PROBABILITY
AND
MATHEMATICAL STATISTICS
Vol. 34, Fasc. 1 (2014), pp. 119–126

MULTIVARIATE PEARSON TYPE II DISTRIBUTION: STATISTICAL AND MATHEMATICAL FEATURES

BY

MAHNAZ KHALAFI (GORGAN) AND MAJID AZIMMOHSENI (GORGAN)

Abstract. In this article, we first review the main characterizations of multivariate Pearson type II distribution as a subclass of multivariate symmetric spherical distributions. Then we try to provide specific mathematical and statistical principles underlying the construction of this subclass of multivariate distributions.

2000 AMS Mathematics Subject Classification: Primary: 62H10; Secondary: 62H05.

Key words and phrases: Multivariate Pearson type II distribution, hypersphere, complex normal distribution, norm-scale invariant, simulation.

1. PRELIMINARIES

The class of multivariate symmetric spherical distributions has received extensive attention as a generalization of multivariate standard normal distribution. The fundamental characterizations and properties of this class of distributions can be found in [3]. The works by Arellano-Valle and Bolfarine [2], Volodin [12], Liang and Bentler [7], Arellano-Valle [1], and references therein have presented more characterizations of the class of multivariate symmetric spherical distributions and some particular subclasses of these distributions. In particular, characterizations of multivariate Pearson type II distribution ($MPII$) have been obtained by Fang et al. [3], Liang and Bentler [7], and Arellano-Valle [1].

The Pearson type II distribution plays a key role in the class of multivariate symmetric spherical distributions. Specifically, every marginal distribution of a multivariate symmetric distribution can be expressed in terms of the corresponding radial distribution and the Pearson type II distribution; see Arellano-Valle [1]. This important property and other mathematical and statistical features motivate us to focus on this distribution. Let us first recall some basic notions.

Multivariate symmetric spherical distributions are defined on the closure of an n-sphere. An n-sphere is a generalization of an ordinary sphere to an arbitrary dimension. From a geometrical point of view, an n-sphere of radius r is a locus of n-dimensional points with maximum distance r from the origin. An n-sphere is represented as follows:

$$S^n = \big\{(x_1, \ldots, x_n) \in R^n, \sum_{i=1}^{n} x_i^2 \leqslant r^2\big\}. \tag{1.1}$$

In a special case where $r = 1$, this shape is called a *unit* n-sphere. The volume $V_n(r)$ and surface area $S_n(r)$ of an n-sphere with radius r are given by

$$V_n(r) = \frac{\pi^{n/2}}{\Gamma(n/2+1)} r^n, \tag{1.2}$$

$$S_n(r) = \frac{2\pi^{n/2}}{\Gamma(n/2)} r^{n-1}. \tag{1.3}$$

Suppose $(U_1, \ldots, U_n)'$ is a random point from a unit n-sphere with joint density function given by

$$\begin{aligned} &f_{U_1,..,U_n}(u_1, \ldots, u_n; a) \\ &\quad = \frac{\Gamma(a+n/2)}{\Gamma(a)} \pi^{-n/2} \big(1 - \sum_{i=1}^{n} u_i^2\big)^{a-1} I_{(0,1)}\big(\sum_{i=1}^{n} u_i^2\big), \quad a > 0. \end{aligned} \tag{1.4}$$

Then $(U_1, \ldots, U_n)'$ follows multivariate Pearson type II distribution with parameters a and n, $MPII_n(a)$, where $\sum_{i=1}^{n} u_i^2$ is called a *spherical kernel* of this multivariate distribution. The class of multivariate Pearson type II distributions formed by their joint density functions is defined as follows:

$$C = \{f_{U_1,\ldots,U_n}(u_1, \ldots, u_n; a),\ a > 0,\ n \in N\}. \tag{1.5}$$

This paper is organized as follows. In Section 2 we present fundamental characterizations of the class of multivariate Pearson type II distributions. Some useful properties are also included in Section 2. In Section 3, we provide some mathematical and statistical features of this class of distributions.

