A general class of trimodal distributions: properties and inference

Roberto Vila a, Victor Serra a, Mehmet Niyazi Çankaya b,c and Felipe Quintino d

aDepartamento de Estatística, Universidade de Brasília, Brasília, Brazil; bFaculty of Applied Sciences, Department of International Trading and Finance, Uşak University, Uşak, Turkey; cFaculty of Art and Sciences, Department of Statistics, Uşak University, Uşak, Turkey; dDepartamento de Estatística, Universidade de Brasília, Ji-Paraná-RO, Brazil

ABSTRACT
The modality is an important topic for modelling. Using parametric models is an efficient way when real data set shows trimodality. In this paper, we propose a new class of trimodal probability distributions, that is, probability distributions that have up to three modes. Trimodality itself is achieved by applying a proper transformation to density function of certain continuous probability distributions. At first, we obtain preliminary results for an arbitrary density function $g(x)$ and, next, we focus on the Gaussian case, studying trimodal Gaussian model more deeply. The Gaussian distribution is applied to produce the trimodal form of Gaussian known as normal distribution. The tractability of analytical expression of normal distribution and properties of the trimodal normal distribution are important reasons why we choose normal distribution. Furthermore, the existing distributions should be improved to be capable of modelling efficiently when there exists a trimodal form in a data set. After new density function is proposed, estimating its parameters is important. Since Mathematica 12.0 software has optimization tools and important modelling techniques, computational steps are performed using this software. The bootstrapped form of real data sets are applied to show the modelling ability of the proposed distribution when real data sets show trimodality.

1. Introduction

The modality is an important topic when the nature of phenomena can be modelled using the function which is capable to have different forms of peaks. The modality can occur when there exists an irregularity in the output of an experiment. In the statistical view point, the random variables are non-identically distributed. In other words, there can be a mixing of some populations even if same experiment is conducted while getting outputs of the corresponding experiment [9,16,20,38]. In such a situation, location and scale parameters of the mixed populations are important to get the central tendency and dispersion (statistics) of the mixed form of the population in the experiment after shape, scale and

CONTACT Mehmet Niyazi Çankaya mehmet.cankaya@usak.edu.tr

Supplemental data for this article can be accessed online at https://doi.org/10.1080/02664763.2023.2207785.

© 2023 Informa UK Limited, trading as Taylor & Francis Group
bimodality parameters of the function in the statistical theory are necessary components of a function which is used for conducting an efficient modelling [11,12,16,36,37,39].

There are many well-known distributions used at the statistical inference in which regression and its counter parts such as time series, design of experiments and structural equation modelling in the applications from social science are applied. There are different techniques to produce a probability density function (PDF) [24]. The analytical tractability and properties of the proposed distribution are important when the new distribution is used for modelling. For example, the existence of moments and entropy function are important when the modelling on the real data sets are performed [7,25,36]. The well-known normal distribution which is also known as Gaussian distribution should be transformed into a trimodal form when data sets show trimodality. The advantage of using a trimodal distribution is that a real data set can be a combination of two, three or more normal distributions with different parameter values for location and scale. The appearance of such mixed distributions can be in a trimodal form. Especially, since the working principle of a phenomena depends on many factors, it is reasonable to expect that a trimodal form for a real data set can occur. In other words, it is assumed that the random variables are identically distributed. However, the identicality is a restrictive assumption for modelling a data set. The parametric models are necessary to perform an efficient modelling when trimodal representation in a data set exists due to the structure of non-identicality (hetero data) or the mixing form of distributions. On the other side, if observations $x_1, x_2, \ldots, x_n$ are distributed as a parametric model such as trimodal form, then they are identical. In other words, it depends on where and how you look at the results of experiments if we assume that the observations are identical, because we do already have finite sample size. For this reason, it is not easy to imply whether or not the observations are identical. While managing an efficient modelling on real data sets, it is reasonable to consult parametric models which can be capable of dealing with different forms of modality. The bimodality parameters $\rho$ and $\delta$ with shape parameter $\alpha$ of Maxwell distribution in Ref. [14] are tools for us to generate different forms of modality. There are different degrees of mixing via Maxwell distribution and the expression which can help us to have different forms of modality of function for generating modality via compounding distributions [12,36]. The bimodal form on the positive part of the real line produces a trimodal form when it is reflected to the negative part of the real line via mirror imaging [4,9,13,31]. The different degree of bimodality, i.e. the length of periodicity of modality, on the positive part of the real line can be constructed by producing a new objective function based on the deformation [6,36], because the deformation which can be regarded as a kind of rescaling can make the different length for periodicity [2,3].

The smooth kernel distribution will be used to fit data sets because the strict and soft forms of trimodal normal distribution should be compared with smooth kernel technique to perform a comparison among them. Three normal distributions, represented by functions $g(x; \mu_1, \sigma_1)$, $g(x; \mu_2, \sigma_2)$ and $g(x; \mu_3, \sigma_3)$, can be mixed to get the trimodal form of normal distribution $g$. Thus, the modelling performance of the trimodal distribution constructed for the normal distribution $g$ can be tested for the mixed data sets. In fact, the mixed form of two or three functions with one mode and symmetric can show the symmetric form with trimodal representation. In other words, the groups around location parameter can be divided into two forms [9,16,20,38].
The main aim in this paper is to propose a distribution with trimodal form on the real line via using a technique, as is given by Refs. [15,20,36,38]. We keep to follow the symmetric case and our aim is to focus on the trimodal form on the real line. When there exists a trimodal form in a data set, the location and scale parameters should be estimated efficiently. In other words, each group coming from groups \( g(x; \mu_1, \sigma_1) \), \( g(x; \mu_2, \sigma_2) \) and \( g(x; \mu_3, \sigma_3) \) has its values for location and scale parameters. In our case, we try to estimate one location parameter and one scale parameter when the mixed data sets for two or three groups are used. It is important to note that the appearance of trimodality can occur via mixing the different values of parameters of functions. Since the true model for the mixed three normal distributions is chosen to estimate the location and scale parameters precisely, the performance of modelling will be increased greatly when the PDFs having modality property are taken into account. For example, the mixed distribution has parameters which are the mixing proportion \( w_1, w_2 \) and \( w_3 = 1 - w_1 - w_2 \) for three groups which are necessary to estimate. For the mixed normal distributions, we have \( \mu_1, \sigma_1, \mu_2, \sigma_2, \mu_3, \sigma_3, w_1 \) and \( w_2 \) (see Section 5). In total, there are eight parameters which have to be estimated. However, in our case we have three main parameters and also \( \mu \) and \( \sigma \) of the distribution. Thus, there will be five parameters which will be estimated. The optimization of the log-likelihood function according to these five parameters of a function can be easy to reach the global point of the log\((f)\) when compared with a function including eight parameters. In addition, we have only one location and scale parameter which can be free from the mixing proportion \( w_1, w_2 \) and \( w_3 = 1 - w_1 - w_2 \). The numerical computation while conducting the optimization of log-likelihood according to parameters can include the less numerical errors. Note that the numerical errors in the function with eight parameters can be bigger than that of five parameters. Thus, the more precise evaluation can be achieved for the numerically precise evaluation of estimating the parameters \( \mu \) and \( \sigma \), which is why we prefer to consider proposing such a trimodality while conducting an efficient fitting on the real data sets. On the other side of the modelling perspective, the structure of grouping cannot be determined only for the used mixing proportions \( w_1, w_2 \) and \( w_3 = 1 - w_1 - w_2 \). There can be irrational proportions such as 1/6, 1/9, etc. for the mixing in a data set. The precise evaluation in computation for the true value of irrational proportion cannot be performed accurately, which makes a disadvantage for us when we use the estimated values of these kinds of the proportions such as irrational ones.

This paper is organized as follows. In Section 2, we define the class of trimodal probabilistic models. In Section 3, some structural properties of the proposed model are examined. We provide a formal proof for the trimodality of a class of symmetric kernel densities, present a stochastic representation and provide the closed formulas for the moments, entropies and stochastic representation. The existence of these expressions is important to use the proposed distribution for fitting the data sets. Section 4 is divided to describe the proposed model when the normal (Gaussian) is applied (see Table 1). For this case, some properties such as modality, moments, Shannon entropy among others also are discussed. In Sections 5, we introduce the mixing form. Section 6 represents the log\(_q\)-likelihood function used as a method for parameter estimation. In Section 7, the real data sets are applied. Section 8 is for conclusion and future works. In order to provide proofs and codes of Mathematica software, a supplementary material is also provided.
2. A class of continuous probability distributions

Let $g : D \subset \mathbb{R} \to [0, \infty)$, $D = \text{supp}(g) \neq \emptyset$, be a kernel density with corresponding cumulative distribution function (CDF) denoted by $G$. The function $g$ can be associated (or not) with an additional parameter $\xi$ (or vector $\xi$). For a random variable $X$, we define the following PDF:

$$f(x; \theta) = \frac{\sigma}{Z_\theta} \left[ \rho + \delta T \left( \frac{x - \mu}{\sigma} \right) ; \alpha, p \right] g \left( \frac{x - \mu}{\sigma} \right), \quad x - \mu \in D,$$

(1)

where $\theta = (\mu, \sigma, \alpha, \rho, \delta)$ is a parameter vector such that $\mu \in \mathbb{R}$ is a location parameter, $\sigma > 0$ is a scale parameter, $\alpha > 0$ is a shape parameter and $\rho \geq 0$ and $\delta \geq 0$ are parameters that control different forms of modality of distribution. Note that $\rho$ and $\delta$ cannot be zero simultaneously. Notice that the model (1) fails to be identifiable because for $\theta = (0, 1, \alpha, 0, \delta) \neq \theta' = (0, 1, \alpha, 0, \delta')$, with $\delta \neq \delta'$ (both non-zero), we have $f(x; \theta) = f(x; \theta')$ for all $x$. Usually a parametric model that fails this property may become identifiable under certain technical restrictions (see, e.g. Ref. [32]). The function $Z_\theta$ appearing in the definition of $f$ is a normalizing factor and the function $T : D \subset \mathbb{R} \to (0, 1)$ is given by

$$T(x; \alpha; p) = \frac{\gamma(p, x^2/\alpha^2)}{\Gamma(p)} \quad \text{with } p > 0 \text{ known.}$$

(2)

Here, $\gamma(p, u) = \int_0^u w^{p-1} e^{-w} dw$ is the incomplete gamma function and $\Gamma(p)$ is the gamma function. For $p = 3/2$ and $D = (0, \infty)$, the function $T(x; \alpha, p)$ on $D$ defines the Maxwell distribution with scale parameter $\alpha$.

