*STATISTICA, anno LXXXI, n. 2, 2021*

# A CLASS OF UNIVARIATE NON-MESOKURTIC DISTRIBUTIONS USING A CONTINUOUS UNIFORM SYMMETRIZER AND CHI GENERATOR

Kamala Naganathan Radhalakshmi[1]
*Department of Statistics, Loyola College, Chennai-600034, India*
Martin Luther William
*Department of Statistics, Loyola College, Chennai-600034, India*

## 1. Introduction

The Normal distribution is important in statistics and is widely used in natural and social sciences to represent the distribution of real-valued random variables and as an approximation to several other distributions. The normal density curve is a symmetric, bell-shaped one with a single peak. Its peak corresponds to the mean, median and mode of the distribution. The normal distribution has come to dominate statistical analysis of real-time data due to empirical evidence for its relevance and theoretical evidence by the Central Limit Theorem. Much of the analytical work in Statistics is centered on this distribution primarily owing to its nice mathematical properties and availability of tables / statistical packages incorporating normal critical values. Even though normal distribution approximates several other distributions well, we encounter situations where the assumption of normality is not satisfied by the original variable nor the transformed one. So, rather than wrongly assuming normality and drawing incorrect conclusions, fitting a non-normal distribution reflecting the characteristics of the variable being studied would lead to appropriate analysis of the data.

### 1.1. *Kurtosis classification*

There are three classes of kurtosis that a distribution can be classified into. The three divisions are made with the standard normal distribution as a benchmark.
MESOKURTIC: A distribution is said to be mesokurtic if its kurtosis value is 3 and the shape of the distribution is close to that of the normal distribution's density curve.

[1] Corresponding Author. E-mail: 20dst001@loyolacollege.edu

LEPTOKURTIC: Leptokurtic distributions are those with kurtosis value greater than 3. Leptokurtic distributions have their density curves with shorter tails.
PLATYKURTIC: The distributions which have a kurtosis value less than 3 fall in this category. The density curve of platykurtic distributions have longer tails.

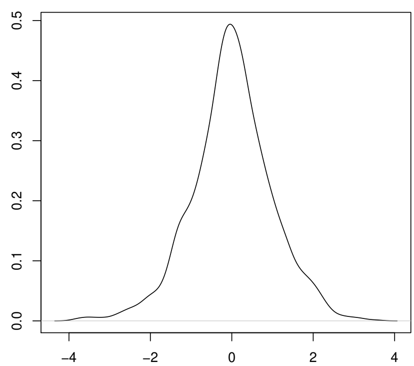


*Figure 1* – A leptokurtic distribution with kurtosis 5.14.

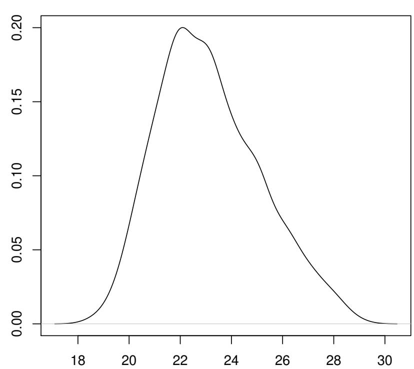


*Figure 2* – A platykurtic distribution with kurtosis 2.7.

Although kurtosis is an important measure of a distribution, it is not given due importance in the literature. This paper aims at dealing with distributions that have the moment-based kurtosis value different from 3; that is, distributions which are non-mesokurtic. In saying non-mesokurtic, we emphasize that both platykurtic and leptokurtic distributions are being dealt with.

In many practical applications involving tests of hypotheses, the analysis of the data is carried out assuming normality. In linear regression, many analytical procedures are based on the assumption of normality of the residuals from the model. However, in several situations, this assumption fails and any follow-up analysis including tests of significance, outlier detection with the normal critical point thresholds, etc. are not justified. The residuals may be symmetrically spread but non-mesokurtic and hence, follow-up analysis and other inferential procedures need to be based on a distribution which is normal-like but non-mesokurtic. These requirements necessitate the development of such distributions.

This paper is an attempt to generate a class of univariate non-mesokurtic symmetric distributions containing both leptokurtic and platykurtic distributions which can be employed in situations where the distribution is symmetric but normal distribution assumption fails. Reference is made to Bagnato *et al.* (2017) for an interesting approach to generate a class of multivariate leptokurtic normal distributions. The approach in the above-mentioned paper is interesting and the generated class of distributions is for a multivariate context. Currently we seek a different route to generate a class of univariate distributions that are symmetric and non-mesokurtic in nature.