2. CHARACTERIZATION OF MULTIVARIATE PEARSON TYPE II DISTRIBUTION

In this section, we present main characterizations of the $MPII$ distribution with detailed description. Some further useful results are also included.

THEOREM 2.1. *Suppose* $\mathbf{U} = (U_1, \ldots, U_n)'$ *follows a multivariate symmetric Pearson type II distribution* $MPII_n(a)$. *Then*:

(C1) $\mathbf{U}$ *has the stochastic representation* $\mathbf{U} \overset{d}{=} \sqrt{V}\mathbf{W}$, *where* $V \overset{d}{=} \|\mathbf{U}\|^2$ *is distributed as beta distribution* $B(n/2, a)$ *and the* n*-dimensional vector* $\mathbf{W}$ *is uniformly distributed on the surface of a unit* n*-sphere. Furthermore*, V *and* $\mathbf{W}$ *are independent.*

(C2) *For each* $1 \leqslant k < n$, $(U_{i_1}, \ldots, U_{i_k})'$ *is distributed as the multivariate Pearson type II distribution* $MPII_k\big(a + (n-k)/2\big)$.

(C3) *Let us assume that* $\mathbf{U} = (\mathbf{U}^{(1)}_{1\times k}, \mathbf{U}^{(2)}_{1\times(n-k)})'$. *Then the conditional distribution* $(1 - \|\mathbf{U}^{(2)}\|^2)^{-1/2}\mathbf{U}^{(1)}$ *given* $\mathbf{U}^{(2)} = \mathbf{u}^{(2)}$ *is distributed as multivariate Pearson type II distribution* $MPII_k(a)$.

(C4) *Let* $\mathbf{S} = (S_1, \ldots, S_n)'$ *be distributed as multivariate symmetric spherical distribution. Then every subvector* $\mathbf{S}^{(k)} = (S_{i_1}, \ldots, S_{i_k})'$, $1 \leqslant k \leqslant n-1$, *has the stochastic representation*

$$\mathbf{S}^{(k)} \stackrel{d}{=} \|\mathbf{S}\|\mathbf{U}^{(k)},$$

where $\mathbf{U}^{(k)}$ *is distributed as* $MPII_k(n-k)$.

P r o o f. See Arellano-Valle [1] and Liang and Bentler [7]. ∎

REMARK 2.1. *Generally, a random vector* $\mathbf{S} = (S_1, \ldots, S_n)'$ *with multivariate spherical symmetric distribution has the stochastic representation*

$$\mathbf{S} \stackrel{d}{=} \|S\|\mathbf{W}, \tag{2.1}$$

where the random vector $\mathbf{W}$ *is uniformly distributed on the surface of an* n*-sphere.*

REMARK 2.2. *If* $n = 2k$, *then two halves* $(U_1, \ldots, U_k)'$ *and* $(U_{k+1}, \ldots, U_{2k})'$ *have the same distribution. The joint density function of the first vector takes the form*

$$\begin{aligned} &f_{U_1,\ldots,U_k}(u_1, \ldots, u_k) \\ &\quad = \pi^{-k/2}\frac{\Gamma(a+3k/2)}{\Gamma(a+k/2)}\big(1 - \sum_{i=1}^{k} u_i^2\big)^{a+k/2-1} I_{(0,1)}\big(\sum_{i=1}^{k} u_i^2\big). \end{aligned} \tag{2.2}$$

REMARK 2.3. *Let us put* $V = \sum_{i=1}^{n} U_i^2$. *Then the conditional distribution of* $(U_1, \ldots, U_n)'$ *given* $\sqrt{V} = \sqrt{v}$ *is uniformly distributed over the surface of an* n*-sphere with radius* $\sqrt{v}$.

