A NEW DEFINITION OF FORM-INVARIANCE MATRIX VARIATE DISTRIBUTIONS

M. Arashi
Department of Statistics, School of Mathematical Sciences, Shahrood University, Iran
Department of Statistics, Faculty of Natural and Agricultural Sciences, University of Pretoria
e-mail: *m_arashi_stat@yahoo.com*
and
A. Bekker [1]
Department of Statistics, Faculty of Natural and Agricultural Sciences, University of Pretoria
e-mail: *andriette.bekker@up.ac.za*
and
M. Ratnaparkhi
Wright State University, Dayton, Ohio

Key words: Dirac delta function, Elliptically contoured distribution, Form-invariance, Inverse Laplace transform.

Summary: Form-invariance is the most practical and desired property of weighted distributions. Weighted distributions commonly arise as univariate probability models in many fields of applications, for example, survival data analysis in medical science. However, the explicit global definition of "form-invariance" seems to be absent (to our knowledge) in spite of the fact that there is a growing use of such distributions during the last three decades. Therefore, in this note an effort is made to provide a new definition of form-invariance with reference to the matrix variate distributions. Obviously this definition and related implications are applicable to the multivariate and univariate distributions (as it should be), even though not necessarily to all families of probability distributions that exist, but definitely to those that are commonly used in practice. To demonstrate the validity of the last statement certain results, based on our new definition of form-invariance, are proved as the characterizations of the class of matrix elliptical distributions. Also considered in this note are certain extensions of our results, as may seem appropriate within the scope of the theory of matrix variate distributions. The concept of form-invariance is extended to marginal and conditional structures, as well as to the sum of independent matrix variates from elliptical distributions.

1. Introduction

During the last three decades the weighted versions of some commonly used univariate distributions were adopted as models in the analysis of size-biased data (meaning that the probability of inclusion of the observation depends on the magnitude (size) of that observation). For example, in medical

[1] Corresponding author.

AMS: PRIMARY 62H10, SECONDARY 62E15

science the researcher often has to deal with data that are available as size-biased or more commonly as the length-biased data. Bivariate weighted distributions are not that common but arise in some applications. There are some practical situations arising from image processing data where there seems to be an application for matrix variate weighted distributions.

The weighted distribution is defined as follows:

Let $X \sim f(x;\theta)$ where θ is a scalar or vector of parameters; $f(x;\theta)$ is referred to as the original distribution. Let $w(x) > 0$ be the function of x with $E[w(X)] < \infty$. Then, the weighted version of the original distribution denoted by $g(x;\theta)$ is defined as $g(x;\theta) \propto w(x)f(x;\theta)$.

In practice, if the collected data are known to be size-biased, then instead of $f(x;\theta)$ the distribution $g(x;\theta)$ is considered as a model with an appropriate weight function $w(x)$ and the inference about the parameter θ is now based on $g(x;\theta)$.

Obviously, unless $g(x;\theta)$ has certain desirable properties it cannot serve the purpose of conducting statistical inference on parameter θ. One such desirable property is referred to as the form-invariance of $g(x;\theta)$ when it is compared with the form of $f(x;\theta)$ as a function of variable X. However, there is no precise global definition or criterion of form-invariance in the related literature.

Therefore, in this note, as mentioned above, in view of the possible occurrence of size-biasedness in matrix variate data, we provide a new definition of the *form-invariance of matrix variate weighted distributions*. This definition, as to be expected, is applicable to the form-invariance of the weighted univariate and multivariate distributions as special cases of the matrix variate distribution. Since the nature of the discussion in this paper is mostly of mathematical nature, we will briefly describe a practical situation, where the weighted matrix variate distribution arises as a model for the available data.

