Some properties of the Pearson type II (power semicircle) distribution

Евелина Велева

Някои свойства на разпределението на Пирсън от тип II: *Разпределението на Пирсън от тип II може да се разглежда като частен случай на разпределението на Пирсън от тип I, известно още и като Бета разпределение с 4 параметъра. В литературата се среща и под името степенно полукръгово разпределение. До интерестни рекурентни свойства на това разпределение се достига при анализа на ковариационни и корелационни матрици на многомерното нормално разпределение.*

Ключови думи: *разпределение на Пирсън от тип I и II, Бета разпределение, степенно полукръгово разпределение*

INTRODUCTION

The system of Pearson distributions was developed by Karl Pearson in 1894 and 1895 to provide flexible descriptions of the non-normal distributions encountered in his biometric research. The original papers are reproduced in [6]. Apart from the fitting of models for observed frequency distributions, the Pearson distributions have also been used to provide approximations to other theoretical distributions, and to study the effect of non-normality on sampling distributions (see [3]).

The Pearson type II distribution has probability density function of the form ([5]):

$$f(x) = C_2(a^2 - x^2)^b, \qquad x \in (-a, a), \tag{1}$$

where $a > 0$ and $b > -1$ are parameters; C_2 is a normalizing constant. It is a particular case of the Pearson type I distribution, which depends on 4 parameters a, b, c, d, $b, d > -1$, $a \geq 0$, $c > 0$ with probability density function of the form

$$f(x) = C_1(a + x)^b (c - x)^d, \qquad x \in (-a, c), \tag{2}$$

where C_1 is a normalizing constant. The Pearson type I distribution is also known as four parameters Beta distribution, because when $a = 0$ and $c = 1$, (2) gives the probability density function of the classical Beta distribution.

The Pearson type II distribution occurs also under the name power semicircle distribution ([2]). When $b = 1/2$, (1) gives the density of the so called semicircle distribution. The other values of b can be presented in the form $b = \theta + 1/2$, thus we obtain from (1) the density of the power semicircle distribution $PS(\theta, a)$

$$f_\theta(x; a) = \frac{\Gamma(\theta + 2)}{\sqrt{\pi} a^2 \Gamma(\theta + 3/2)} (a^2 - x^2)^{\theta + 1/2}, \qquad x \in (-a, a), \tag{3}$$

$a > 0$, $\theta > -3/2$.

The Pearson type II distribution or the power semicircle distribution is the distribution of the sample correlation coefficient for a sample of observations on two independent random variables with a bivariate normal distribution. The distribution $\psi_n(m)$, $m \geq n$ of the sample correlation matrix for a sample of $m + 1$ observations on n independent random variables with multivariate normal distribution has probability density function of the form (see [9]):

$$f(\mathrm{U}) = \frac{\left(\Gamma\left(m/2\right)\right)^n}{\Gamma_n\left(m/2\right)}\left(\det \mathrm{U}\right)^{(m-n-1)/2}, \tag{4}$$

for every positive definite matrix $\mathrm{U} = (u_{i,j})$ with units on the main diagonal, where $\Gamma_n(\cdot)$ is the multivariate gamma function defined as $\Gamma_n\left(\gamma\right) = \pi^{n(n-1)/4}\prod_{j=1}^{n}\Gamma[\gamma + (1-j)/2]$. When $n = 2$ from (4) we obtain the density of the sample correlation coefficient for the 2 observed variables in the form

$$f(u) = \frac{\Gamma\left(m/2\right)}{\sqrt{\pi}\Gamma\left((m-1)/2\right)}\left(1-u^2\right)^{(m-3)/2}, \qquad u \in (-1,1). \tag{5}$$

According to (3), this is the density of the distribution $PS\left((m-4)/2,1\right)$.

In this paper we consider properties of the power semicircle distribution resulting from its special role in the analysis of sample correlation and covariance matrices.