P r o o f. For the proof note that $V \sim B(n/2, a)$. Therefore,

$$f_{\sqrt{V}}(\sqrt{v}) = \frac{2}{B(n/2, a)} v^{(n/2-1/2)}(1-v)^{a-1},$$

so we have

$$\begin{aligned} &f_{U_1,\ldots,U_n|\sqrt{V}}(u_1, u_2, \ldots, u_n|\sqrt{v}) \\ &= \left[\frac{\Gamma(a+n/2)}{\Gamma(a)}\pi^{-n/2}(1-v)^{a-1}\right]\left[\frac{2}{B(n/2,a)}v^{(n/2-1/2)}(1-v)^{a-1}\right]^{-1} \\ &= \frac{1}{S_n(\sqrt{v})}, \end{aligned}$$

where $S_n(\cdot)$ is given by (1.3). ∎

3. SOME OTHER MATHEMATICAL AND STATISTICAL FEATURES OF MULTIVARIATE PEARSON TYPE II DISTRIBUTION

In this section, we obtain some results that lead us to more distinguished features of $MPII$ distributions.

3.1. More characterizations. Let us provide some properties of the class of $MPII$ distributions through the following theorem.

THEOREM 3.1. *Suppose* $\mathbf{U} = (U_1, U_2, \ldots, U_n)' \sim MPII_n(a)$. *Then*:

(P1) (Norm-scale invariance property). *The transformation of*

$$(T_{i_1}, \ldots, T_{i_k})' = \left(\frac{U_{i_1}}{\|\mathbf{U}\|}, \ldots, \frac{U_{i_k}}{\|\mathbf{U}\|}\right)', \quad 1 \leqslant k \leqslant n-1,$$

is distributed as $MPII_k\big(\frac{1}{2} + (n-1-k)/2\big)$.

(P2) *If* $\mathbf{W} = (W_1, \ldots, W_n)'$ *is uniformly distributed on the volume of a unit* n*-sphere*, *then*, *for* $1 \leqslant k \leqslant n-1$,

$$(W_{i_1}, \ldots, W_{i_k})' \sim MPII_k\left(1 + \frac{n-k}{2}\right).$$

Proof. (P1) By Theorem 2.1 (C2), the random vector $(T_{i_1}, \ldots, T_{i_k})'$ is an exchangeable random vector, so

$$(T_{i_1}, \ldots, T_{i_k})' \stackrel{d}{=} (T_1, \ldots, T_k)'$$

for any choice of $(i_1, \ldots, i_k)$. Therefore, it suffices to show that $(T_1, \ldots, T_{n-1})'$ has $MPII$ distribution. Put $V = \sum_{i=1}^n U_i^2$ and $Y = (T_1, \ldots, T_{n-1}, V)'$. Then the Jacobian of this transform is as follows:

$$J = \left|\frac{\partial u}{\partial y}\right| = \begin{vmatrix} \sqrt{v} & 0 & \ldots & 0 & \frac{t_1}{2\sqrt{v}} \\ 0 & \sqrt{v} & \ldots & 0 & \frac{t_2}{2\sqrt{v}} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \ldots & \sqrt{v} & \frac{t_{n-1}}{2\sqrt{v}} \\ \frac{-t_1\sqrt{v}}{\sqrt{1-\sum_{i=1}^{n-1} t_i^2}} & \frac{-t_2\sqrt{v}}{\sqrt{1-\sum_{i=1}^{n-1} t_i^2}} & \ldots & & \frac{\sqrt{1-\sum_{i=1}^{n} t_i^2}}{2\sqrt{v}} \end{vmatrix}$$

$$= \frac{v^{n/2-1}}{\sqrt{1-\sum_{i=1}^{n-1} t_i^2}}.$$

Hence we have

$$f_{T_1,\dots,T_{n-1},V}(t_1,\dots,t_{n-1},v)$$
$$= \frac{\Gamma(a+n/2)}{\Gamma(a)}\pi^{-n/2}\frac{v^{n/2-1}(1-v)^{a-1}}{\sqrt{1-\sum_{i=1}^{n-1}t_i^2}}I_{(0,1)}\big(\sum_{i=1}^{n-1}t_i^2\big)I_{(0,1)}(v).$$

It is obvious that V and $(T_1,\dots,T_{n-1})'$ are independent. Moreover, it follows that V and $(T_1,\dots,T_{n-1})'$ are distributed as $B(n/2,a)$ and $MPII_{n-1}\big(\frac{1}{2}\big)$, respectively.

