR* (total mortality rate), *SMEAN* (arithmetic mean of biweekly sulfate readings ( $\mu_g/m^3 \times 10$ )), and *PMEAN* (arithmetic mean of biweekly suspended particulate reading ( $\mu_g/m^3 \times 10$ )). The observed values are in table 1.

	<i>TMR</i>	<i>SMEAN</i>	<i>PMEAN</i>
Buffalo	1012	114	131
San Francisco	925	62	70
Jackson	928	52	77
Nashville	919	160	130
Fresno	845	34	119
Allentown	1059	146	135
El Paso	618	87	150
Toledo	1031	86	104
Columbus	877	161	119
Birmingham	943	145	146
Las Vegas	727	79	145
South Bend	888	77	90

Table 1. Air Pollution Data

We do not assume that observations are independent among cities, but only that they are non correlated, and check them for normality. More precisely, we assume that the joint distribution of the whole set of observations in the 12 cities is  $\text{MPE}_{3 \times 12}(\mu v', \Sigma, I_{12}, \beta)$ , where  $\mu$  is an unknown tridimensional vector and  $v$  is the 12-dimensional vector  $v = (1, \dots, 1)'$  (this involves that the means are the same for each city). The unknown parameters  $\Sigma$  and  $\beta$  are as in the definition 2. As for the parameter  $\Phi$ , we have taken  $\Phi = I_{12}$  as the simplest way to reflect the lack of correlation; other diagonal matrices may also be taken. Thus, we deal with the parameters  $\mu$ ,  $\Sigma$  and  $\beta$ . We are interested mainly in  $\beta$ , which reflects the non normality of the distribution.

We start by calculating the maximum likelihood estimate of  $\beta$ .

It is not possible to estimate the three parameters at once by maximum likelihood, because the size of the sample is only 1 (see [13], section 7.1 for a similar case). So we decide to proceed in two steps. In the first stage a previous pilot sample is taken, that is, another

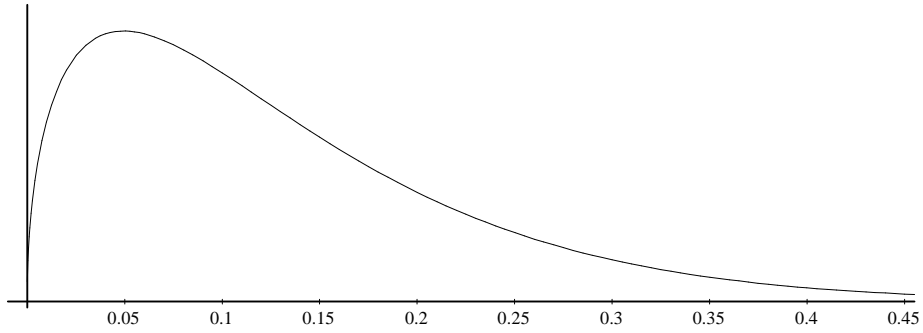


Figure 1: Logarithm of the likelihood function of  $\beta$ .

set of 12 random cities from the original data is used; we considered  $\beta$  fixed and estimated  $\mu$  and  $\Sigma$  as functions of  $\beta$  by using theorem 7.1.4 from [13]. We obtained

$$\hat{\mu} = \begin{pmatrix} 897.67 \\ 100.25 \\ 118.0 \end{pmatrix},$$

$$\hat{\Sigma} = 3 \left( I_{12} - \frac{1}{12} vv' \right) \left( \frac{\beta}{36} \right)^{\frac{1}{\beta}}.$$

In the second stage, we substitute  $\hat{\mu}$ ,  $\hat{\Sigma}$  and the original sample data (table 1) for  $\mu$ ,  $\Sigma$  and  $X$  in the density function (5) and we obtain the likelihood function of  $\beta$ ; it is proportional to

$$f(x|\beta) = \frac{\left(\frac{18}{\beta}\right)^{\frac{18}{\beta}}}{\Gamma\left(1 + \frac{18}{\beta}\right)} \exp\left\{-\frac{18}{\beta} (2.8339)^\beta\right\}; \quad (15)$$

its logarithm has the shape shown in figure 1.