PROPERTIES

Let $P(n,\mathbb{R})$ be the set of all real symmetric positive definite matrices of order n. Let us denote by $D(n,\mathbb{R})$ the set of all real symmetric matrices of order n, with positive diagonal elements, which off-diagonal elements are in the interval (-1,1). There exist a bijection $h\colon D(n,\mathbb{R}) \to P(n,\mathbb{R})$, considered in [7] - [11]. The image of an arbitrary matrix $X = (x_{i,j})$ from $D(n,\mathbb{R})$ by the bijection h, is a matrix $Y = (y_{i,j})$ from $P(n,\mathbb{R})$, such that

$$y_{j,j} = x_{j,j},\ j = 1,\ldots,n, \tag{6}$$

$$y_{1,j} = x_{1,j}\sqrt{x_{1,1}x_{j,j}},\ j = 2,\ldots,n, \tag{7}$$

$$y_{i,j} = \sqrt{x_{i,i}x_{j,j}}\left[\sum_{r=1}^{i-1}\left(x_{r,i}x_{r,j}\prod_{q=1}^{r-1}\sqrt{(1-x_{q,i}^2)(1-x_{q,j}^2)}\right) + x_{i,j}\prod_{q=1}^{i-1}\sqrt{(1-x_{q,i}^2)(1-x_{q,j}^2)}\right], \tag{8}$$
$$2 \le i < j \le n.$$

The next Proposition is proved in [8].

Proposition 1. *Let $\boldsymbol{\xi} = (\xi_{i,j})$ be a random symmetric matrix of order n with units on the main diagonal. Suppose that $\xi_{i,j}$, $1 \le i < j \le n$ are independent and $\xi_{i,j} \sim PS\left((m-i-3)/2,1\right)$, where m is an integer, $m \ge n$. Let $\mathbf{V}$ be the matrix $\mathbf{V} = h(\boldsymbol{\xi})$, where h is the bijection, defined by (6) – (8). Then the matrix $\mathbf{V}$ has distribution $\psi_n(m)$.*

Theorem 1 below gives an interesting property of the power semicircle distribution, which follows from the representation (8) for an arbitrary element of a positive definite matrix.

Theorem 1. *Let ζ_i, η_i, $i = 1,\ldots,k$ be independent random variables, ζ_i, $\eta_i \sim PS\left((m-i-3)/2,1\right)$, where m is an integer, $m > k > 1$. Then the random variable*

$$\zeta_1\eta_1 + \zeta_2\eta_2\sqrt{(1-\zeta_1^2)(1-\eta_1^2)} + \zeta_3\eta_3\sqrt{(1-\zeta_1^2)(1-\eta_1^2)(1-\zeta_2^2)(1-\eta_2^2)}$$

$$+\dots+\zeta_{k-1}\eta_{k-1}\sqrt{(1-\zeta_1^2)(1-\eta_1^2)\dots(1-\zeta_{k-2}^2)(1-\eta_{k-2}^2)}$$
$$+\eta_k\sqrt{(1-\zeta_1^2)(1-\eta_1^2)\dots(1-\zeta_{k-1}^2)(1-\eta_{k-1}^2)}$$

is distributed $PS\big((m-4)/2,1\big)$.

Proof. Let us have $m+1$ observations on $k+1$ independent random variables with multivariate normal distribution. The distribution of the sample correlation matrix $\mathbf{V}=(V_{i,j})$ is $\psi_{k+1}(m)$ ([9]). Each element $V_{i,j}$ of $\mathbf{V}$ is the sample correlation coefficient, calculated on the basis of $m+1$ observations on the corresponding pair of independent factors. Hence $V_{i,j}$, $1\le i<j\le n$ are identically distributed, $V_{i,j}\sim PS\big((m-4)/2,1\big)$, with density function given by (5). From Proposition 1 we have that $\mathbf{V}=h(\boldsymbol{\xi})$, where $\boldsymbol{\xi}=(\xi_{i,j})$ is a random matrix with units on the main diagonal. The off-diagonal elements of $\boldsymbol{\xi}$ are independent and $\xi_{i,j}\sim$ $PS\big((m-i-3)/2,1\big)$. Then the element $V_{k,k+1}$ of $\mathbf{V}$, according to (8), can be presented as

$$V_{k,k+1}=\sum_{r=1}^{k-1}\left(\xi_{r,k}\xi_{r,k+1}\prod_{q=1}^{r-1}\sqrt{(1-\xi_{q,k}^2)(1-\xi_{q,k+1}^2)}\right)+\xi_{k,k+1}\prod_{q=1}^{k-1}\sqrt{(1-\xi_{q,k}^2)(1-\xi_{q,k+1}^2)}\,.$$

Now, if we denote $\zeta_i=\xi_{i,k}$, $i=1,\dots,k-1$, $\eta_i=\xi_{i,k+1}$, $i=1,\dots,k$, the Theorem follows. □

The sample covariance matrix for a sample from a multivariate normal distribution has Wishart distribution (see [1]). A $n\times n$ random matrix $\mathbf{W}$ with Wishart distribution $W_n(m,\Sigma)$, where $n\le m$ and Σ is a positive definite $n\times n$ matrix, has probability density function of the form

$$f_{n,m,\Sigma}(\mathrm{W})=\frac{1}{2^{nm/2}\Gamma_n\big(m/2\big)(\det\Sigma)^{m/2}}(\det\mathrm{W})^{(m-n-1)/2}e^{-tr(\mathrm{W}\Sigma^{-1})/2}\,, \qquad (9)$$

for any real $n\times n$ positive definite matrix W, where $tr(\cdot)$ denotes the trace of a matrix.