Hereafter, we will denote $X \sim \text{TD}(\theta)$ for a random variable $X$ that follows the trimodal distribution (1).

When $\delta = 0$ and $\rho \neq 0$ fixed in (1), the original density $g$ is recovered.

Since $0 < T(x; \alpha, p) < 1$ for almost all $x \in D$, we have $0 < Z_\theta \leq (\rho + \delta)\sigma$. A simple calculation shows that (see Corollary 3.6)

$$Z_\theta = (\rho + \delta)\sigma + \delta \sigma \left\{ \mathbb{E} \left[ G(-\alpha \sqrt{Y}) \right] - \mathbb{E} \left[ G(\alpha \sqrt{Y}) \right] \right\},$$

(3)

where $Y \sim \text{Gamma}(p, 1)$ and $G$ is the corresponding CDF of $g$. Furthermore, the CDF of $X \sim \text{TD}(\theta)$, denoted by $F(x; \theta)$, is written as

$$F(x; \theta) = \frac{\rho\sigma}{Z_\theta} G \left( \frac{x - \mu}{\sigma} \right) + \delta \sigma \left\{ \mathbb{E} \left[ G(-\alpha \sqrt{Y}) \right] + T \left( \frac{x - \mu}{\sigma} ; \alpha, p \right) G \left( \frac{x - \mu}{\sigma} \right) \right\}$$

$$- \frac{\delta \sigma}{Z_\theta} \left\{ \mathbb{E} \left[ \mathbb{1}_{\{Y \leq ((x - \mu)/\sigma \alpha)\}} G(-\alpha \sqrt{Y}) \right] \mathbb{1}_{\{x < -\mu\}} \right\}$$

$$+ \mathbb{E} \left[ \mathbb{1}_{\{Y \leq ((x - \mu)/\sigma \alpha)\}} G(\alpha \sqrt{Y}) \mathbb{1}_{\{x \geq \mu\}} \right],$$

(4)

for each $x \in \mathbb{R}$. For more details, see Corollary 3.6.

Taking $x = \mu$ in (4), we get $F(\mu; \theta) = \sigma \left( \rho G(0) + \delta \mathbb{E} [G(-\alpha \sqrt{Y})] \right)/Z_\theta$. Let $W$ be a random variable with corresponding CDF $G$. If the distribution $G$ of $W$ is symmetric about zero, then $Z_\theta = \rho \sigma + 2\delta \sigma \mathbb{E} [G(-\alpha \sqrt{Y})], G(0) = 1/2$ and then $F(\mu; \theta) = 1/2$. So, in this
Table 1. Some kernel densities \((g)\) that generate multimodality in the model \((1)\).

| Distribution       | \(g\)                                      | \(G\)                                      | \(\xi\) | \(D\)  |
|--------------------|--------------------------------------------|--------------------------------------------|---------|--------|
| Trimodal Gumbel    | \(e^{-x} - e^{-x}\)                        | \(1 + \frac{1}{\psi}e^{-x}\)              |         | \(\mathbb{R}\) |
| Trimodal Laplace   | \(\frac{1}{2}e^{-|x|}\)                    | \(1 + \frac{1}{\psi}e^{-|x|}\)            |         | \(\mathbb{R}\) |
| Trimodal Logistic  | \(\frac{1}{1+e^{-x^2}}\)                   | \(\frac{1}{\pi e^{-x^2}}\)                |         | \(\mathbb{R}\) |
| Trimodal Cauchy    | \(\frac{1}{\pi e^{-x^2}}\)                | \(\frac{1}{\pi}e^{-x^2}\)                |         | \(\mathbb{R}\) |
| Trimodal Student-t | \(\frac{\Gamma((v+1)/2)}{\sqrt{\pi}}\)\left(1 + \frac{x^2}{\nu}\right)^{-(v+1)/2}\) | \(\frac{1}{2} + \frac{1}{\psi}F_{1}(1/2, (v+1)/2, 2)\) | \(v > 0\) | \(\mathbb{R}\) |
| Trimodal Normal    | \(\phi(x) = \frac{1}{\sqrt{\pi}}e^{-x^2/2}\) | \(\Phi(x) = \int_{-\infty}^{x} \phi(t) dt = \frac{1}{2}[1 + \text{erf}(x/\sqrt{\pi})]\) | \(-\infty, \infty\) | \(\mathbb{R}\) |

case, \(\mu\) is location parameter of \(X\). Moreover, by letting \(x \to \infty\) in \((4)\), a simple observation shows that \(F(x; \theta)\) tends to 1, showing that the parametric function in \((1)\) is in fact a PDF.

Some natural examples of kernel densities \(g\) to be plugged into \((1)\), with \(p\) given, where trimodality shape is observed, are presented in Table 1. Here, \(2F_1\) is the hypergeometric function and \(\text{erf}(x) = 2\int_0^x e^{-t^2} dt/\sqrt{\pi}\) is the error function (also called the Gauss error function).

3. Structural properties

In this section, some basic properties such as trimodality for symmetric kernels, moments and truncated moments, and entropies for \(X \sim \text{TD}(\theta)\) are discussed in detail.

3.1. Trimodality for a class of symmetric kernel densities

In this subsection, we suppose that the kernel density \(g\) in \((1)\) has the following form:

\[
g(x) = g(0) e^{-\int_0^x \tilde{\eta}(t^2) dt}, \quad x \in \mathbb{R},
\]  

(5)

for some decreasing and positive real function \(\tilde{\eta}\) such that the integral \(\int_0^x \tilde{\eta}(t^2) dt\) exists. It is clear that \(g(0)\) is the normalization constant so that \(g\) is a PDF. Notice that \((5)\) is equivalent to

\[
g'(x) = -x\tilde{\eta}(x^2)g(x), \quad x \in \mathbb{R}.
\]  

(6)

It is immediate to verify that \(g\), as defined in \((5)\), is symmetric about zero, that is, \(g(x) = g(-x)\) on the real line \(D = \mathbb{R}\). For example, in the Laplace, Cauchy, Student-\(t\) and Normal kernel densities (see Table 1), we have \(\tilde{\eta}(y) = 1/\sqrt{\pi},\ \tilde{\eta}(y) = 2/(1 + y),\ \tilde{\eta}(y) = (\nu + 1)/(\nu + y)\) and \(\tilde{\eta}(y) = 1\ \forall y > 0,\) respectively.

Moreover, we assume that \(\tilde{\eta}\) has the following form:

\[
\tilde{\eta}(y) = \frac{C}{(\alpha^2 A + y)^{\beta - p}} \quad \text{for } C \geq 1, A \geq 0 \text{ and } \beta > p.
\]  

(7)

That is, \(\tilde{\eta}(y), y > 0,\) decays polynomially. The function \(\tilde{\eta}(y) = 1\) corresponding to the Normal distribution is not of the form \((7)\), then the next result cannot be applied and a separate study must be carried out. In this paper, the Gaussian case will be studied in detail in Section 4.
For a formal proof of the following lemma, see Lemma 1.1 of Supplementary Material I of this work.

Lemma 3.1: Let \( \mathcal{H} \) be as in (7). For some \( \rho > 0 \), the function \( \mathcal{R} \), defined by

\[
\mathcal{R}(y) = 2\delta \left( \frac{1}{\alpha^2} \right)^p \frac{y^{p-1} e^{-y/\alpha^2}}{\Gamma(p)} - \left[ \rho + \delta \frac{\gamma(p, y/\alpha^2)}{\Gamma(p)} \right] h(y), \quad y > 0,
\]

has at most two real roots.

Proposition 3.2: Let \( g \) be a kernel density as in (5), with \( \mathcal{H} \) as in (7). A point \( x \in \mathbb{R} \) is a critical point of density (1) if \( x = \mu \) or \( R\left(\frac{(x - \mu)^2}{\sigma^2}\right) = 0 \), where \( R \) is as in Lemma 3.1.

Proof: The proof is immediate since, by using Equation (6), the first-order derivative of \( f(x; \theta) \), with respect to \( x \), is given by

\[
f'(x; \theta) = \left( \frac{x - \mu}{\sigma} \right) g((x - \mu)/\sigma) R\left(\frac{(x - \mu)^2}{\sigma^2}\right)/(\sigma Z_{\theta}).
\]

■

Theorem 3.3 (Uni-, bi- or trimodality): If \( X \sim \text{TD}(\theta) \), then the following holds:

1. If \( \mathcal{R} \) has no real roots, then \( f(x; \theta) \) is unimodal with mode \( x = \mu \).
2. If \( \mathcal{R} \) has one real root, then \( f(x; \theta) \) is bimodal with minimum point \( x = \mu \).
3. If \( \mathcal{R} \) has two distinct real roots, then \( f(x; \theta) \) is trimodal where \( x = \mu \) is one of the modes.

Proof: It is clear that if \( \mathcal{R} \) has no real roots, by Proposition 3.2, \( x = \mu \) is the only critical point of the density \( f \). Since \( \lim_{x \to \pm \infty} f(x; \theta) = 0 \), the point \( x = \mu \) is a mode. This proves Item (1).

In order to prove Item (2), we suppose that \( \mathcal{R} \) has one real root, denoted by \( a \). By Proposition 3.2, it follows that \( x = \mu \) and \( x = \mu \pm \sigma \sqrt{a} \) are three critical points of \( f \). Since \( \lim_{x \to \pm \infty} f(x; \theta) = 0 \), the point \( x = \mu \) is a minimum and \( x = \mu \pm \sigma \sqrt{a} \) are two symmetrical modes. This proves the second item.