The value of  $\beta$  that maximizes the likelihood function turned out to be  $\beta = 0.049$ . Hence, the distribution seems to be far from normality (which corresponds to  $\beta = 1$ ) and its kurtosis is much higher.

Now, we test the point null hypothesis  $H_0 : \beta = 1$  against  $H_1 : \beta \neq 1$  from a Bayesian point of view. We take again the likelihood function of  $\beta$  as in (15) and use a generic prior

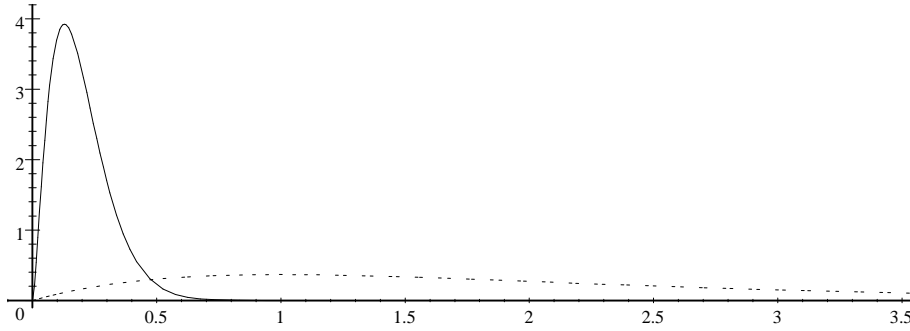


Figure 2: Prior (dotted line) and posterior distributions (solid line) of  $\beta$ .

giving a probability  $\pi_0$  to the point  $\beta = 1$  and spreading the remaining  $1 - \pi_0$  according to the gamma density  $g(\beta) = \beta \exp\{-\beta\} I_{(0, \infty)}(\beta)$  (whose mode is 1), shown by the dotted line in figure 2.

The Bayes factor

$$B = \frac{f(x|\beta = 1)}{\int_0^\infty g(\beta) f(x|\beta) d\beta},$$

turned out to be  $B = 7.566 \times 10^{-3}$ , a very small value: the data provide evidence against the null hypothesis. The posterior probability of the point  $\beta = 1$  is

$$\alpha_0 = \frac{1}{1 + \frac{1 - \pi_0}{\pi_0} \frac{1}{B}}.$$

If we set the usual value  $\pi_0 = 1/2$  (see [19], p. 151), we obtain  $\alpha_0 = 7.509 \times 10^{-3}$ . On the other hand, we can use

$$\pi_0 = \int_{1-\epsilon}^{1+\epsilon} \beta \exp\{-\beta\} d\beta,$$

according to [20]; by taking  $\epsilon = 0.05$  we have  $\pi_0 = 0.037$ ; the posterior probability is now  $\alpha_0 = 2.906 \times 10^{-4}$ . These results indicate the non normality of the distribution.

In the case  $\pi_0 = 1/2$  we had the prior probabilities  $P((0, 1)) = 0.132$ ,  $P(\{1\}) = 0.5$ ,  $P((1, \infty)) = 0.368$ , whereas the posterior ones are  $P((0, 1)|X) = 0.99249$ ,  $P(\{1\}|X) = 7.509 \times 10^{-3}$  and  $P((1, \infty)|X) \cong 4 \times 10^{-6}$ . Actually, the remaining  $1 - 7.509 \times 10^{-3} = 0.99249$

is spread according to the density shown by the solid line in figure 2: nearly all the mass is in the interval  $(0, 1)$ . Since  $P((0, 1) | X)$  is so large, the distribution seems to be more leptocurtic than the normal distribution.