To prove (P2) note that

$$f_{W_1,\dots,W_n}(w_1,\dots,w_n) = \frac{1}{V_n(1)}I_{(0,1]}\big(\sum_{i=1}^{n}w_i^2\big).$$

Then the result follows by the fact that $(W_1,\dots,W_{n-1})' \sim MPII_{n-1}\big(\frac{3}{2}\big)$. ■

3.2. Multivariate Pearson type II transform of a normal random sample. Since multivariate standard normal distribution plays an important role in the class of multivariate symmetric spherical distributions, in this section we present special transformations of normal and complex normal random samples that follow multivariate Pearson type II distribution.

Let $X_1,\dots,X_{n+1}$ be a random sample from normal distribution with mean zero and variance σ^2. Now define the transforms R_{n+1} and $U_i, i=1,\dots,n+1$, as follows:

$$R_{n+1} = \sqrt{X_1^2+X_2^2+\ldots+X_{n+1}^2}, \tag{3.1}$$

$$U_i = \frac{X_i}{R_{n+1}}, \quad i=1,\dots,n+1. \tag{3.2}$$

The random vector $(U_1,\dots,U_{n+1})'$ is distributed uniformly over the surface of a unit $(n+1)$-sphere; see Rubinstein and Kroese [11] and Marsaglia [8]. Using the following theorem we show that all subvectors $(U_{i_1},\dots,U_{i_k})'$, $k=1,\dots,n$, of $(U_1,\dots,U_{n+1})'$ are distributed as multivariate Pearson type II distribution. According to Theorem 2.1 (C2), it suffices to show that $(U_1,\dots,U_n)'$ has multivariate Pearson type II distribution.

THEOREM 3.2. *For fixed n, assume R_{n+1} and $(U_1,\dots,U_n)'$ are defined as in* (3.1) *and* (3.2), *respectively.*

(i) $(U_{i_1},U_{i_2},\dots,U_{i_k})'$, $k=1,\dots,n$, *and R_{n+1} are independent.*

(ii) $(U_1,\dots,U_n)'$ *is distributed as $MPII_n\big(\frac{1}{2}\big)$ and*

$$\left(\frac{X_{i_1}}{\sqrt{\sum_{j=1}^{n+1}X_j^2}},\dots,\frac{X_{i_k}}{\sqrt{\sum_{j=1}^{n+1}X_j^2}}\right)' \sim MPII_k\left(\frac{1}{2}+\frac{n-k}{2}\right)$$

for $k=1,\dots,n-1$.

P r o o f. (i) By similar arguments to those in the proof of Theorem 2.1 (C2) we can obtain

$$\begin{aligned} f_{U_1,U_2,\dots,U_n,R_{n+1}}(u_1,\dots,u_n,r) &= f_{(X_1,\dots,X_{n+1})}(u_1,\dots,u_n,r)\cdot|J| \\ &= \frac{2\exp\{-r^2/2\sigma^2\}r^n}{(2\pi\sigma^2)^{(n+1)/2}\sqrt{1-\sum_{i=1}^n u_i^2}} I_{(0,\infty)}(r) I_{(0,1)}\big(\sum_{i=1}^n u_i^2\big). \end{aligned}$$

Hence R_{n+1} and $(U_1,\dots,U_n)'$ are independent.

(ii) It is easy to see that $(U_1,\dots,U_n)'$ is distributed as $MPII_n\big(\frac{1}{2}\big)$. We have

$$f_{U_1,\dots,U_n}(u_1,\dots,u_n) = \frac{\Gamma(n/2+1/2)}{\pi^{(n+1)/2}\sqrt{1-\sum_{i=1}^n u_i^2}} I_{(0,1)}\big(\sum_{i=1}^n u_i^2\big). \tag{3.3}$$

Comparing (3.3) to (1.4), we can conclude that $(U_1,\dots,U_n)' \sim MPII_n\big(\frac{1}{2}\big)$. ∎