The clustering of objects/subjects is required in many fields of applications such as medical science, psychology, marketing, physiology and many others. A typical scenario, in general, is as follows:

Suppose p-dimensional data are available on each of the N subjects that are to be classified into clusters and suppose the number of clusters is k_1. Various methods are used for clustering, for example, the k-means method (refer to Johnson and Wichern (2006) for clustering methods). Suppose the researcher is interested in using the k-means (nearest -neighbour) method. In order to apply the k-means method, the researcher is required to specify the number of variables to be used for clustering. Unfortunately, the choice of variables and the related precision of the clustering scheme, present a complicated problem, if the number of variables p is large. In particular, different methods are suggested for the choice of variables that should be used for clustering of subjects. Therefore, the researcher has to settle on a subset of variables in one way or another, hoping that this subset will work for the true clusters of N subjects. One simple solution is to use all the p-variables and obtain k_1 clusters. However, this is not an optimal solution due to the fact that, if the clustering problems are undertaken in a completely new field of research and if the implications of the variables in the clustering task is an important part to the experiment conducted for collecting the data on these p-variables, then there is a need for pinning down the most appropriate minimal subset of variables. (For example, such situations arise in the cluster analysis of different types of body tissues known to be responding to different kinds of impulses applied to them in neurological studies.) To arrive at such a minimal subset of q-variables $(q < p)$, the following data analysis method is considered.

- Step I. Use all the p-variables and obtain the fixed number of clusters ($k = k_1$) and the frequency distribution (# of subjects in each cluster).

- Step II. Select a subset of the p-variables consisting of q-variables using a suitable criterion supported by the objective of the study and the experimental conditions.

  Note that the selection of variables automatically causes selection biasedness, due to the experimental conditions and the nature of the collected /available data.

  Cluster analysis with k_1 clusters is conducted for these q-variables.

- Step III. Now, from the analyses done in Steps I and II for each of N subjects, we have two observations for each cluster frequency (namely the frequency with p-variables and the other with q-variables).

- Step IV. The cross-classification for each of the N subjects results in a $k_1 \times k_1$ contingency table- in which the cell frequencies are the values of the discrete random variables (r.v.s.), say X_{ij}'s. The resulting frequencies are the realizations of the r.v.s. arranged as a $k_1 \times k_1$ matrix, having a certain matrix variate distribution. Further, these frequencies, when expressed as proportions or the normalized r.v.s., are transformed into continuous variables defined on the corresponding support space.

Now, if we can identify the original distribution of such a discrete matrix variate distribution or of the corresponding matrix of proportions, we can associate a matrix weight function and derive a weighted version of the original distribution. This can be used for further analysis as may seem appropriate for the practical situation under consideration.

Remark 1 The contingency table (similar to those referred to in Step IV) and its variants are studied in machine learning (artificial intelligence) in order to minimize the misclassification errors. Such a matrix is referred to as a *confusion matrix*. The Bayesian analysis of the confusion matrix is often considered in related studies. The discussion about confusion matrices can be found in text books on machine learning and related applications. Here, we are suggesting the use of weighted matrix variate distributions as a method of correcting the misclassification errors. We may further add that in certain applications of univariate weighted distributions, one can consider the posterior distribution as a weighted version of the original distribution. In this respect, the above discussion, with reference to a confusion matrix, is an extension of similar univariate cases (see Bayarri and Berger, 1998).

For the convenience of the readers, the main features of our new definition are listed below.

(a) It is valid for scalar as well as vector variates with no change in their structure or basis.

(b) It can be used for non-positive definite matrices. Considering item (a), it can even be applied to negative random variables.

(c) The Student's t-distribution, commonly used in statistical inference, belongs to the form-invariance class.

Some preliminary results from the theory of matrix variate distributions are listed in Section 2. The main definition and related implications and applications are listed in Section 3. A few relevant extensions of the results obtained in Sections 2 and 3 are discussed in Section 4.

2. Preliminaries

For the purpose of this note we consider the matrix variate elliptically contoured (MEC) distributions.

Let $\mathbf{X}$ be an $n \times p$ random matrix, which can be expressed in terms of its elements, column and rows as

$$\mathbf{X} = (x_{ij}) = (\mathbf{x}_1, \cdots, \mathbf{x}_p) = (\mathbf{x}_{(1)}, \cdots, \mathbf{x}_{(n)})'. \tag{1}$$

Here $\mathbf{x}_{(1)}, \cdots, \mathbf{x}_{(n)}$ can be regarded as a sample of size n from a p-dimensional population.