This paper is organized as follows: After this introductory section, in Section 2 we give a brief review of the literature on leptokurtic distributions. In Section 3, we present an alternate approach to generate a class of univariate symmetric non-mesokurtic distributions with a different symmetrizer, namely a continuous uniform random variable. The class of distributions so generated contains both leptokurtic and platykurtic distributions and the mesokurtic (normal) distribution as well. In Section 4, we derive the density functions of the distributions in the generated class. In Section 5 we bring out some interesting properties of the distributions in the class. In Section 6, we derive the estimators of the parameters using the method of moments. In section 7, we give the concluding remarks.

## 2. A REVIEW OF CURRENT LITERATURE ON LEPTOKURTIC DISTRIBUTIONS

The present study has been inspired by two research papers: (1) "On the Theory of Elliptically Contoured Distributions" by Cambanis *et al.* (1981); (2) "The multivariate leptokurtic-normal distribution and its application in model-based clustering" by Bagnato *et al.* (2017).

The former paper discusses the important theoretical aspects of elliptically contoured distributions, of which multivariate normal distributions constitute a sub-class. Statistical inference dealing with continuous multivariate data is commonly focused on elliptical distributions but, among them, the normal distribution is the most widely used one because of computational and theoretical convenience. Even though there is a rich theory for elliptically contoured distributions, this class of distributions has not been generally applied in practical situations owing to lack of awareness and absence of computing software that support the computations related to this class of distributions.

The latter paper draws inspiration from the former and deals specifically with leptokurtic multivariate normal distribution and its application in model based clustering. In several practical situations, it is found that the kurtosis of the observed distribution is different from that of the normal distribution even though thecolorred distribution is symmetric. For instance, the movement in stock indices may be much concentrated about a central value with a high-peaked and thin-tailed distribution which may appear like a normal distribution but not exactly normal as found in illustrations in Figure 1 and Figure 2. Reference is made to Szegö (2004) for such instances in finance data. Unlike the concepts of location, spread and skewness, the meaning of kurtosis is a topic of considerable debate. For interesting expositions on kurtosis, we refer to Balanda and MacGillivray (1988, 1990) and Wang and Zhou (2012) besides others.

According to Arevalillo and Navarro (2012) the statistical concept behind kurtosis is concerned with the curvature, the amount of peakedness, and the tail weight of a distribution.The classical notion of univariate kurtosis is moment-based and given by the standardized fourth central moment. A natural multivariate extension for a random vector X with mean vector $\mu$ and covariance matrix $\Sigma$ is

$$\text{Kurt}(X) = E[(X-\mu)'\Sigma^{-1}(X-\mu)]^2.$$

We refer to Mardia (1970) for details.

Bagnato *et al.* (2017) have proposed a class of multivariate leptokurtic-normal (MLN) distributions to handle the high peakedness or high value of kurtosis of the data. Their formulation is a multivariate Gram-Charlier expansion of the Multivariate Normal (MN) distribution. The MLN distribution is obtained by reshaping the generating variable of its elliptical representation. We refer to Cambanis *et al.* (1981) for details. The result is a distribution characterized by one additional parameter corresponding to the excess kurtosis vis-a-vis the original MN distribution. According to the authors of that article, the distributions thus proposed by them appear suitable in fitting several real data sets showing different levels of excess kurtosis.

Here, we give a brief review of the literature on generating elliptical leptokurtic distributions.

A d-variate continuous random vector X, with mean vector $\mu$ , has an elliptical distribution if and only if it can be written as

$$X = \mu + R\Lambda U, \tag{1}$$

where $U$ is a $d$-variate random vector uniformly distributed on the unit hypersphere with $d=1$ dimensions $\{u \in R^d :\| u \|= 1\}$, $R$ is a non-negative random variable - called a generating variate - stochastically independent of $U$ and $\Lambda$ is a $d \times d$ matrix satisfying the condition

$$\Lambda\Lambda' = \frac{d}{E(R^2)} VC(X), \tag{2}$$

where $VC(X)$ denotes the variance-covariance matrix of $X$. With this condition, Bagnato *et al.* (2017) developed a family of multivariate leptokurtic normal distributions. In this paper we seek to take an alternate approach towards developing a family of non-mesokurtic distributions for dimension $d = 1$. For this case, the Equation (2) reduces to

$$\Lambda^2 = \frac{V(X)}{E(R^2)}. \tag{3}$$

Noting that the set $\{u \in R^d :\| u \|= 1\}$, reduces to a two-point set $\{˘1, +1\}$ and, hence that $E(U^2) = 1$, the Equation (3) is mathematically consistent with Equation (1) for $d = 1$. However, for the approach to be proposed in the next section, a modification to the condition laid out in Equation (3) is given and the development of the univariate non-mesokurtic family is presented.