REMARK 3.1. *Note that R_{n+1} is said to have a* generalized Rayleigh distribution. *In general, a random variable X follows a generalized Rayleigh distribution with parameters k and θ, $GR(k,\theta)$, if the density function of X is given by*

$$f(x) = \frac{2\theta^{k+1}}{\Gamma(k+1)} x^{2k+1}\exp\{-\theta x^2\}, \qquad x \geqslant 0,\ \theta > 0,\ k \geqslant 0$$

(see Johnson et al. [6]). *By this notation, R_{n+1} follows $GR\big((n-1)/2, 1/2\sigma^2\big)$. For $n=1$, R_{n+1} has an ordinary Rayleigh distribution.*

Moreover, suppose $X_1,\dots,X_n$ are i.i.d. complex normal random variables $N_c(0,\sigma^2)$. Let

$$\Re(\mathbf{X}) = \big(\Re(X_1),\dots,\Re(X_n)\big)'_{n\times 1} \quad \text{and} \quad \Im(\mathbf{X}) = \big(\Im(X_1),\dots,\Im(X_n)\big)'_{n\times 1}.$$

It is well known (see Goodman [4]) that

$$\big(\Re(\mathbf{X}),\Im(\mathbf{X})\big)' \sim N_{2n}(\mathbf{0},\mathbf{\Sigma}), \qquad \text{where } \mathbf{\Sigma} = \begin{bmatrix} \sigma^2 I & \mathbf{0} \\ \mathbf{0} & \sigma^2 I \end{bmatrix}.$$

Let

$$\mathbf{R} = \frac{\Re(\mathbf{X})}{\|\mathbf{X}\|} \quad \text{and} \quad \mathbf{I} = \frac{\Im(\mathbf{X})}{\|\mathbf{X}\|},$$

where $\|\mathbf{X}\| = \sqrt{\sum_{i=1}^n \big(\Re(X_i)\big)^2 + \sum_{i=1}^n \big(\Im(X_i)\big)^2}$. By Remark 2.2, one can conclude that $\mathbf{R} \overset{d}{=} \mathbf{I}$ and both are distributed as $MPII_n\big(\frac{1}{2}+n/2\big)$.

3.3. Simulation procedure. A common procedure to simulate random vectors from a prescribed multivariate distribution is generating their elements sequentially from their corresponding conditional distributions; see Ross [10]. Using this method and considering Remark 2.3, we can present the following procedure to simulate a random vector $(U_1, \ldots, U_n)$ with $MPII_n(a)$ distribution.

For fixed a and $n \geqslant 2$ take the following steps:

Step 1. Generate a random variable W from $B(n/2, a)$.

Step 2. Choose a random point $(U_1, \ldots, U_n)$ uniformly on the surface of an n-sphere with radius $\sqrt{w}$ generated in Step 1.

Several methods have been obtained to generate a random point uniformly on the surface of an n-sphere with radius r; see, e.g., Muller [9], Marsaglia [8], Guralnik et al. [5], and Rubinstein and Kroese [11]. According to these works, an efficient approach is to generate the vector

$$\Big(\frac{rX_1}{\sqrt{(X_1^2 + \ldots + X_n^2)}}, \ldots, \frac{rX_n}{\sqrt{(X_1^2 + \ldots + X_n^2)}}\Big),$$

where $X_1, \ldots, X_n$ is a random sample from $N(0, 1)$.

According to Theorem 2.1 (C4), we can use this procedure to generate random vectors from the marginal distributions of an arbitrary multivariate symmetric spherical distribution. Also, a conditioning technique can be utilized to generate a random vector from an arbitrary spherical distribution.

3.4. An extension. In this section, we introduce an extension to the class of multivariate Pearson type II distribution, namely hyperspherical beta distribution $HB_n(a, b)$. The joint density function of $HB_n(a, b)$ for $a, b > 0$ is given by

$$f_{U_1,\ldots,U_n}(u_1, \ldots, u_n) = A_{a,b,n}\big(\sum_{i=1}^{n} u_i^2\big)^{b-1}\big(1 - \sum_{i=1}^{n} u_i^2\big)^{a-1} I_{(0,1)}\big(\sum_{i=1}^{n} u_i^2\big). \tag{3.4}$$