Four classes of matrix variate elliptical distributions are defined and discussed in Dawid (1977) and Gupta and Varga (1993). In view of the objective of this paper we specifically consider the following case:

Definition 1 *The $n \times p$ random matrix $\mathbf{X}$ has a MEC distribution if its density has the form*

$$g(\mathbf{X}) = d_{n,p}|\mathbf{\Phi}|^{-\frac{p}{2}}|\Sigma|^{-\frac{n}{2}} f\left\{\mathbf{\Sigma}^{-1}(\mathbf{X}-\mathbf{1}\boldsymbol{\mu}')'\Phi^{-1}(\mathbf{X}-\mathbf{1}\mu')\right\} \tag{2}$$

where $\boldsymbol{\mu} \in R^p$ and $\mathbf{\Sigma}$ is a $p \times p$ semi-positive definite matrix. This distribution is denoted by $\mathbf{X} \sim E_{n,p}(\boldsymbol{\mu}, \Phi \otimes \Sigma, f)$. For notational convenience we may also use $E_{n,p}(\boldsymbol{\mu}, \Phi, \mathbf{\Sigma}, f)$ when needed.

Definition 1 imposes the condition $f(\mathbf{AB}) = f(\mathbf{BA})$ on the density generator f for any $p \times p$ positive definite symmetric matrices $\mathbf{A}$ and $\mathbf{B}$. Throughout this paper, without loss of generality, we will be using the full form of the tr operation from the argument of $f(\cdot)$.
Two well-known examples of MEC distributions are:

(a) Matrix Variate Normal (MN) Distribution
$\mathbf{X} \in \mathbb{R}^{n \times p}$ has MN distribution, with mean $\mathbf{M} = \mathbf{1}\boldsymbol{\mu}'$, row and column covariance matrices Φ and $\mathbf{\Sigma}$ respectively, denoted by $\mathbf{X} \sim N_{n,p}(\mathbf{M}, \Sigma, \mathbf{\Phi})$, if its density is given by

$$f(\mathbf{X}) = \frac{|\Phi|^{-\frac{p}{2}}|\mathbf{\Sigma}|^{-\frac{n}{2}}}{(2\pi)^{\frac{np}{2}}} \exp\left\{-\frac{1}{2}tr\left[\mathbf{\Sigma}^{-1}(\mathbf{X}-\mathbf{M})'\Phi^{-1}(\mathbf{X}-\mathbf{M})\right]\right\}.$$

(b) Matrix Variate Student-t (MT) Distribution
$\mathbf{X} \in \mathbb{R}^{n \times p}$ has MT distribution, with mean $\mathbf{M} = \mathbf{1}\mu'$, row and column scale matrices $\mathbf{\Phi}$ and $\mathbf{\Sigma}$ respectively and ν degrees of freedom denoted by $\mathbf{X} \sim T_{n,p}(\mathbf{M}, \Sigma, \mathbf{\Phi}, \nu)$, if its density is given by

$$f(\mathbf{X}) = \frac{|\Phi|^{-\frac{p}{2}}|\Sigma|^{-\frac{n}{2}}}{g_{n,p}} \left|\mathbf{I}_n + \mathbf{\Sigma}^{-1}(\mathbf{X}-\mathbf{M})'\Phi^{-1}(\mathbf{X}-\mathbf{M})\right|^{-\frac{n+p+\nu-1}{2}},$$

where

$$g_{n,p} = \frac{(\pi)^{\frac{np}{2}}\Gamma_p\left(\frac{\nu+p-1}{2}\right)}{\Gamma_p\left(\frac{\nu+n+p-1}{2}\right)}.$$

Remark 2 For $\mathbf{\Phi} = \frac{\beta}{2}\mathbf{I}$, $n = 1$, the distribution in (b) above simplifies to the new mixture representation for multivariate t, proposed by Iranmanesh et al. (2012). For other possible examples, refer to Arashi et al. (2013).