## 3. Generating a univariate non-mesokurtic normal family

In the existing literature on the development of multivariate leptokurtic normal family,a discrete uniform symmetrizer is required in the case of dimension $d = 1$.The objective of the present paper is to provide an alternate development of a univariate family that contains leptokurtic as well as platykurtic distributions. The approach proposed here uses Equation (1) but, with a difference in the distributional aspects of the variables used to define the random variable X.

Consider $R$, a random variable which follows Chi-distribution; that is, $R^2$ follows a Chi-square distribution with $m$ degrees of freedom and let $U$ be a continuous random variable uniformly distributed in the interval (–1, +1) independent of $R$. That is, we use a Chi generator as suggested by Cambanis *et al.* (1981) and Bagnato *et al.* (2017), but the symmetrizer is a continuous uniform variable. The proposed family of non-mesokurtic distributions is constructed with the following ‘generating equations:

$$X = \mu + R\Lambda U,$$

where

$$\Lambda^2 = \frac{Var(X)}{E(R^2)E(U^2)}. \tag{4}$$

It may be noted that the above formulation for $\Lambda^2$, works well not only for our choice of the symmetrizer $U$, but also for the discrete uniform symmetrizer of Bagnato *et al.* (2017).

Now, $\beta_2(X)$, the moment based kurtosis of $X$ is

$$\beta_2(X) = \frac{E(X-\mu)^4}{[V(X)]^2} = \frac{E(R^4)E(U^4)}{[E(R^2)]^2[E(U^2)]^2}.$$

We have $R^2 \sim X^2(m)$ and therefore

$$E(R^2) = m \text{ and } E(R^4) = 2m + m^2.$$

Also,

$$E(U^2) = \frac{1}{3} \text{ and } E(U^4) = \frac{1}{5}.$$

Hence

$$\beta_2(X) = \frac{9}{5}\left(\frac{m+2}{m}\right). \tag{5}$$

Special Cases:

When $m = 1$, $\beta_2(X) = 5.4$;

When $m = 2$, $\beta_2(X) = 3.6$;

When $m = 3$, $\beta_2(X) = 3$;

When $m = 4$, $\beta_2(X) = 2.7$.

From the above calculations, we observe that for $m = 1, 2$, the distribution of $X$ is leptokurtic. For $m = 3$, we get mesokurtic nature. Higher values of $m$ above 3 (degrees of freedom of Chi-squared variable), gives kurtosis less than 3, i.e., a platykurtic distribution. Thus, the proposed approach, gives a way to generate a wide class of leptokurtic, mesokurtic and platykurtic distributions. It is also interesting that the distributions in the class are symmetric as evidenced by the zero value for the third central moments.

It is easy to note that as $m$ increases, the value of $\beta_2(X)$ approaches 1.8. The class that we generate in this approach encompasses a wide class of distributions with the moment-based kurtosis values ranging from 1.8 to 5.4. While Bagnato *et al.* (2017) have taken up the generation of leptokurtic distributions for multivariate set-up, we have considered only the univariate case but encompassed leptokurtic and platykurtic distributions in the class being generated. Furthermore, we have used a continuous uniform symmetrizer along with a chi generator for generating the class.

## 4. Derivation of Density Functions

In this Section, we shall derive the density functions of the distributions in the class. Basically, we need the density functions of $X - \mu = R\Lambda U = W$, say, where $\Lambda$ is a constant depending on the distribution of $R$ and distribution of $U$ and takes care of the variance of $X$. As specified earlier, $R^2$ is a chi square random variable with $m$ degrees of freedom. We denote $R^2$ as $S$ in the sequel.