Now, we assume that \( \mathcal{R} \) has two distinct real roots, denoted by \( a \) and \( b \). Without loss of generality, we can assume that \( a < b \). Again, by Proposition 3.2, we have that \( x = \mu, x = \mu \pm \sigma \sqrt{a} \) and \( x = \mu \pm \sigma \sqrt{b} \) are five critical points of \( f \). Since \( \lim_{x \to \pm \infty} f(x; \theta) = 0 \) and \( a < b \), the critical points \( x = \mu \) and \( x = \mu \pm \sigma \sqrt{b} \) are modes and \( x = \mu \pm \sigma \sqrt{a} \) are minimum points. Hence, the proof of Item (3) follows. ■

3.2. Moments

Theorem 3.4: Let \( X \sim \text{TD}(\theta) \) and \( L : \mathbb{R} \to \mathbb{R} \) be a Borel-measurable function. Then, the expectation of random variable \( L(X) \) with \( X \leq b \) and \( b \in \mathbb{R} \), is given by

\[
\mathbb{E}\left[ \mathbb{I}_{\{X \leq b\}} L(X) \right] = \frac{\rho \sigma}{Z_{\theta}} \mathbb{E}\left[ \mathbb{I}_{\{W \leq (b-\mu)/\sigma\}} L(W_{\mu, \sigma}) \right] + \frac{\delta \sigma}{Z_{\theta}} \mathbb{E}\left[ \mathbb{I}_{\{Y \leq ((b-\mu)/\alpha)^2, W \leq (b-\mu)/\sigma\}} L(W_{\mu, \sigma}) \right]
\]
\[\begin{align*}
&+ \mathbb{E} \left[ \mathbbm{1}_{\{Y \geq \frac{(b-\mu)}{\alpha \sigma} \}, W \leq -\alpha \sqrt{Y}} L(W_\mu, \sigma) \right] \mathbbm{1}_{\{b < \mu\}} \\
&+ \frac{\delta \sigma}{Z_\theta} \left\{ \mathbb{E} \left[ \mathbbm{1}_{\{W \leq -\alpha \sqrt{Y}} L(W_\mu, \sigma) \right] \\
&+ \mathbb{E} \left[ \mathbbm{1}_{\{Y \leq \frac{(b-\mu)}{\alpha \sigma}, \alpha \sqrt{Y} \leq W \leq \frac{(b-\mu)}{\sigma} \}} L(W_\mu, \sigma) \right] \mathbbm{1}_{\{b \geq \mu\}}, \right. \\
\end{align*}\]

where \( W_{\mu, \sigma} = \sigma W + \mu \), \( W \) is a continuous random variable with CDF \( G \) (that for brevity we write \( W \overset{d}{=} G \)), \( Y \sim \text{Gamma}(p, 1) \), and \( W \) and \( Y \) are independent.

**Proof:** By using the definition of expectation and by taking the change of variables \( w = (x - \mu) / \sigma \) and \( dx = \sigma \, dw \), we have

\[\begin{align*}
\mathbb{E} \left[ \mathbbm{1}_{\{X \leq b\}} L(X) \right] &= \frac{\rho}{Z_\theta} \int_{\sigma D + \mu} \mathbbm{1}_{\{x \leq b\}} L(x) g \left( \frac{x - \mu}{\sigma} \right) \, dx \\
&\quad + \frac{\delta}{Z_\theta} \int_{\sigma D + \mu} \mathbbm{1}_{\{x \leq b\}} L(x) \tau \left( \frac{x - \mu}{\sigma}; \alpha, p \right) g \left( \frac{x - \mu}{\sigma} \right) \, dx \\
&= \frac{\rho \sigma}{Z_\theta} \int_D \mathbbm{1}_{\{w \leq \frac{(b-\mu)}{\sigma} \}} L(\sigma w + \mu) g(w) \, dw \\
&\quad + \frac{\delta \sigma}{Z_\theta} \int \int_{w \leq \frac{(b-\mu)}{\sigma}} \mathbbm{1}_{\{w \leq \frac{(b-\mu)}{\sigma} \}} \tau(w, y) \, dy \, dw, \quad (8) \\
\end{align*}\]

where, for notational simplicity, we denote

\[\tau(w, y) = L(\sigma w + \mu) g(w) \left[ \frac{y^{\rho-1} e^{-y}}{\Gamma(p)} \right].\]

There are two cases to consider according to whether \( \xi := \frac{(b-\mu)}{\sigma} < 0 \) or \( \xi := \frac{(b-\mu)}{\sigma} \geq 0 \); see Figure 1(a,b). In the former case,

\[\begin{align*}
\int \int_{w \leq \frac{(b-\mu)}{\sigma}} &\mathbbm{1}_D(\omega) \tau(w, y) \, dy \, dw = \int_0^{\frac{(b-\mu)}{\alpha \sigma}} \left( \int_{-\infty}^{\frac{(b-\mu)}{\sigma}} \mathbbm{1}_D(\omega) \tau(w, y) \, dw \right) \, dy \\
&\quad + \int_{\frac{(b-\mu)}{\alpha \sigma}}^{\infty} \left( \int_{-\infty}^{-\alpha \sqrt{y}} \mathbbm{1}_D(\omega) \tau(w, y) \, dw \right) \, dy \\
\end{align*}\]

and in the latter case,

\[\begin{align*}
\int \int_{w \leq \frac{(b-\mu)}{\sigma}} &\mathbbm{1}_D(\omega) \tau(w, y) \, dy \, dw = \int_0^{\infty} \left( \int_{-\infty}^{-\alpha \sqrt{y}} \mathbbm{1}_D(\omega) \tau(w, y) \, dw \right) \, dy \\
&\quad + \int_0^{\frac{(b-\mu)}{\alpha \sigma}} \left( \int_{\alpha \sqrt{y}}^{\frac{(b-\mu)}{\sigma}} \mathbbm{1}_D(\omega) \tau(w, y) \, dw \right) \, dy. \\
\end{align*}\]
Figure 1. (a) \( w \leq \xi, 0 < y \leq w^2/\alpha^2, \xi < 0 \) and (b) \( w \leq \xi, 0 < y \leq w^2/\alpha^2, \xi \geq 0 \).

Hence, by combining the last two integral identities with (8), when \( b < \mu \), we have

\[
\mathbb{E} \left[ \mathbf{1}_{\{X \leq b\}} L(X) \right] = \frac{\rho \sigma}{Z_\theta} \int_D \mathbf{1}_{\{w \leq (b - \mu)/\alpha\}} L(\sigma w + \mu) g(w) \, dw \\
+ \frac{\delta \sigma}{Z_\theta} \int_0^{(b - \mu)/\alpha^2} \left( \int_{-\infty}^{(b - \mu)/\alpha} \mathbf{1}_D(\omega) \tau(w, y) \, dw \right) dy \\
+ \frac{\delta \sigma}{Z_\theta} \int_{(b - \mu)/\alpha^2}^{\infty} \left( \int_{-\infty}^{-\alpha \sqrt{y}} \mathbf{1}_D(\omega) \tau(w, y) \, dw \right) dy
\]

and, when \( b \geq \mu \),

\[
\mathbb{E} \left[ \mathbf{1}_{\{X \leq b\}} L(X) \right] = \frac{\rho \sigma}{Z_\theta} \int_D \mathbf{1}_{\{w \leq (b - \mu)/\alpha\}} L(\sigma w + \mu) g(w) \, dw \\
+ \frac{\delta \sigma}{Z_\theta} \int_0^{\infty} \left( \int_{-\infty}^{-\alpha \sqrt{y}} \mathbf{1}_D(\omega) \tau(w, y) \, dw \right) dy \\
+ \frac{\delta \sigma}{Z_\theta} \int_0^{(b - \mu)/\alpha^2} \left( \int_{\alpha \sqrt{y}}^{(b - \mu)/\alpha} \mathbf{1}_D(\omega) \tau(w, y) \, dw \right) dy.
\]

Then there are \( W \overset{d}{=} G \) and \( Y \sim \text{Gamma}(\rho, 1) \) independent so that, for \( b < \mu \),

\[
\mathbb{E} \left[ \mathbf{1}_{\{X \leq b\}} L(X) \right] = \frac{\rho \sigma}{Z_\theta} \mathbb{E} \left[ \mathbf{1}_{\{W \leq (b - \mu)/\alpha\}} L(W_{\mu, \sigma}) \right] \\
+ \frac{\delta \sigma}{Z_\theta} \left\{ \mathbb{E} \left[ \mathbf{1}_{\{(Y/(b - \mu)/\alpha)^2, W \leq (b - \mu)/\alpha\}} L(W_{\mu, \sigma}) \right] \\
+ \mathbb{E} \left[ \mathbf{1}_{\{(Y/(b - \mu)/\alpha)^2, W \leq -\alpha \sqrt{Y}\}} L(W_{\mu, \sigma}) \right] \right\}
\]
and, for \( b \geq \mu \),
\[
\mathbb{E} \left[ I_{\{X \leq b\}} L(X) \right] = \frac{\rho \sigma}{Z_{\theta}} \mathbb{E} \left[ I_{\{W \leq (b-\mu)/\sigma\}} L(W, \sigma) \right] + \frac{\delta \sigma}{Z_{\theta}} \mathbb{E} \left[ I_{\{W \leq -\alpha \sqrt{Y}\}} L(W, \sigma) \right] + \mathbb{E} \left[ I_{\{Y \leq ((b-\mu)/\alpha)^2, \alpha \sqrt{Y} \leq \frac{(b-\mu)}{\sigma}\}} L(W, \sigma) \right].
\]

This completes the proof. \( \blacksquare \)

By taking \( b \to \infty \) in Theorem 3.4, with \((X - \mu)/\sigma\) instead of \(X\), we get

**Corollary 3.5:** Under the hypotheses of Theorem 3.4,
\[
\mathbb{E} \left[ L \left( \frac{X - \mu}{\sigma} \right) \right] = \frac{(\rho + \delta)\sigma}{Z_{\theta}} \mathbb{E}[L(W)] + \frac{\delta \sigma}{Z_{\theta}} \left( \mathbb{E} \left[ I_{\{W \leq -\alpha \sqrt{Y}\}} L(W) \right] - \mathbb{E} \left[ I_{\{W \leq \alpha \sqrt{Y}\}} L(W) \right] \right).
\]

**Corollary 3.6:** Under the hypotheses of Theorem 3.4, if

(a) \( b \to \infty \) and \( L(x) = 1 \forall x \in \sigma D + \mu \), then the formula (3) for the normalizing factor \( Z_{\theta} \) is obtained.