The MEC distribution can always be expressed in an integral form of a matrix variate normal distribution with *weight function* as given in the following theorem.

Theorem 1 *Let $\mathbf{X} \sim E_{n,p}(\boldsymbol{\mu}, \mathbf{I}_n \otimes \mathbf{\Sigma}, f)$ where the density $g(\mathbf{X})$ of $\mathbf{X}$ is defined by*

$$g(\mathbf{X}) = |\mathbf{\Sigma}|^{-\frac{n}{2}} h\left[tr\mathbf{\Sigma}^{-1}(\mathbf{X}-\mathbf{1}\mu')'(\mathbf{X}-\mathbf{1}\mu')\right].$$

If $h(t), t \in [0,\infty)$ has the inverse Laplace transform (denoted by $\mathscr{L}[h(t)]$), then we have

$$g(\mathbf{X}) = \int_0^\infty w(z) f_{N(\boldsymbol{\mu},z^{-1}\mathbf{I}_n\otimes\mathbf{\Sigma})}(\mathbf{X})dz \tag{3}$$

where $f_{N(\mu,z^{-1}\mathbf{I}_n\otimes\Sigma)}(\mathbf{X})$ stands for the density of the $n \times p$ random matrix $\mathbf{X}$ distributed as a matrix variate normal with the mean matrix $\mathbf{1}\boldsymbol{\mu}'$ and the covariance matrix $z^{-1}\mathbf{I}_n \otimes \mathbf{\Sigma}$, and $w(z)$ is the weight function given by

$$w(z) = (2\pi)^{np/2} z^{-np/2} \mathscr{L}[h(2z)]. \tag{4}$$

For the proof refer to Gupta and Varga (1993).

Remark 3 In Theorem 1 it is stated that f is the density of $N_{n,p}\left(\boldsymbol{\mu}, z^{-1}\mathbf{I}_n \otimes \mathbf{\Sigma}\right)$, it can also be verified from the proof that f can have one of the following densities:

(a) $N_{n,p}\left(\boldsymbol{\mu}, z^{-1}\mathbf{I}_n, \mathbf{\Sigma}\right)$ or

(b) $N_{n,p}\left(\boldsymbol{\mu}, \mathbf{I}_n, z^{-1}\mathbf{\Sigma}\right)$ or

(c) $N_{n,p}\left(\boldsymbol{\mu}, z^{-\frac{1}{2}}\mathbf{I}_n, z^{-\frac{1}{2}}\mathbf{\Sigma}\right)$.

This fact enables us to adopt any of these representations when needed for practical use.

3. Characterization

In this section, we define a form-invariance structure for matrix variate distributions. Then we obtain a characterization of the class of distributions that satisfies our definition.

Definition 2 *The random matrix $\mathbf{X} \in \mathbb{R}^{n\times p}$, having the density function g, has a form-invariance distribution if its density can be expressed as*

$$g(\mathbf{X};\Sigma) = \int_0^\infty \delta(t-1) g(\mathbf{X};t^{-1}\Sigma)dt, \tag{5}$$

where $\boldsymbol{\Sigma}$ is the column covariance matrix and $\delta(\cdot)$ is the Dirac delta or impulse function having the property $\int_{\mathbb{R}} h(x)\delta(x)dx = h(0)$, for every Borel-measurable function $h(\cdot)$.

In the following result we show that all matrix variate elliptical distributions satisfy Definition 2. In other words we characterize the class of distributions that synchronizes Definition 2.