The p.d.f of $R$ is

$$f_R(r) = \frac{1}{\Gamma(m/2)2^{\frac{m}{2}}} e^{-\frac{r^2}{2}} r^{m-1}, r > 0.$$

We shall first find the conditional density of $W$ given $U$ $f_W(W|U)$. Consider,

$$W = R\Lambda U,$$

$$R = \frac{W}{\Lambda U},$$

$$|\frac{dR}{dW}| = \frac{1}{\Lambda |U|},$$

$$f_W(W|U) = \frac{2|w|^{(m-1)} \exp(-\frac{w^2}{2\Lambda^2 u^2})}{\Gamma \frac{m}{2} u^m 2^{\frac{m}{2}}}, w > 0, u > 0 \,(or)\, w < 0, u < 0.$$

The joint probability density function of $U$ and $W$ is

$$f(u,w) = f_w(w|u).f(u) = \frac{|w|^{(m-1)} \exp(-\frac{w^2}{2\Lambda^2 u^2})}{2^{\frac{m}{2}} \Gamma \frac{m}{2} \Lambda^m |u|^m}, w > 0, u > 0 \,(or)\, w < 0, u < 0.$$

Hence, the marginal density of $W$ is

$$f_W(w) = \frac{|w|^{m-1}}{\Lambda^m \Gamma \frac{m}{2} 2^{\frac{m}{2}}} \int_0^1 \frac{\exp(-\frac{w^2}{2\Lambda^2 u^2})}{u^m} du, \; -\infty < w < \infty. \tag{6}$$

## 5. Properties of the proposed class of distributions

In this Section, we provide some interesting results on the class of distributions generated. The first result shows that when we choose $m = 3$, the distribution is normal. The second result gives an understanding of the shape of the distributions in the class.

THEOREM 1. *When the Chi generator has degrees of freedom $m = 3$, the distribution generated is normal*

PROOF. As derived above, the density function of $W$ is given as

$$f_W(w) = \frac{|w|^{m-1}}{\Lambda^m \Gamma \frac{m}{2} 2^{\frac{m}{2}}} \int_0^1 \frac{\exp(-\frac{w^2}{2\Lambda^2 u^2})}{u^m} du, \; -\infty < w < \infty.$$

Taking $m = 3$, we get

$$f_W(w) = \int_0^1 \frac{w^2 \exp(-\frac{w^2}{2\Lambda^2 u^2})}{\Lambda^3 \Gamma \frac{3}{2} 2^{\frac{3}{2}} u^3} du, \; -\infty < w < \infty.$$

Taking,

$$\frac{1}{u^2} = v \text{ and } \frac{w^2}{2\Lambda^2} = l$$

and carrying out the integration, we get

$$f_W(w) = \frac{1}{\Lambda\sqrt{2\pi}} e^{-\frac{w^2}{2\Lambda^2}}, \; -\infty < w < \infty,$$

which is the density function of a normal random variable with mean 0 and variance $\Lambda^2$. $\square$

THEOREM 2. *The density functions of the generated distributions has two points of inflexion equidistant on either side of the mean when $m > 2$.*

PROOF. For w> 0, the density function of the generated distributions can be written as

$$f_W(w) = c\int_0^1 \frac{w^{m-1}}{u^m} e^{-\frac{w^2}{2\Lambda^2 u^2}},$$

where

$$c = \frac{1}{\Lambda^m \Gamma\frac{m}{2} 2^{\frac{m}{2}}}.$$

Now taking

$$\frac{w^2}{\Lambda^2 u^2} = z,$$

we get

$$f_W(w) = \frac{c\Lambda^{m-1}}{2}\int_{\frac{w^2}{\Lambda^2}}^{\infty} z^{\frac{m-3}{2}}\, dz,$$

$$f'(w) = c e^{\frac{w^2}{2\Lambda^2}} w^{m-2}$$

$$f''(w) = -c \; \exp\left\{\frac{w^2}{2\Lambda^2}\left[(m-2) - \frac{w^2}{\Lambda^2}\right]\right\},$$

$$\begin{cases} < 0, \; if \; w^2 \; < \; \Lambda^2(m-2) \\ = 0, if \; w^2 \; = \; \Lambda^2(m-2) \\ > 0, if \; w^2 \; > \; \Lambda^2(m-2). \end{cases}$$

From the above equations, we observe the following about the shape of the density curve: When $m \geq 2$, the curve is convex throughout $w > 0$. The same convex shape exists on the negative side of the real line also. And, when $m > 2$, there is a point of inflexion at $w = \Lambda\sqrt{(m-2)}$ on the positive side. By symmetricity, the point $w = -\Lambda\sqrt{(m-2)}$ is also a point of inflexion. Between these points, the curve is concave and outside this it is convex. It is clearly seen that when $m = 3$, the points of inflexion are $\pm\Lambda$ where $\Lambda$ is the standard deviation of normal $N(0, \Lambda^2)$. $\square$