(b) \( b = x \in \mathbb{R} \) fixed and \( L(x) = 1 \forall x \in \sigma D + \mu \), then the formula (4) of the CDF \( F(x; \theta) \) is obtained.

By taking \( b \to \infty \) in Theorem 3.4, with \( L(x) = x^n \forall x \in \sigma D + \mu \) and \( n \geq 1 \) integer, by a binomial expansion, we have the next formula for the moments.

**Corollary 3.7:** The \( n \)th moment of \( X \sim \text{TD(} \theta \text{)} \) is written as
\[
\mathbb{E}(X^n) = \sum_{k=0}^{n} \binom{n}{k} \mu^{n-k} \sigma^{k} \left\{ \frac{(\rho + \delta)\sigma}{Z_{\theta}} \mathbb{E}(W^k) + \frac{\delta \sigma}{Z_{\theta}} \left( \mathbb{E} \left[ I_{\{W \leq -\alpha \sqrt{Y}\}} W^k \right] - \mathbb{E} \left[ I_{\{W \leq \alpha \sqrt{Y}\}} W^k \right] \right) \right\}.
\]

The above formula informs that the moments of \( X \) (whenever they exist) depend on the existence of moments of \( I_{\{W \leq \pm \alpha \sqrt{Y}\}} W \), with \( W \overset{d}{=} G \) and \( Y \sim \text{Gamma}(p, 1) \) independent.
3.3. Entropies

The Tsallis [34] entropy associated with the random variable $X \sim \text{TD}(\theta)$ is defined as

$$S_q(X) = \begin{cases} -\int_{\sigma \Delta + \mu} f_q(x; \theta) \log_q f(x; \theta) \, dx & \text{if } q \neq 1, \\ -\int_{\sigma \Delta + \mu} f(x; \theta) \log f(x; \theta) \, dx & \text{if } q = 1, \end{cases}$$

where, for $x > 0$,

$$\log_q(x) = \begin{cases} x^{1-q} - 1 & \text{if } q \neq 1, \\ \frac{1-q}{1-q} \log(x) & \text{if } q = 1 \end{cases}$$

(9)

represents a Box-Cox transformation in statistics (often called deformed logarithm \[18,35\]). Since $\log_q(x) \to \log(x)$ when $q \to 1$, we have $S_q(X) \to S_1(X)$ when $q \to 1$. That is, when $q \to 1$, the usual definition of Shannon’s entropy $S_1(X)$ [33] is recovered.

Proposition 3.8: Under the hypotheses of Theorem 3.4,

$$\mathbb{E} \left[ f_q^{-1}(X; \theta) \right] = \frac{(\rho + \delta)\sigma}{Z_\theta} \mathbb{E} \left[ f_q^{-1}(\sigma W + \mu; \theta) \right] + \frac{\delta \sigma}{Z_\theta} \left\{ \mathbb{E} \left[ \mathbb{1}_{W \leq -\alpha \sqrt{Y}} f_q^{-1}(\sigma W + \mu; \theta) \right] - \mathbb{E} \left[ \mathbb{1}_{W \leq \alpha \sqrt{Y}} f_q^{-1}(\sigma W + \mu; \theta) \right] \right\},$$

where $W \overset{d}{=} G, Y \sim \text{Gamma}(p, 1)$, and $W$ and $Y$ are independent. Consequently, the Tsallis entropy of $X$ (whenever it exists) depends on the existence of truncated moments of $f_q^{-1}(W; \theta)$.

Proof: By taking $b \to \infty$ in Theorem 3.4, with $L(x) = f_q^{-1}(x; \theta)$ $\forall x \in \sigma \Delta + \mu$, the proof follows. ■

For a formal proof of next result, see Proposition 1.2 of Supplementary Material I of this paper.

Proposition 3.9: If $S_q(W)$ exists and $q > 0$, then $S_q(X)$, with $X \sim \text{TD}(\theta)$, also exists.

Remark 3.1: As a consequence of Proposition 3.9, by letting $q \to 1$, the Shannon entropy $S_1(X)$ exists whenever $S_1(W)$ also exists.

For a rigorous proof of next result, see Proposition 1.4 of Supplementary Material I of this paper.
**Proposition 3.10:** Under the hypotheses of Theorem 3.4, the Shannon entropy of $X \sim TD(\theta)$ is written as

$$S_1(X) = \log(Z_\theta) - \frac{(\rho + \delta)\sigma}{Z_\theta} \mathbb{E} \left[ \log(\rho + \delta T(W; \alpha, p)) \right]$$

$$- \frac{\delta \sigma}{Z_\theta} \left\{ \mathbb{E} \left[ \mathbb{1}_{\{W \leq -\alpha \sqrt{T}\}} \log(\rho + \delta T(W; \alpha, p)) \right] \right\}$$

$$+ \frac{(\rho + \delta)\sigma}{Z_\theta} S_1(W) + \frac{\delta \sigma}{Z_\theta} \left\{ \mathbb{E} \left[ S_1 \left( \mathbb{1}_{\{W \leq -\alpha \sqrt{T}\}} W \right) \right] \right\},$$

where $T$ was defined in (2).

### 3.4. Stochastic representation

Let $h(u)$, $0 < u < 1$, be a PDF with corresponding CDF $H$. Let $S : D \rightarrow (0, 1)$ be an injective and increasing transformation, where $D$ is a non-empty set of $\mathbb{R}$. We consider the following CDF:

$$F(z) = \int_0^z h(u) \, du = H(S(z)), \quad z \in D. \quad (10)$$

We also define by $f$ to the corresponding PDF of $F$. That is, $F'(z) = f(z) = h(S(z))S'(z)$ for almost all $z \in D$.

We define the PDF $h$ as follows:

$$h(u) = \frac{\alpha}{Z_\theta} \left[ \rho + \delta T \left( G^{-1}(u); \alpha, p \right), \quad 0 < u < 1, \rho \geq 0, \delta \geq 0, \alpha > 0, \sigma > 0, \quad (11) \right.$$ 

where $G$ and $G^{-1}$, respectively, are the CDF defined in Section 2 and its inverse function, and $T$ is as in (2). When $\delta = 0$, $h$ reduces to the continuous uniform distribution on the interval $(0, 1)$. The CDF $H$ of $h$ is given by

$$H(u) = \frac{\rho}{Z_\theta} u + \frac{\delta}{Z_\theta} \left\{ \mathbb{E} \left[ H(G(-\alpha \sqrt{Y}) + S \left( G^{-1}(u); \alpha, p \right) u \right] \right\}$$

$$- \frac{\delta}{Z_\theta} \left\{ \mathbb{E} \left[ \mathbb{1}_{\{Y \leq [G^{-1}(u)/\alpha]^2\}} G(-\alpha \sqrt{Y}) \right] \mathbb{1}_{\{G^{-1}(u) < 0\}} \right\}$$

$$+ \mathbb{E} \left[ \mathbb{1}_{\{Y \leq [G^{-1}(u)/\alpha]^2\}} G(\alpha \sqrt{Y}) \right] \mathbb{1}_{\{G^{-1}(u) \geq 0\}} \right\},$$

where $Y \sim \text{Gamma}(p, 1)$.

If $S : D \rightarrow (0, 1)$ is defined by $S(z) = G(z) \forall z \in D$, by (10), the family of trimodal distributions in (1) is obtained. That is,

$$F(x; \theta) = F \left( \frac{x - \mu}{\sigma} \right) = H \left( G \left( \frac{x - \mu}{\sigma} \right) \right) \quad (12)$$

and $f(x; \theta) = (1/\sigma) f((x - \mu)/\sigma)$. 

If $U$ is distributed according to (11) and $X \sim \text{TD}(\theta)$, then, by (12), the random variable $X$ admits the following stochastic representation:

$$X = \mu + \sigma G^{-1}(U).$$

### 4. The Gaussian case of trimodality

In this section, the standard normal kernel density $g(x) = \phi(x), x \in \mathbb{R}$, is plugged into (1). Some structural properties as modality, moments, entropies and rate of the distribution are discussed.

The CDF corresponding to the normal kernel density $g$ is $G(x) = \Phi(x)$, with

$$\Phi(x) = \int_{-\infty}^{x} \phi(t) \, dt = \frac{1}{2} \left[ 1 \pm \text{erf} \left( \frac{x}{\sqrt{2}} \right) \right],$$

where $\text{erf}(x)$ is the error function. So, in this section, we consider the following PDF:

$$f(x; \theta) = \frac{1}{Z_\theta} \left[ \rho + \delta T \left( \frac{x-\mu}{\sigma} ; \alpha, p \right) \right] \phi \left( \frac{x-\mu}{\sigma} \right), \quad x \in \mathbb{R},$$

where $T$ is as in (2) and $Z_\theta$ is the normalizing factor (3), which is given by

$$Z_\theta = (\rho + \delta) \sigma + \delta \sigma \left\{ E \left[ \Phi(-\alpha \sqrt{Y}) \right] - E \left[ \Phi(\alpha \sqrt{Y}) \right] \right\}.$$ (15)

We will denote $X \sim \text{TD}_\Phi(\theta)$ for a random variable $X$ that follows (14). Plots of the TD$_\Phi$ density, where trimodality is observed, are given in Figure 2(a,b).

#### 4.1. A study on the modality

For a mathematical proof of next result, see Lemma 1.5 of Supplementary Material I of this work.

**Lemma 4.1:** The function $R$, defined by

$$R(y) = 2\delta \left( \frac{1/\alpha^2}{\Gamma(p)} y^{-1/\alpha^2} \right) e^{-y/\alpha^2} - \left[ \rho + \delta \gamma(p, y/\alpha^2) \Gamma(p) \right], \quad y > 0,$$

has at most two real roots.

![Figure 2. PDFs and CDFs of trimodal normal distribution. (a) The strict form of modality for trimodal normal distribution and (b) the soft form of modality for trimodal normal distribution.](image-url)
Remark 4.1: Notice that when \( \rho \) is sufficiently large, the function \( R \) has no roots. For sufficiently small \( \rho \), \( R \) have one, or two roots depending on whether \( p \leq 1 \) or \( p > 1 \). Furthermore, when \( \delta = 0 \) (with \( \rho > 0 \)) or \( \alpha \to \infty \), the function \( R \) has no roots.