## 7. CONCLUSIONS

The generalization of the power exponential distribution to a matrix distribution has been shown analytically tractable. We think it is useful to model random phenomena whose distributions have different tails than those of the normal distribution, such as the distribution of multivariate samples which are not independent from each other (due, for instance, to time or spatial dependency). It thus permits to pose robust alternatives to many processes as well as to test the normality of distributions, as shown in the example of Section 6. We hope that, just like to the multivariate power exponential distribution, the matrix variate version is used, since it is easy to be simulated and programmed.

## ACKNOWLEDGEMENTS

This research was supported in part by the Ministerio de Educación y Cultura, DGEIC (Spain) PB98-0797.

We would like to thank the editor and an anonymous referee for their very valuable comments.

## BIBLIOGRAPHY

- [1] Subbotin, M. On the law of frequency of errors. *Mathematicheskii Sbornik*, **1923**, 296–301.
- [2] Box, G. A note on regions for tests of kurtosis. *Biometrika*, **1953**, *40*, 465–468.
- [3] Box G.; Tiao, G. *Bayesian Inference in Statistical Analysis*, Addison-Wesley, Reading, **1973**.
- [4] Gómez E.; Gómez-Villegas M.A.; Marín J.M. A Multivariate Generalization of the Power Exponential Family of Distributions. *Comm. in Statist. Th. and M.*, **1998**, *27*, 589–600.

- [5] Lindsey J.K. Multivariate elliptically-contoured distributions for repeated measurements. *Biometrics*, **1999**, *56*, 1277–1280.
- [6] Lindsey J.K. and Lindsey P.J. Multivariate distributions with correlation matrices for nonlinear repeated measurements. **2002**, Tech. Report.
- [7] Basu, S. and Miccheli C. A., Parametric density estimation for the classification of acoustic feature vectors in speech recognition, in *Nonlinear Modeling: Advanced Black-Box Techniques* (Eds. J.A.K. Suykens and J. Vandewalle), Kluwer Acad. Press **1998**, 87–118.
- [8] Basu, S.; Miccheli C. A. and Olsen, P. Power exponential densities for the training and classification of acoustic feature vectors in speech recognition. *J. of graph. and comp. stat.* Jan. **2001**.
- [9] Aiazzi, B.; Alparone, L. and Baronti, S. Estimation based on entropy matching for generalized gaussian PDF modelling. *IEEE Sign. Processing Let.*, **1999**, *6*,(6) 138–140.
- [10] Sharifi, K. and León-García, A. Estimation of shape parameter for generalized gaussian distributions in subband decompositions of video. *IEEE trans. on circuits and systems for video Tech.*, **1995**, *5*,(1) 52–56.
- [11] Huang, J. and Mumford D. Statistics of natural images and models. *CVPR*, **1999**, 541–547.
- [12] Fang, K.; Zhang Y. *Generalized Multivariate Analysis*. Springer-Verlag, Beijing. **1990**.
- [13] Gupta, A.K.; Varga, T. *Elliptically Contoured Models in Statistics*. Kluwer Ac. Press. London. **1993**.
- [14] Johnson, M. (1987). *Multivariate Statistical Simulation*. Wiley. New York. **1987**.
- [15] Gómez E.; Gómez-Villegas M.A.; Marín J.M. On Absolutely Continuous Elliptical Distributions. *Techn. Report*, Universidad Complutense de Madrid, **1998**.
- [16] Anderson, T. W. *An introduction to Multivariate Statistical Analysis*. Wiley. New York.



**1958.**

[17] Gibbons D.; McDonald G.C.; Gunst R.F. The complementary use of regression diagnostics and robust estimators. *Naval Research Logistics*, **1987**, *34*, 1, Feb.

[18] Jobson, J.D. *Applied Multivariate Data Analysis*. Springer-Verlag. New York, **1992**.

[19] Berger, J.O. *Statistical Decision Theory and Bayesian Analysis*. Springer-Verlag. New York, **1985**.

[20] Gómez-Villegas M.A.; Gómez E. Bayes Factor in Testing Precise Hypotheses. *Comm. in Statist. Th. and M.*, **1992**, *21*, 1707-1715.