Wishart distribution can be also produced from the bijection h (see [8], [9]). Here the power semicircle distribution also plays a key role. In the proposition below we denote by $\chi^2(m)$ the chi-square distribution with density function

$$f(x)=\frac{1}{2^{m/2}\Gamma\big(m/2\big)}x^{m/2-1}e^{-x/2}\,, \qquad x>0\,.$$

Proposition 2. Let $\boldsymbol{\xi}=(\xi_{i,j})$ be a random symmetric matrix of order n. Suppose that $\xi_{i,j}$, $1\le i\le j\le n$ are independent, $\xi_{i,j}\sim PS\big((m-i-3)/2,1\big)$ for $1\le i<j\le n$ and $\xi_{i,i}\sigma_i^{-2}\sim\chi^2(m)$, $i=1,\dots,n$. The parameters σ_1^2, …, σ_n^2 are arbitrary positive numbers. Let $\mathbf{W}$ be the matrix $\mathbf{W}=h(\boldsymbol{\xi})$, where h is the bijection, defined by (6) – (8). Then the matrix $\mathbf{W}$ has distribution $W_n(m,\Sigma)$, $\Sigma=diag(\sigma_1^2,\dots,\sigma_n^2)$.

This representation of the Wishart distribution leads to another property of the power semicircle distribution.

Theorem 2. *Let* ζ_i, $i=1,\ldots,k$ *be independent random variables,* $\zeta_i \sim PS\left((m-i-3)/2,1\right)$, $i=1,\ldots,k-1$, $\zeta_k \sim \chi^2(m)$, *where* m *is an integer,* $m>k>1$. *Let* ν_i, $i=1,\ldots,k$ *be the random variables*

$$\nu_1=\zeta_1\sqrt{\zeta_k}\,,\ \nu_i=\zeta_i\sqrt{\zeta_k}\sqrt{(1-\zeta_1^2)\ldots(1-\zeta_{i-1}^2)}\,,\ i=2,\ldots,k-1, \tag{10}$$

$$\nu_k=\sqrt{\zeta_k}\sqrt{(1-\zeta_1^2)\ldots(1-\zeta_{k-1}^2)}\,. \tag{11}$$

Then the random variables ν_i**,** $i=1,\ldots,k$ **are independent and** $\nu_i \sim N(0,1)$**,** $i=1,\ldots,k-1$**,** $\nu_k \sim \chi^2(m-k+1)$**.**

To proof Theorem 2 we need the next Lemma, which can be easily checked, using the equalities (6) – (8), defining the bijection h.

Lemma 1. *Let* $\mathrm{X}=(x_{i,j})$ *be a matrix from* $D(n,\mathbb{R})$ *and* $\mathrm{Y}=h(\mathrm{X})$ *be the corresponding positive definite matrix from* $P(n,\mathbb{R})$. *Let* U *be the lower triangular matrix*

$$\mathrm{U}=\begin{pmatrix} \sqrt{x_{1,1}}s_{1,1} & 0 & \cdots & 0 \\ \sqrt{x_{2,2}}s_{2,1} & \sqrt{x_{2,2}}s_{2,2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \sqrt{x_{n,n}}s_{n,1} & \sqrt{x_{n,n}}s_{n,2} & \cdots & \sqrt{x_{n,n}}s_{n,n} \end{pmatrix}, \tag{12}$$

where

$$s_{1,1}=1,\ s_{j,1}=x_{1,j},\ j=2,\ldots,n \tag{13}$$

$$s_{j,i}=x_{i,j}\sqrt{(1-x_{1,j}^2)\ldots(1-x_{i-1,j}^2)},\ 2\le i<j\le n \tag{14}$$

$$s_{j,j}=\sqrt{(1-x_{1,j}^2)\ldots(1-x_{j-1,j}^2)},\ j=2,\ldots,n. \tag{15}$$

Then $\mathrm{Y}=\mathrm{U}\mathrm{U}^t$.