Theorem 2 *As a result of Definition 2, all matrix variate elliptical distributions defined in Definition 1, satisfy the regularity condition (5).*

Proof. Suppose that $\mathbf{X} \sim E_{n,p}(\boldsymbol{\mu}, \boldsymbol{\Phi}\otimes\boldsymbol{\Sigma}, f)$. Then using Theorem 1 we have,

$$\begin{aligned}
&\textstyle\int_0^\infty \delta(t-1) g_{E(\mu,t^{-1}\boldsymbol{\Phi}\otimes\Sigma,f)}(X)dt \\
&= \textstyle\int_{t=0}^\infty \delta(t-1)\left(\int_{z=0}^\infty w(z) g_{N(\mu,t^{-1}z^{-1}\boldsymbol{\Phi}\otimes\boldsymbol{\Sigma})}(X)dz\right)\, dt \\
&= \textstyle\int_{z=0}^\infty w(z)\left(\int_{t=0}^\infty \delta(t-1) g_{N(\boldsymbol{\mu},t^{-1}z^{-1}\boldsymbol{\Phi}\otimes\boldsymbol{\Sigma})}(X)dt\right)\, dz \\
&= \textstyle\int_{z=0}^\infty w(z) g_{N(\boldsymbol{\mu},z^{-1}\boldsymbol{\Phi}\otimes\Sigma)}(X)dz \\
&= g(\mathbf{X};\boldsymbol{\Sigma}),
\end{aligned}$$

where the third equality comes from the fact that, based on Theorem 1, the weight function for the MN distribution is the Dirac delta function. The proof is complete. ∎

Theorem 2 shows that all matrix variate elliptical distributions are form-invariant. Thus all matrix spherical distributions (taking $\boldsymbol{\mu} = \mathbf{0}_{n\times p}$, $\Phi = \mathbf{I}_n$ and $\boldsymbol{\Sigma} = \mathbf{I}_p$) are also form-invariant in the sense of Definition 2. Next, we record some further results similar to those of Theorem 2.

4. More Properties

In this section, we extend the result for form-invariance to marginal and well as conditional structures. The sum of independent matrix elliptical variates is also considered.

Theorem 3 *If $\mathbf{X} \sim E_{n,p}(\boldsymbol{\mu}, \boldsymbol{\Phi}\otimes\boldsymbol{\Sigma}, f)$ is form-invariant, then so is its transform, $\mathbf{X}'$.*

The proof follows directly using Theorem 2.1.3 of Gupta and Varga (1993) and Theorem 2.
Using Theorem 3, we will obtain more properties for $\mathbf{X} \sim E_{n,p}(\boldsymbol{\mu}, \boldsymbol{\Sigma}\otimes\Phi, f)$.

Theorem 4 *If $\mathbf{X} \sim E_{n,p}(\mu, \boldsymbol{\Phi}\otimes\Sigma, f)$ is form-invariant and partition $\mathbf{X}, \boldsymbol{\mu}$ and Σ as*

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix}, \boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \ and\ \boldsymbol{\Sigma} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix},$$

where $\mathbf{X}_1$ is $q\times n$, $\boldsymbol{\mu}_1$ is $q\times n$, and Σ_{11} is $q\times q$, $1 \le q < p$, then $\mathbf{X}_1$ is also form-invariant.

Proof. By making use of Theorem 2.3.1 of Gupta and Varga (1993), $\mathbf{X}_1 \sim E_{n,q}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}\otimes\boldsymbol{\Phi}, f)$. Then the result follows from Theorem 1. ∎

A similar result using Theorem 4 can be also given by partitioning $\boldsymbol{\Phi}$.

Remark 4 Based on the results of Theorems 3 and 4, we have the following conclusions:

(a) Let $n = q = 1$, the scalar elliptical variate is form-invariant.

(b) Let $q = 1$, the vector elliptical variate is form-invariant.

Remark 5 As in Theorems 3 and 4, to show that a random matrix or its variate is form-invariant, it is sufficient to show that its distribution is matrix elliptical and then use Theorem 2. In other words, by making use of Theorem 2, if $\mathbf{X} \in \mathbb{R}^{n\times p}$ is form-invariant, then it has a matrix elliptical distribution and in order to show that $\mathbf{Y} = \delta(\mathbf{X}), \delta : \mathbb{R}^{n\times p} \to \mathbb{R}^{k\times p}$ is form-invariant, it is sufficient to demonstrate that $\mathbf{Y}$ has a matrix variate elliptical distribution.