It is obvious that for $b = 1$ the joint density function (3.4) reduces to joint density function of $MPII_n(a)$. Including one more parameter makes it being more flexible than the $MPII$ distribution. Also, the class of hyperspherical beta distributions can be studied in several statistical views as a generalization of beta distributions over a unit hypersphere. Let us summarize some features of this class through the following statements:

(i) The coefficient $A_{a,b,n}$ in (3.4) is given by

$$A_{a,b,n} = \frac{\pi^{-n/2}\Gamma(n/2)\Gamma(n/2 + a + b - 1)}{\Gamma(a)\Gamma(n/2 + b - 1)}.$$

(ii) The stochastic representation of the random vector

$$HB = (U_1, \ldots, U_n)' \sim HB_n(a, b)$$

can be given by $HB \overset{d}{=} \sqrt{V}W$, where V is distributed as $B(n/2 + b - 1, a)$.

(iii) For each $1 \leqslant k < n$, we have $(U_{i_1}, \ldots, U_{i_k})' \overset{d}{=} \sqrt{V}W^{(k)}$, where V and $W^{(k)}$ are distributed as $B(n/2 + b - 1, a)$ and $MPII_k(n - k)$, respectively. Moreover, they are independent.

4. DISCUSSION

There is a lot of information cited in many textbooks and articles about univariate distributions and their particular mathematical and statistical features. Nevertheless, for multivariate distributions, one can only find multivariate normal and some related distributions. Having a lot of mathematical and statistical features, the class of hyperspherical distributions, especially $MPII$, can be taken into account to cite in many multivariate textbooks as a class of multivariate distributions with several features. Also, there are wide-ranging multivariate data recorded in many applied fields that follow spherical distributions. The main aim of this article was to provide more details of this important class of multivariate distributions.

REFERENCES

[1] R. B. Arellano-Valle, *On some characterizations of spherical distributions*, Statist. Probab. Lett. 54 (2001), pp. 227–232.
[2] R. B. Arellano-Valle and H. Bolfarine, *On some characterizations of the t-distribution*, Statist. Probab. Lett. 25 (1995), pp. 79–85.
[3] K.-T. Fang, S. Kotz, and K.-W. Ng, *Symmetric Multivariate and Related Distributions*, Chapman and Hall, London 1990.
[4] N. R. Goodman, *Statistical analysis based on a certain multivariate complex Gaussian distribution (An introduction)*, Ann. Math. Statist. 34 (1) (1963), pp. 152–177.
[5] G. Guralnik, C. Zemach, and T. Warnock, *An algorithm for uniform random sampling of points in and on a hypersphere*, Inform. Process. Lett. 21 (1985), pp. 17–21.
[6] N. L. Johnson, S. Kotz, and N. Balakrishnan, *Continuous Univariate Distributions, Volume 2*, second edition, Wiley, New York 1994.
[7] J. J. Liang and P. M. Bentler, *Characterizations of some subclasses of spherical distributions*, Statist. Probab. Lett. 40 (1998), pp. 155–164.
[8] G. Marsaglia, *Choosing a point from the surface of sphere*, Ann. Math. Statist. 43 (2) (1972), pp. 645–646.
[9] M. E. Muller, *A note on a method for generating points uniformly on n-dimensional spheres*, Commn. ACM 2 (4) (1959), pp. 19–20.
[10] S. M. Ross, *Simulation, Fourth Edition (Statistical Modeling and Decision Science)*, Elsevier Inc., 2006.
[11] R. Y. Rubinstein and D. P. Kroese, *Simulation and Monte Carlo Methods*, second edition, Wiley, 2008.
[12] N. A. Volodin, *Spherically symmetric logistic distribution*, J. Multivariate Anal. 70 (1999), pp. 202–206.

Mahnaz Khalafi (corresponding author)
Department of Statistics
Faculty of Science
Golestan University
P.O. Box 155, Gorgan, Iran
E-mail: m.khalafi@gu.ac.ir

Majid Azimmohseni
Department of Statistics
Faculty of Science
Golestan University
P.O. Box 155, Gorgan, Iran
E-mail: m.azim@gu.ac.ir

Received on 11.2.2013;
revised version on 29.11.2013