Proposition 4.2: A point \( x \in \mathbb{R} \) is a critical point of density (14) if \( x = \mu \) or \( R[(x - \mu)^2/\sigma^2] = 0 \), where \( R \) is as in Lemma 4.1.

Proof: The proof follows from identity \( f'(x; \theta) = \frac{1}{(\sigma Z_{\theta})} \frac{g((x - \mu)/\sigma)}{\sigma Z_{\theta}} R[(x - \mu)^2/\sigma^2] \).

Theorem 4.3 (Uni-, bi- or trimodality): Let \( X \sim TD_{\Phi}(\theta) \) and \( R \) be the function defined in Lemma 4.1. The following holds:

1. If \( R \) has no real roots, then \( f(x; \theta) \) is unimodal with mode \( x = \mu \).
2. If \( R \) has one real root, then \( f(x; \theta) \) is bimodal with minimum point \( x = \mu \).
3. If \( R \) has two distinct real roots, then \( f(x; \theta) \) is trimodal where \( x = \mu \) is one of the modes.

Proof: The proof follows the same steps of proof of Theorem 3.3 by taking \( R \) instead of \( R \), and by using Proposition 4.2 instead of Proposition 3.2.

4.2. The normalizing factor

In what follows we find a closed expression for the normalizing factor \( Z_{\theta} \) in (14). By (15), it is necessary to calculate \( \mathbb{E}[\Phi(\pm \alpha \sqrt{Y})] \), where \( Y \sim \text{Gamma}(p, 1) \). Indeed, using the formula (13) and then taking the change of variables \( z = \sqrt{y} \) and \( dy = 2z \, dz \), we obtain

\[
\mathbb{E}[\Phi(\pm \alpha \sqrt{Y})] = \int_0^{\infty} \frac{\Phi(\pm \alpha \sqrt{y}) y^{p-1} e^{-y}}{\Gamma(p)} \, dy = \frac{1}{2} \left[ \frac{1}{2} + \frac{1}{2} \int_0^{\infty} \text{erf} \left( \pm \frac{\alpha \sqrt{y}}{\sqrt{2}} \right) \frac{y^{p-1} e^{-y}}{\Gamma(p)} \, dy \right].
\]

By using the formula (see Item 8 in Subsection 4.3 of Ref. [27]):

\[
\int_0^{\infty} \text{erf}(ax)x^p e^{-b^2 x^2} \, dx = \frac{a}{\sqrt{\pi}} b^{p-2} \Gamma \left( \frac{p}{2} + 1 \right) {}_2F_1 \left( \frac{1}{2}, \frac{p}{2} + 1; \frac{3}{2}; -\frac{a^2}{b^2} \right), \quad b^2 > 0, p > -2,
\]

where \( {}_2F_1(a_1, a_2; b_1; x) \) is the generalized Hypergeometric function, we have

\[
\int_0^{\infty} \text{erf} \left( \pm \frac{\alpha z}{\sqrt{2}} \right) \frac{z^{2p-1} e^{-z^2}}{\Gamma(p)} \, dz = \pm \frac{\alpha \Gamma(p + \frac{1}{2})}{\sqrt{2\pi} \Gamma(p)} {}_2F_1 \left( \frac{1}{2}, \frac{p + 1}{2}; \frac{3}{2}; -\frac{\alpha^2}{2} \right).
\]

Therefore,

\[
\mathbb{E}[\Phi(\pm \alpha \sqrt{Y})] = \frac{1}{2} \left[ \frac{1}{2} + \frac{\alpha \Gamma(p + (1/2))}{\sqrt{2\pi} \Gamma(p)} {}_2F_1 \left( \frac{1}{2}, \frac{p + 1}{2}; \frac{3}{2}; -\frac{\alpha^2}{2} \right) \right].
\]
Replacing (17) in (15), we obtain the following closed expression for the normalizing factor \( Z_\theta \):

\[
Z_\theta = (\rho + \delta)\sigma - \frac{2 \delta \sigma \alpha \Gamma(p + (1/2))}{\sqrt{2\pi} \Gamma(p)} \cdot 2F_1 \left( \frac{1}{2}, p + \frac{3}{2}; \frac{-\alpha^2}{2} \right).
\]  

(18)

The flexibility of the TD model is shown. Note that the TD model can be unimodal, bimodal or trimodal. Figure 2(a,b) shows how the TD density function is influenced by parameters \( \alpha, \rho \) and \( \delta \).

4.3. CDF

To determine the CDF of \( X \sim TD_\Phi(\theta) \), we use formula (4). So, by (4), it is essential to calculate the expectation \( \mathbb{E}[1_{\{Y \leq ((x - \mu)/\alpha \sigma)^2\}} \Phi(\pm \alpha \sqrt{Y})] \), \( x \in \mathbb{R} \), where \( Y \sim \text{Gamma}(p, 1) \). Indeed, similarly to the one done in (16),

\[
\mathbb{E}\left[ I_{\{Y \leq ((x - \mu)/\alpha \sigma)^2\}} \Phi(\pm \alpha \sqrt{Y}) \right] = \frac{1}{2} \cdot T \left( \frac{x - \mu}{\sigma}; \alpha, p \right) + \frac{1}{\Gamma(p)} I \left( \left| \frac{x - \mu}{\alpha \sigma} \right| \pm \frac{\alpha}{\sqrt{2}}, 2p - 1, 1 \right),
\]

(19)

where \( I(u; a, b, c) = \int_0^u \text{erf}(ax)x^{b}e^{-c^2x^2} \, dx \), \( u > 0, a \in \mathbb{R}, b > 0, c > 0 \). Since \( |\text{erf}(x)| \leq 1 \), we get \( |I(u; a, b, c)| \leq I(u; 0, b, c) < \infty \), and therefore, \( I(u; a, b, c) \) always exists. In general, closed form solutions for the definite integral \( I(u; a, b, c) \) are not available in terms of commonly used functions.

Replacing the formulas (17) and (19) in (4), we obtain the following closed formula for the CDF of \( X \sim TD_\Phi(\theta) \):

\[
F(x; \theta) = \begin{cases} 
\sigma \left[ \rho + \delta T \left( \frac{x - \mu}{\sigma}; \alpha, p \right) \right] \Phi \left( \frac{x - \mu}{\sigma} \right) \\
+\frac{\delta \sigma}{Z_\theta} \left[ \frac{1}{2} - \frac{\alpha \Gamma(p + 12)}{\sqrt{2\pi} \Gamma(p)} \cdot 2F_1 \left( \frac{1}{2}, p + \frac{3}{2}; \frac{-\alpha^2}{2} \right) \right] \\
-\frac{\delta \sigma}{Z_\theta} \left[ \frac{1}{2} T \left( \frac{x - \mu}{\sigma}; \alpha, p \right) - \frac{1}{\Gamma(p)} I \left( \left\| \frac{x - \mu}{\alpha \sigma} \right\| \pm \frac{\alpha}{\sqrt{2}}, 2p - 1, 1 \right) \right] & \text{if } x < \mu, \\
\sigma \left[ \rho + \delta T \left( \frac{x - \mu}{\sigma}; \alpha, p \right) \right] \Phi \left( \frac{x - \mu}{\sigma} \right) \\
+\frac{\delta \sigma}{Z_\theta} \left[ \frac{1}{2} - \frac{\alpha \Gamma(p + 1/2)}{\sqrt{2\pi} \Gamma(p)} \cdot 2F_1 \left( \frac{1}{2}, p + \frac{3}{2}; \frac{-\alpha^2}{2} \right) \right] \\
-\frac{\delta \sigma}{Z_\theta} \left[ \frac{1}{2} T \left( \frac{x - \mu}{\sigma}; \alpha, p \right) + \frac{1}{\Gamma(p)} I \left( \left\| \frac{x - \mu}{\alpha \sigma} \right\| \pm \frac{\alpha}{\sqrt{2}}, 2p - 1, 1 \right) \right] & \text{if } x \geq \mu,
\end{cases}
\]

where \( Z_\theta \) and \( T \) are as in (18) and (2), respectively.

Remark 4.2: As expected, since \( \Phi \) has a symmetric distribution around 0, \( F(\mu; \theta) = 1/2 \) and then, \( Q_2 = \mu \) is the median and the mean for \( X \sim TD_\Phi(\theta) \).

Remark 4.3: Some examples where \( I((x - \mu)/\alpha \sigma; \pm (\alpha/\sqrt{2}), 2p - 1, 1) \) in (19) admits a closed form are:
(1) By taking $\alpha = \sqrt{2}$ and by using the following formula (see Item 4 in Subsection 1.5.3, p. 31, of Ref. [30]):

$$
\int_0^\infty \text{erf}(ax)x^\lambda e^{-a^2x^2} \, dx = \frac{2a}{\sqrt{\pi}(\lambda + 2)} u^{\lambda+2} 2F_2 \left( 1, \frac{\lambda}{2} + 1; \frac{3}{2}, \frac{\lambda}{2} + a^2 u^2 \right), \quad \lambda > -2,
$$

where $2F_2(a_1,a_2;b_1,b_2;x)$ is the generalized hypergeometric, we have

$$
I \left( \frac{x - \mu}{\alpha \sigma}; \pm \frac{\alpha}{\sqrt{2}}, 2p - 1, 1 \right) = \frac{2}{\sqrt{\pi}(2p + 1)} \left( \frac{x - \mu}{\alpha \sigma} \right)^{2p+1} 2F_2 \left( 1, p + \frac{1}{2}; \frac{3}{2}, p + \frac{3}{2}; \left( \frac{\mu - x}{\alpha \sigma} \right)^2 \right).
$$

(2) By taking $\alpha = \sqrt{2}$ and $p = 3/2$,

$$
I \left( \frac{x - \mu}{\alpha \sigma}; \pm \frac{\alpha}{\sqrt{2}}, 2p - 1, 1 \right) = \frac{1}{2} \left( \frac{x - \mu}{\alpha \sigma} \right) \text{erf} \left( \frac{x - \mu}{\alpha \sigma} \right) \exp \left[ - \left( \frac{x - \mu}{\alpha \sigma} \right)^2 \right] 
$$

$$
\pm \frac{1}{4\sqrt{\pi}} \left\{ 1 - \exp \left[ -2 \left( \frac{x - \mu}{\alpha \sigma} \right)^2 \right] \right\} \pm \frac{\sqrt{\pi}}{8} \left[ \text{erf} \left( \frac{x - \mu}{\alpha \sigma} \right) \right]^2.
$$