Proof of Theorem 2. Let $\mathbf{W}$ be a random symmetric matrix with Wishart distribution $W_n(m,\mathrm{I}_n)$, where I_n is the identity matrix of size n. From Proposition 2 we have that $\mathbf{W}=h(\boldsymbol{\xi})$, where $\boldsymbol{\xi}=(\xi_{i,j})$ is a random matrix with independently distributed elements, $\xi_{i,j} \sim PS\left((m-i-3)/2,1\right)$, $1\le i<j\le n$ and $\xi_{i,i} \sim \chi^2(m)$, $i=1,\ldots,n$. The random variable ζ_i is distributed as $\xi_{i,k}$ for $i=1,\ldots,k$.

Let $\mathbf{U}=(U_{i,j})$ be the lower triangular matrix, constructed from the matrix $\boldsymbol{\xi}=(\xi_{i,j})$, according to Lemma 1. From (12) – (15) it can be seen that $U_{k,1}=\xi_{1,k}\sqrt{\xi_{k,k}}$, $U_{k,i}=\xi_{i,k}\sqrt{\xi_{k,k}}\sqrt{(1-\xi_{1,k}^2)\ldots(1-\xi_{i-1,k}^2)}$, $i=2,\ldots,k-1$, $U_{k,k}=\sqrt{\xi_{k,k}}\sqrt{(1-\xi_{1,k}^2)\ldots(1-\xi_{k-1,k}^2)}$. Consequently, the variables ν_i, $i=1,\ldots,k$ are distributed as the elements $U_{k,1}$, …, $U_{k,k}$ of $\mathbf{U}$. Bartlett ([4]), using a different approach, proves that the elements $U_{k,1}$, …, $U_{k,k}$ of a lower triangular matrix $\mathbf{U}$, such that $\mathbf{W}=\mathbf{U}\mathbf{U}^t$, are independent, $U_{k,i}\sim N(0,1)$, $i=1,\dots,k-1$ and $U_{k,k}\sim\chi^2(m-k+1)$. □

CONCLUSIONS

The Pearson type II ((power semicircle) distribution plays a key role in generation by the bijection h of positive definite random matrices with different distributions (see [7] – [11]). More properties, similar to Theorem 1 and 2 can be derived from the representations of the corresponding distributions of positive definite random matrices.

It is interesting whether similar properties hold for the Pearson type I distribution or in particular for the classical Beta distribution.

ЛИТЕРАТУРА

[1] T. W. Anderson. An Introduction to Multivariate Statistical Analysis. John Wiley & Sons, N. Y., 3ed., 2003.
[2] O. Arizmendi, V. P´erez-Abreu. On the non-classical infinite divisibility of power semicircle distributions. Commun. Stoch. Anal. 4 (2010), 161–178, in honor of Gopinath Kallianpur.
[3] P. Armitage. Pearson Distributions. Encyclopedia of Biostatistics, 2005.
[4] M.S. Bartlett. On the theory of statistical regression. Proc. Roy. Soc. Edinb., 53 (1933), 260 – 283.
[5] L. Devroye. Non-Uniform Random Variate Generation. New York, Springer-Verlag, 1986.
[6] K. Pearson. Early Statistical Papers. Cambridge University Press, Cambridge, 1948.
[7] E. Veleva. Uniform random positive definite matrix generation. Math. and Education in Math. 35 (2006), 321 – 326.
[8] E. Veleva. A representation of the Wishart distribution by functions of independent random variables. Annuaire Univ. Sofia, Fac. of Econ. & Buss. Admin., 6 (2007), 59 – 68.
[9] E. Veleva. Some New Properties of Wishart Distribution. Applied Mathematical Sciences, Vol. 2, 2008, no. 54, 2673 - 2682.
[10] E. Veleva. Uniformly Distributed Positive Definite Matrices With Bounded Trace. Journal of the Technical University at Plovdiv “Fundamental Sciences and Applications”, Vol. 14, 2009, 455-460.
[11] E. Veleva. Stochastic representations of the Bellman gamma distribution. Proc. of Int. Conf. on Theory and Applications of Mathematics and Informatics, ICTAMI 2009, Alba Iulia, Romania, 463-474.

За контакти:

гл.ас.д-р Евелина Велева, Катедра “Приложна математика и статистика”, Русенски университет “Ангел Кънчев”, e-mail: eveleva@uni-ruse.bg

Докладът е рецензиран.