In the sequel, we propose a series of properties based on the result of Gupta and Varga (1993).

Theorem 5 *(a) Let $\mathbf{X}$ be form-invariant. For any $\mathbf{A}$ and $\mathbf{B}$ of sizes $q \times p$ and $m \times n$ respectively, the distribution of $\mathbf{Y} = \mathbf{AXB}'$ is form-invariant.*

(b) Under the assumptions of Theorem 4, suppose rank($\mathbf{\Sigma}_{22}$) ≥ 1. Then $\mathbf{X}_1 | \mathbf{X}_2$ is form-invariant.

(c) Let $\mathbf{X} \sim E_{n,p}(\mu, \mathbf{\Phi} \otimes \mathbf{\Sigma}, f), i = 1,\ldots,k$ be k independent form-invariant random elliptical matrices. Then $\mathbf{Y} = \sum_{i=1}^{k} \mathbf{X}_i$ is form-invariant.

5. Discussion

Patil and Ord (1976) introduced a form-invariance property for the univariate case. If a weighted distribution is form-invariant then the statistical inference (for example, the properties of estimators) based on such distribution will not be unduly distorted. This is observed for some known univariate distributions discussed in Patil and Ord (1976) and in many other papers on weighted univariate distributions.

The concept of form-invariance for the matrix environment has not been addressed before. Therefore, we proposed in this paper a new form-invariance definition for *matrix variate* distributions. In general, our definition is not comparable to the one in Patil and Ord (1976) for the univariate case.

However, due to the characterization result given as Theorem 2, for a random variable having a form-invariant distribution, it is necessary that the inverse Laplace transform exists for the density generator of its distribution. According to Definition 2, the form-invariance concept is an integral operator which is applied to the original distribution.

The advantages of using this new definition are:

(a) there is no restrictions in our definition;

(b) it holds for scalar as well as vector variates with no change in its structure or basis and

(c) the support space for the random variable is not restricted to a non-negative variable.

Further research in this area is needed and is in progress. But, in view of the scope of this paper the related results are not included here for brevity.

Acknowledgements

This work is based upon research supported by the National Research Foundation, South Africa (Incentive Funding For Rated Researchers), VC Post-Doctoral Fellowship of the University of Pretoria, as well as Shahrood University, Iran. The authors would like to thank the anonymous referee and the associate editor's significant comments that have helped in improving the paper, as well as valuable discussions with Professor J.J.J. Roux.

References

ARASHI, M., IRANMANESH, ANIS, NOROUZIRAD, M., AND SALARZADEH JENATABADI, H. (2013). Bayesian analysis in multivariate regression models with conjugate priors. *Statistics*. DOI: 10.1080/02331888.2013.809720.

BAYARRI, M. J. AND BERGER, J. (1998). Robust Bayesian analysis of selection models. *Annals of Statistics*, **26**, 645–659.

DAWID, A. P. (1977). Spherical matrix distributions and a multivariate model. *Journal of the Royal Statistical Society, Series B*, **39** (2), 254–261.

GUPTA, A. K. AND VARGA, T. (1993). *Elliptically Contoured Models in Statistics*. Academic Press, Kluwer.

IRANMANESH, A., ARASHI, M., NAGAR, D. K., NADARAJAH, S., AND TABATABAEY, S. M. M. (2012). A new mixture representation for multivariate t. *Journal of Multivariate Analysis*, **107**, 227–231.

JOHNSON, R. A. AND WICHERN, D. W. (2006). *Applied Multivariate Statistical Analysis*. Prentice Hall, New Jersey.

PATIL, G. P. AND ORD, J. K. (1976). On size-biased sampling and related form-invariant weighted distributions. *Sankhyā*, **38**, 48–61.

Manuscript received, 2013-11-03, revised, 2014-05-10, accepted, 2014-06-16.