### 4.4. Moments and Shannon entropy

The existence of moments and entropy functions can guarantee to use the maximum log$_q$-likelihood estimation (ML$q$E) and maximum likelihood estimation (MLE) for the estimation of parameters of the proposed distribution [7,25,36]. A routine calculation shows that the moments of $X \sim \text{TD}_\Phi(\theta)$ can be written as (for more details, see Supplementary Material II of this work)

$$
\mathbb{E}(X^n) = \frac{(\rho + \delta) \sigma}{Z_\theta} \sum_{k=0}^n \binom{n}{k} \mu^{n-k} \sigma^k \frac{2^{-k/2} k!}{(k/2)!} 1_{[k \text{ even}]}
$$

$$
+ \frac{\delta \sigma}{Z_\theta \sqrt{2\pi} \Gamma(p)} \sum_{k=0}^n \binom{n}{k} \mu^{n-k} \sigma^k \left[ 1 - (-1)^{k-1} \right] \sum_{m=0}^{[k/2]} \sum_{j=0}^{[m/2]} \frac{c_{k,m} d_{k,m,j} \alpha^{m-2j}}{(1 + \alpha^2)^{(m-2j)/2} + p} \Gamma \left( \frac{k_m - 2j}{2} + p \right),
$$

where $p \geq 1$ is an integer, $k_m = k - 2m - 1$, $Z_\theta$ is as in (18), $c_{n,m}$ and $d_{n,m,j}$ are constants explicitly defined in the Supplementary Material II. In particular, $\mathbb{E}(X) = \mu$ and

$$
\text{Var}(X) = (\mu^2 + \sigma^2) \left[ \frac{\sigma (\rho + \delta)}{Z_\theta} - \frac{2\alpha \sigma \delta \Gamma(p + (1/2))}{\sqrt{2\pi} \Gamma(p) Z_\theta} \right] 2F_1 \left( 1/2, p + 1/2; 3/2; -\frac{\alpha^2}{2} \right) - \mu^2.
$$

As a consequence, from Corollary 3.5, the closed expressions for skewness and kurtosis of random variable $X \sim \text{TD}_\Phi(\theta)$ are easily obtained.
On the other hand, a formula for the Shannon entropy is given by (for more details, see Supplementary Material III of this paper)

\[
S_1(X) = \log(Z_\theta) - \frac{2(\rho + \delta)\sigma}{Z_\theta} \sum_{k=0}^{\infty} \frac{1}{(2k+1)} \left[ 1 + \frac{1}{(\rho + 1)2^{2k+1}} \sum_{i=p}^{\infty} \tilde{c}_{i,k} \frac{2^{-i}(2i)!}{i!} \right]
\]

\[
+ \frac{2\delta\sigma}{Z_\theta} \sum_{k=0}^{\infty} \frac{1}{(2k+1)} \left[ 2E[X^{2i}/Phi1(\theta)] - 1 + \frac{1}{(\rho + 1)2^{2k+1}} \sum_{i=p}^{\infty} \tilde{c}_{i,k} \frac{2^{-i}(2i)!}{i!} \right]
\]

\[
+ \frac{(\rho + \delta)\sigma}{Z_\theta} \left[ \log(\sqrt{2\pi}) + \frac{1}{2} \right] + \delta\sigma \left\{ \frac{\alpha \Gamma(p + (1/2))}{\sqrt{2\pi} \Gamma(p)} \left( \frac{1}{1 + (\alpha^2/2)^p + (1/2)} \right) \right.
\]

\[
\left. - \left[ \log(\sqrt{2\pi}) + \frac{1}{2} \right] \left( 2E[X^{2i}]/Phi1(\theta) \right) \right] ,
\]

where \(Z_\theta\) is as in (18), the coefficients \(\tilde{c}_{i,k}\)'s are determined in the Supplementary Material III of this paper, \(E[X^{2i}/Phi1(\theta)]\) is given in (17) and \(E[I_{\{X \leq \alpha\sqrt{Y}\}}Z^{2i}] = E[I_{\{X \leq \alpha\sqrt{Y}\}}Z^{2i}] - E[I_{\{X \leq -\alpha\sqrt{Y}\}}Z^{2i}]\).

**4.5. Rate of the random variable \(X \sim TD_\phi(\theta)\)**

Following Ref. [23], the rate of a continuous random variable \(X\) is given by

\[
\tau_X = - \lim_{x \to \infty} \frac{d \log f_X(x)}{dx},
\]

where \(f_X(x)\) denotes its respective PDF.

A simple computation shows that the rate of \(X \sim TD_\phi(\theta)\) is

\[
\tau_{TD_\phi(\theta)} = \frac{1}{\sigma} \lim_{x \to \infty} \left[ \frac{x - \mu}{\sigma} - \frac{2}{\alpha \Gamma(p)} \left( \frac{x - \mu}{\alpha \sigma} \right)^{2p-1} e^{-((x-\mu)/\alpha \sigma)^2} \rho + T \left( \frac{x - \mu}{\sigma}; \alpha, p \right) \right]
\]

\[
= \infty.
\]

Then, far enough out in the tail, the distribution of \(X \sim TD_\phi(\theta)\) looks like a Normal distribution, as expected. In addition, we have some comparisons between the rate of \(X \sim TD_\phi(\theta)\) with the rates of some random variables with known distributions in the literature: Inverse-gamma, Log-normal, Generalized-Pareto, BGumbel [28], BWeibull [36], BGamma [37], exponential and Normal:

\[
\tau_{InvGamma}(\alpha, \beta) = \tau_{LogNorm}(\mu, \sigma^2) = \tau_{GenPareto}(\alpha, \beta, \delta) = \tau_{BWeibull}(\alpha < 1, \beta, \delta) = 0
\]

\[< \tau_{BGumbel}(\mu, \beta, \delta) = \tau_{BWeibull}(\alpha = 1, \beta, \delta) = \tau_{BGamma}(\alpha, 1/\beta, \delta) = \tau_{exp}(1/\beta) = 1/\beta
\]

\[< \tau_{BWeibull}(\alpha > 1, \beta, \delta) = \tau_{TD_\phi(\theta)} = \tau_{Normal}(\mu, \sigma^2) = \infty.
\]

In other words, the tail of the normal distribution of \(X \sim TD_\phi(\theta)\) and of BWeibull(\(\alpha > 1, \beta, \delta\)) are lighter than the tail of the other distributions specified above.
5. The model arising from a generalized mixture

The weighted distributions of random variable \((W - \mu)/\sigma\), with weight function \(w_k\), have its PDF defined by

\[
f_k(x; \mu, \sigma) = \frac{1}{\int_D w_k(y) g(y) dy} w_k \left( \frac{x - \mu}{\sigma} \right) g \left( \frac{x - \mu}{\sigma} \right), \quad \frac{x - \mu}{\sigma} \in D, \quad k = 0, 1, \ldots
\]

By using the powers series expansion of the incomplete gamma function, note that the PDF \(f(x; \theta)\) in (1) interprets as an infinite (generalized) mixture of weighted distributions of \((W - \mu)/\sigma\) with the weight functions \(w_k(y) = y^{2k}, k = 0, 1, \ldots\), and with same parameter vector \((\mu, \sigma)\). That is,

\[
f(x; \theta) = c_0 f_0(x; \mu, \sigma) + \sum_{k=p}^{\infty} c_k f_k(x; \mu, \sigma), \quad (20)
\]

where the constants \(c_0, c_k\), with \(k = p, p + 1, \ldots\), that depends only on \(\theta\), respectively, are given by

\[
c_0 = \frac{\rho \sigma}{Z_\theta}; \quad c_k = \frac{\delta \sigma}{\Gamma(p) Z_\theta} \left( -1 \right)^{k-p} \frac{\mathbb{E}(W^{2k})}{k(k-p)! \alpha^{2k}}, \quad k = p, p + 1, \ldots
\]

It is straightforward to verify that \(c_0 + \sum_{k=p}^{\infty} c_k = 1\). Note that the weights \(c_k\) can take on negative values.

If \(X \sim \text{TD}_\phi(\theta)\) and \(L : \mathbb{R} \to \mathbb{R}\) is a Borel-measurable function, then, by representation (20), the expectation of truncated random variable \(1_{\{X \leq b\}} L(X)\), with \(b \in \mathbb{R}\), is given by

\[
\mathbb{E} \left[ 1_{\{X \leq b\}} L(X) \right] = c_0 \mathbb{E}_0 \left[ 1_{\{X \leq b\}} L(X) \right] + \sum_{k=p}^{\infty} c_k \mathbb{E}_k \left[ 1_{\{X \leq b\}} L(X) \right].
\]

But this representation as infinite sum is not as informative as that representation of Theorem 3.4. For this reason in the paper, we had been concerned with providing mathematical representations that allow us to find the closed expressions for some characteristics of the distribution (such as normalizing factor, CDF, moments and entropies) as a function of the known mathematical functions.

6. Inference

The ML\(_q\)E and Fisher information (FI) matrix for this estimation method are introduced. ML\(_q\)E drops to the MLE if \(q \to 1\) [8,18,19,35].

6.1. The \(\log_q\)-likelihood function

For a random sample \(X_1, \ldots, X_n\) from the random variable \(X \sim \text{TD}(\theta)\), with parameter vector \(\theta = (\mu, \sigma, \alpha, \rho, \delta)\), let us suppose that \(x_1, \ldots, x_n\) are the observed values of
The definition of FL based on $\log_q$ is given in Refs. [8, 29]. The elements of FI matrix are defined by

$$[I(\theta)]_{j,k} = \sum_{i=1}^{n} \mathbb{E} \left[ f^{1-q}(X_i; \theta) \frac{\partial \log f(X_i; \theta)}{\partial \theta_j} \frac{\partial \log f(X_i; \theta)}{\partial \theta_k} \right], \quad \theta_j, \theta_k \in \{\mu, \sigma, \alpha, \rho, \delta\},$$

where $\log f(x; \theta)$ is given in (22) and $f(x; \theta)$ is the parametric model in (1). When the inverse of FI matrix exists, it is well known that the diagonal elements of inverse of FI give $\text{Var}(\hat{\theta})$ based on $\log_q$ [3, 8]. In general, there is no closed form expression for the FI matrix (see Supplementary Material IV and Subsection 6.6 of Supplementary Material VI). It is clear that when $q = 1$ in the above identity, under standard regularity conditions, we obtain the classical FI matrix [19, 25].