## 6. ESTIMATING THE PARAMETERS $\mu$, $m$ and $\Lambda^2$

From Section 2, we know that $X-\mu = R\Lambda U = W$
From the density function of the random variable W, we find E(W) = 0 and

$$E(W^2)=\frac{\Lambda^2 m}{3}, \tag{7}$$

$$E(W^4)=\frac{\Lambda^4 m(m+2)}{5}. \tag{8}$$

From (7), we get

$$\Lambda^2=\frac{3E(W^2)}{m}=\frac{3V(X)}{m}. \tag{9}$$

Substituting the value of $\Lambda^2$ in (8), we get

$$\frac{5}{9}\beta_2(W)=\frac{m+2}{m}.$$

Solving further, we get

$$m=\frac{18}{5\beta_2(W)-9}=\frac{18V(X)^2}{5E(X-\mu)^4-9V(X)^2}. \tag{10}$$

Denoting a random sample from the distribution as $X_1, X_2,...,X_n$, the moment estimator of the mean $\mu$ is the usual sample first moment given by

$$\hat{\mu}=\bar{X}. \tag{11}$$

The moment estimators of $m$ and $\Lambda^2$ are obtained from Equations (9) and (10) as

$$\hat{m}=\frac{18[\Sigma_{i=1}^n(X_i-\bar{X})^2]^2}{5n\Sigma_{i=1}^n(X_i-\bar{X})^4-9[\Sigma_{i=1}^n(X_i-\bar{X})^2]^2}, \tag{12}$$

$$\hat{\Lambda}^2=\frac{5n\Sigma_{i=1}^n(X_i-\bar{X})^4-9[\Sigma_{i=1}^n(X_i-\bar{X})^2]^2}{6\Sigma_{i=1}^n(X_i-\bar{X})^2}. \tag{13}$$

Using the above equations (11), (12) and (13), one can get the estimates of the parameters.

## 7. CONCLUDING REMARKS

In many real life situations, we come across datasets which are not exactly normal but normal like. In such situations, assuming normality and proceeding leads to wrong conclusions. This has been the motivation of the theoretical development in the present paper. By following an approach different from existing approaches, we have developed a family of univariate non-mesokurtic distributions. The moment estimators of the parameters have been derived. The prospective applications of the family of distributions so developed include outlier detection in real time data analysis. The extension to a family of multivariate non-mesokurtic distributions is being addressed by the authors and will be communicated for publication.

## REFERENCES

J. M. AREVALILLO, H. NAVARRO (2012). *A study of the effect of kurtosis on discriminant analysis under elliptical populations*. Journal of Multivariate Analysis, 107, pp. 53–63.

L. BAGNATO, A. PUNZO, M. G. ZOIA (2017). *The multivariate leptokurtic-normal distribution and its application in model-based clustering*. Canadian Journal of Statistics, 45, no. 1, pp. 95–119.

K. P. BALANDA, H. MACGILLIVRAY (1988). *Kurtosis: a critical review*. The American Statistician, 42, no. 2, pp. 111–119.

K. P. BALANDA, H. L. MACGILLIVRAY (1990). *Kurtosis and spread*. Canadian Journal of Statistics, 18, no. 1, pp. 17–30.

S. CAMBANIS, S. HUANG, G. SIMONS (1981). *On the theory of elliptically contoured distributions*. Journal of Multivariate Analysis, 11, no. 3, pp. 368–385.

K. V. MARDIA (1970). *Measures of multivariate skewness and kurtosis with applications*. Biometrika, 57, no. 3, pp. 519–530.

G. P. SZEGÖ (2004). *Risk measures for the 21st century*, vol. 1. Wiley New York.

J. WANG, W. ZHOU (2012). *A generalized multivariate kurtosis ordering and its applications*. Journal of Multivariate Analysis, 107, pp. 169–180.

SUMMARY

In a good number of real life situations, the observations on a random variable of interest tend to concentrate either too closely or too thinly around a central point but symmetrically like the normal distribution. The symmetric structure of the density function appears like that of a normal distribution but the concentration of the observations can be either thicker or thinner around the mean. This paper attempts to generate a family of densities that are symmetric like normal but with different kurtosis. Drawing inspiration from a recent work on multivariate leptokurtic normal distribution, this paper seeks to consider the univariate case and adopt a different approach to generate a family to be called 'univariate non-mesokurtic normal' family.The symmetricity of the densities is brought out by a uniform random variable while the kurtosis variation is brought about by a chi generator. Some of the properties of the resulting class of distributions and the pameter estimation are discussed.
*Keywords:* Kurtosis; Moment estimators; Non-mesokurtic distributions