7. Application on real data sets

We present applications to illustrate the performance of the trimodal normal model compared with smooth kernel distribution as semiparametric distribution (SmoothKernel Distribution) and estimation of distribution which is performed using a function named as FindDistribution embedded into Mathematica 12.0 software to find
an appropriate distribution for the data set. Since Mathematica software is capable for performing the optimization, bootstrap and also includes the numerical evaluation of hypergeometric function \((_{2}F_{1})\) while conducting the modelling the data sets, practitioners can use these codes for their aims in researches. The supplementary materials provide the building codes for practitioners \([1,17,26]\).

Examples 1 and 2 represent the real data called as heterodatatain$V5 and heterodata$V4, respectively, in the ‘Rmixmod’ package at R software with version 4.1.3. The numbers of sample size \(n\) are 300 and 200 for Examples 1 and 2, respectively. The package ‘multilevel’ includes data called as ‘bh1996’. The columns 11 and 13 of bh1996 data represent the modality. Examples 3 and 4 are for data sets with sample sizes \(n = 7382\). The salary and campus police can show the trimodality behaviour. When it is considered the experimental nature of salary and police reality as phenomena on the universe or world, it is reasonable to observe the fluctuation for the economical and secure and non-secure events in the experiment. Examples 5 and 6 representing salary and campus police provided by ‘wooldridge’ package with ceosal2—lsalary and campus—lpolice data sets in R, respectively, have the sample sizes as 177 and 97, respectively. The \(q\) values for Examples 1–6 are chosen as 0.98,0.95,0.98,0.99,0.98, \(1 − 10^{-10}\), respectively, according to the goodness of fit tests (GOFs).

The location \((\mu)\) and scale \((\sigma)\) are important parameters to summarize the data set. The efficient estimations of these parameters depend on the chosen function used for modelling. Table 2 provides them and other statistics for testing the modelling competence of the used functions. When all of PM (\(^{q}\)TD\(_{\Phi}\), SK, EstD and TD\(_{\Phi}\) ) in Table 2 are compared, SK and EstD are rival ones among models. On the other side, the statistics and information criteria of \(^{q}\)TD\(_{\Phi}\) and TD\(_{\Phi}\) can be near to SK and EstD as alternative models (see also discussion in Section 6.3 of Supplementary Material VI). Table 3 introduces the basic statistics to see the role of distribution with one mode property and trimodality in Table 2. Since EstD and SK are based on the distributions as approach from the parametric and the non-parametric cases, respectively, it is necessary to know the existence of moments and entropy functions which are root of the estimation methods such as MLqE and MLE \([3,25]\), the results show that there is an infinity case in the numerical results of EstD in Table 2. Such a result means that the theoretical properties of the proposed parametric function are important tools which can guarantee to use the function for modelling the data set. Furthermore, existing the analytical expression for CDF is also important instead of using the numerical integration rule for getting the values of CDF (see Ref. [10]). From such results, it is reasonable to imply that even if the data set cannot show an exact pattern in the trimodality, it is necessary to have a form from trimodality case to perform a modelling at each time needed. Since the real data sets have eventually finite sample size, it is reasonable to expect that the nature of fluctuation cannot be strictly same pattern given by parametric model such as TD\(_{\Phi}\). In the general setting, since it is observed that SK which accommodates the fluctuation well shows the better performance when compared with TD\(_{\Phi}\), the role of TD\(_{\Phi}\) in modelling the data set as well is observed and can be concluded that the fluctuation in three modes can be observed more exactly and well-fitted by means of TD\(_{\Phi}\) if the sample size can get more larger values; however, since SK is a non-parametric approach, it is generally logical to expect that it will be rival one at each time while performing to fit the data set eventually well or not-well. On the other side, note that the efficient fitting is an important issue; because when looking at the results of \(\hat{\sigma}\) from \(^{q}\)TD\(_{\Phi}\) and TD\(_{\Phi}\),
Thus, the estimation of scale parameter representing the variability can be determined well.

**Table 2.** The models, the estimates of parameters, statistics and information criteria for assessment of models.

| PM       | Estimates | GOFTs | Log(L) and information criteria |
|----------|-----------|-------|---------------------------------|
|          | $\hat{\mu}$ | $\hat{\sigma}$ | KS | CVM | AD | Log(L) | AIC | BIC |
| Example 1 | $9^{*}$TD$_{0}$ | $-2.25117$ | $1.32990$ | | | | | |
| SK       | $-2.24486$ | $1.62363$ | | | | | | |
| EstD     | $-2.22717$ | $1.61703$ | | | | | | |
| TD$_{0}$ | $-2.25803$ | $1.38998$ | | | | | | |
| Example 2 | $9^{*}$TD$_{0}$ | $-1.94546$ | $1.29762$ | | | | | |
| SK       | $-1.98323$ | $1.63448$ | | | | | | |
| EstD     | $-1.96097$ | $1.65964$ | | | | | | |
| TD$_{0}$ | $-1.91852$ | $1.47189$ | | | | | | |
| Example 3 | $9^{*}$TD$_{0}$ | $2.77324$ | $0.856429$ | | | | | |
| SK       | $2.78055$ | $0.915343$ | | | | | | |
| EstD     | $2.78047$ | $0.909687$ | | | | | | |
| TD$_{0}$ | $2.77950$ | $0.882003$ | | | | | | |
| Example 4 | $9^{*}$TD$_{0}$ | $-0.001341870$ | $0.865658$ | | | | | |
| SK       | $-0.000056456$ | $0.889187$ | | | | | | |
| EstD     | $0.000581656$ | $0.884030$ | | | | | | |
| TD$_{0}$ | $-0.002632980$ | $0.869382$ | | | | | | |
| Example 5 | $9^{*}$TD$_{0}$ | $6.59421$ | $0.512779$ | | | | | |
| SK       | $6.58597$ | $0.621550$ | | | | | | |
| EstD     | $6.58414$ | $0.629930$ | | | | | | |
| TD$_{0}$ | $6.58931$ | $0.535466$ | | | | | | |
| Example 6 | $9^{*}$TD$_{0}$ | $2.73408$ | $0.696568$ | | | | | |
| SK       | $2.73093$ | $0.829185$ | | | | | | |
| EstD     | $2.73345$ | $0.845120$ | | | | | | |
| TD$_{0}$ | $2.73054$ | $0.690970$ | | | | | | |

Notes: PM: (semi)-parametric models; $9^{*}$TD$_{0}$: objective function log$_{q}$ (MLqE) from trimodal normal distribution; SK: the smooth kernel distribution based on Gaussian (normal) distribution (semiparametric model); EstD: the automatically chosen function by 'FindDistribution' in Mathematica software; TD$_{0}$: objective function log$_{q}$ (MLE) from trimodal normal distribution; KS: Kolmogorov-Smirnov; CVM: Cramér-von Mises; AD: Anderson-Darling; AIC: Akaike information criterion; BIC: Bayesian information criterion.

Italic represents closeness to the values produced by SK and EstD or almost best ones.

Bold represents the smallest values of statistics of GOFTs, Log(L) and information criterion.

**Table 3.** The estimates for $\hat{\mu}$ and $\hat{\sigma}$ from normal distribution and robust form.

| Example 1 | Example 2 | Example 3 | Example 4 | Example 5 | Example 6 |
|-----------|-----------|-----------|-----------|-----------|-----------|
| N         | $\hat{\mu}$ | $-2.24483$ | $-1.96589$ | $2.77989$ | $1.24627 \times 10^{-12}$ | $6.58483$ | $2.727990$ |
|           | $\hat{\sigma}$ | $1.58461$ | $1.58535$ | $0.908344$ | $0.882701$ | $0.601046$ | $0.793457$ |
| R         | Median | $-2.34344$ | $-2.23693$ | $2.83333$ | $0.0309632$ | $6.56997$ | $2.760020$ |
|           | MAD | $1.08095$ | $0.981103$ | $0.611111$ | $0.607600$ | $0.43837$ | $0.578178$ |

Note: N: normal distribution; R: robust statistics; M: median; MAD: median absolute deviation computed by Median(|x-Median(x)|).

the values of $\hat{\sigma}$ from all of examples are smaller than the estimates of $\hat{\sigma}$ from SK and EstD. Thus, the estimation of scale parameter representing the variability can be determined well and this estimate of $\hat{\sigma}$ can be used when the artificial data sets generated will be used to make a comparison between the real data and the artificial data in order to test whether or
not a system works safely, correctly, etc. Even if the modelling performance for the finite sample size can be observed for the SK and EsTD which can give similar results to $^q \text{TD}_\Phi$ and $\text{TD}_\Phi$ cases, the results of Example 6 for EstD can give the infinite results because of the theoretical properties such as finiteness of moment or entropy functions for the mixed distributions used by EstD in Mathematica software could not be satisfied.

The estimates of $\hat{\mu}$ and $\hat{\sigma}$ from different PM show that we can have a clue to imply that the existence of modality can be observed, because the estimates of $\hat{\mu}$ and $\hat{\sigma}$ from SK as a smooth kernel technique based on working on the data-adaptive approach (which is capable to fit the modality whether or not it exists in the reality—see also discussions on Section 5.2 of Supplementary Material V) instead of parametric approach for modelling can be close to the estimates of $\hat{\mu}$ and $\hat{\sigma}$ from $^q \text{TD}_\Phi$ and $\text{TD}_\Phi$. The kernel estimation method as a smoothing technique is the best one generally. Even if 1000 replication for the different design of samplings constructed by use of bootstrap technique is applied, the numerical error(s) in computation for optimization can be trick to consider and make an accurate judgement among the modelling performance of the used four models. For example, CDF and PDF of $\text{TD}_\Phi$ depend on the Hypergeometric2F1 in Mathematica. Soft forms of PDF of $\text{TD}_\Phi$ in Figure 2(b) and smooth kernel technique in Ref. [21] can be alternative to each other when the estimates of $\hat{\mu}$ and $\hat{\sigma}$, the statistics from GOFTs, the values of log(L) and IC are taken into account. On the other side, it is very difficult to know which function will be the best one for modelling when the data sets in the finite sample size are tried to be fitted by the functions. Even if the sample sizes of Examples 3 and 4 are 7382, eventually we have finite sample size whatever it is. The population in reality will not known exactly. Consequently, an alternative function can be necessary for driving modality via parameters $\rho$ and $\delta$. This is the reason why we make a comparison between the SK and other parametric models to observe what and how the estimates of $\hat{\mu}$ and $\hat{\sigma}$ will be changed if PDF are changed [22]. Note that the modelling and numerical error(s) are topics which can affect each other.

8. Conclusions

Recent times show that an increasing popularity has been observed in the modelling for data sets having modality, producing the trimodal form of any PDF has been proposed. The trimodal form is constructed using the technique which includes the cumulative function of Maxwell distribution, the existing unimodal distribution and the corresponding normalizing constant of the proposed distribution. The properties of trimodality have been examined. The application of producing the trimodality has been conducted for the normal distribution which is symmetric and unimodal form with two parameters which are location and scale. The properties of trimodal normal distribution have been examined. Thus, the applicability of this distribution have been tested. The trimodal normal distribution can have different forms such as strict and soft modalities to perform a precise fitting when there exist three modes in the empirical distribution of the data sets.

A comparison among trimodal normal, the kernel type estimation method and the probable parametric distribution driven by Mathematica software has been performed in order to make applications for numerical evaluation of $\text{TD}_\Phi$. The log$_q$-likelihood estimation method and its special form with $q \to 1$ have been used to estimate the parameters of trimodal normal distribution. The proofs, properties of $\text{TD}_\Phi$ distribution and codes used
for application have been given by Supplementary Materials if the researchers perform to model the data sets by use of TDΦ distribution.

The future will be an application on the different areas of statistics such as regression modelling, the tools in the multivariate statistics and other tools based on the distribution theory. The order statistic form of TDΦ in the least informative distribution will be studied for the trimodal forms of the existing distributions in the applied field of science. Additionally, the precise modelling for inlier into data sets can also be performed by use of trimodality, the generalized logarithms, entropy functions, order statistic and different estimation methods all together [5,6]. A package in R software will be prepared for practitioners after the special functions in R software are improved.

Acknowledgments

We acknowledge the anonymous referees for their helpful comments, suggestions and references provided in their reports. R. V. thanks A. V. Medino, J. Roldan and E. M. M. Ortega for partial discussions of Theorem 3.4 and for general paper questions.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES)—Finance Code 001.

ORCID

Roberto Vila http://orcid.org/0000-0003-1073-0114
Victor Serra http://orcid.org/0000-0003-3061-2094
Mehmet Niyazi Çankaya http://orcid.org/0000-0002-2933-857X
Felipe Quintino http://orcid.org/0000-0003-0286-0541

References

[1] J.A. Baglivo, Mathematica Laboratories for Mathematical Statistics: Emphasizing Simulation and Computer Intensive Methods, Society for Industrial and Applied Mathematics, Pennsylvania, 2005.
[2] J.F. Bercher, A simple probabilistic construction yielding generalized entropies and divergences, escort distributions and q-Gaussians, Phys. A Stat. Mech. Appl. 391 (2012), pp. 4460–4469.
[3] J.F. Bercher, Some properties of generalized Fisher information in the context of nonextensive thermostatistics, Phys. A Stat. Mech. Appl. 392 (2013), pp. 3140–3154.
[4] M.N. Çankaya, Asymmetric bimodal exponential power distribution on the real line, Entropy 20 (2018), p. 23.
[5] M.N. Çankaya, M-estimations of shape and scale parameters by order statistics in least informative distributions on q-deformed logarithm, J. Inst. Sci. Technol. 10 (2020), pp. 1984–1996.
[6] M.N. Çankaya, Derivatives by ratio principle for q-sets on the time scale calculus, Fractals 29 (2021), p. 2140040. 10.1142/S0218348X21400405
[7] M.N. Çankaya and O. Arslan, On the robustness properties for maximum likelihood estimators of parameters in exponential power and generalized T distributions, Commun. Stat. Theory Methods 49 (2020), pp. 607–630.
[8] M.N. Çankaya and J. Korbel, *Least informative distributions in maximum q-log-likelihood estimation*, Physica A. 509 (2018), pp. 140–150.

[9] M.N. Çankaya, Y.M. Bulut, F.Z. Doğru, and O. Arslan, *A bimodal extension of the generalized gamma distribution*, Rev. Colomb. Estad. 38 (2015), pp. 371–384.

[10] M.N. Çankaya, A. Yalçınkaya, Ö. Altındağ, and O. Arslan, *On the robustness of an epsilon skew extension for Burr III distribution on the real line*, Comput. Stat. 34 (2019), pp. 1247–1273.

[11] F.Z. Doğru, Y.M. Bulut, and O. Arslan, *Doubly reweighted estimators for the parameters of the multivariate t-distribution*, Commun. Stat. Theory Methods 47 (2018), pp. 4751–4771.

[12] F. Domma, B.V. Popović, and S. Nadarajah, *An extension of Azzalini’s method*, J. Comput. Appl. Math. 278 (2015), pp. 37–47.

[13] F. Domma, F. Condino, and B.V. Popović, *A new generalized weighted Weibull distribution with decreasing, increasing, upside-down bathtub, N-shape and M-shape hazard rate*, J. Appl. Stat. 44 (2017), pp. 2978–2993.

[14] R.C. Dunbar, *Deriving the Maxwell distribution*, J. Chem. Edu. 59 (1982), p. 22.

[15] D. Elal-Olivero, *Alpha-skew-normal distribution*, Proyecciones (Antofagasta) 29 (2010), pp. 224–240.

[16] B.S. Everitt and D.J. Hand, *Finite Mixture Distributions*, Springer Science & Business Media, New York, 2013.

[17] J.F. Feagin, *Quantum Methods with Mathematica®*, Springer Science & Business Media, New York, 2002.

[18] D. Ferrari and Y. Yang, *Maximum Lq-likelihood estimation*, Ann. Stat. 38 (2010), pp. 753–783.

[19] R.A. Fisher, *Theory of Statistical Estimation*, Mathematical Proceedings of the Cambridge Philosophical Society, Vol. 22 (5), Cambridge University Press, 1925, pp. 700–725.

[20] E. Gómez-Déniz, J.M. Sarabia, and E. Calderín-Ojeda, *Bimodal normal distribution: extensions and applications*, J. Comput. Appl. Math. 388 (2021), pp. 113292.

[21] W. Härdle, M. Müller, S. Sperlich, and A. Werwatz, *Nonparametric and Semiparametric Models* (Vol. 1), Springer, Berlin, 2004.

[22] P.J. Huber, *Robust estimation of a location parameter*, Ann. Math. Stat. 35 (1964), pp. 73–101.

[23] S. Klugman, H. Panjer, and G. Willmot, *Loss Models: From Data to Decisions*, Wiley, New York, 1998.

[24] C. Lee, F. Famoye, and A.Y. Alzaatreh, *Methods for generating families of univariate continuous distributions in the recent decades*, Wiley Interdiscipl. Rev. Comput. Stat. 5 (2013), pp. 219–238.

[25] E.L. Lehmann and G. Casella, *Theory of Point Estimation*, Vol. 589, Wadsworth & Brooks/Cole, Pacific Grove, CA, 1998.

[26] J.G.S. León, *Mathematica® Beyond Mathematics: The Wolfram Language in the Real World*, Chapman and Hall/CRC, Boca Raton, 2017.

[27] E.W. Ng and M. Geller, *A table of integrals of the error functions*, J. Res. Natl. Bur. Stand. Sec. B Math. Sci. 73B (1969), pp. 1–20.

[28] C.E.G. Otiniano, R. Vila, P.C. Brom, and M. Bourguignon, *On the bimodal Gumbel model with application to environmental data*, Aust. J. Stat. 52 (2023), pp. 45–65.

[29] A. Plastino, A.R. Plastino, and H.G. Miller, *Tsallis nonextensive thermostatistics and Fisher’s information measure*, Physica A. 235 (1997), pp. 577–588.

[30] A.P. Prudnikov, I.U.A. Brychkov, and O.I. Marichev, *Integrals and Series. Vol 2. Special Functions*, Taylor & Francis, London, 2002.

[31] M. Rahman, B. Al-Zahrani, and M.Q. Shahbaz, *Cubic transmuted Pareto distribution*, Ann. Data Sci. 7 (2020), pp. 91–108.

[32] T.J. Rothenberg, *Identification in parametric models*, Econometrica 39 (1971), pp. 577–591.

[33] C.E. Shannon, *A mathematical theory of communication*, Bell Labs. Tech. J. 27 (1948), pp. 623–656.

[34] C. Tsallis, *Possible generalization of Boltzmann–Gibbs statistics*, J. Stat. Phys. 52 (1988), pp. 479–487.

[35] C. Tsallis, *Introduction to Nonextensive Statistical Mechanics: Approaching a Complex World*, Springer, New York, 2009.
[36] R. Vila and M.N. Çankaya, *A bimodal Weibull distribution: properties and inference*, J. Appl. Stat. 49 (2022), pp. 3044–3062.

[37] R. Vila, L. Ferreira, H. Saulo, F. Prataviera, and E.M.M. Ortega, *A bimodal gamma distribution: properties, regression model and applications*, Statistics 54 (2020), pp. 469–493.

[38] R. Vila, H. Saulo, and J. Roldan, *On some properties of the bimodal normal distribution and its bivariate version*, Chil. J. Stat. 12 (2021), pp. 125–144.

[39] R. Vila, L. Alfaia, A.F. Menezes, M.N. Çankaya, and M. Bourguignon, *A model for bimodal rates and proportions*, J. Appl. Stat. (2022), pp. 1–18. doi:10.1080/02664763.2022.2146661