Journal of Statistical Theory and Applications, Vol. 14, No. 1 (March 2015), 28-44

# A generalization of Tukey's $g-h$ family of distributions

J.A. Jiménez[a], V. Arunachalam[b] and G.M. Serna[c]

[a] *Department of Mathematics, Universidad Nacional de Colombia, Bogotá, Colombia.*
[b] *Department of Statistics, Universidad Nacional de Colombia, Bogotá, Colombia.*
[c] *Department of Business Studies, University of Alcalá de Henares, España.*
[a] *josajimenezm@unal.edu.co,*[b] *varunachalam@unal.edu.co,*[c] *gregorio.serna@uah.es*

Received 30 May 2013

Accepted 12 December 2014

*A new class of distribution function based on the symmetric densities is introduced, these transformations also produce nonnormal distributions and its* pdf *and* $cdf$ *can be expressed in parametric form. This class of distributions depend on the two parameters, namely g and h which controls the skewness and the elongation of the tails, respectively. This class of skewed distributions is a generalization of Tukey's* $g-h$ *family of distributions. In this paper, we calculate a closed form expression for the density and distribution of the Tukey's* $g-h$ *family of generalized distributions, which allows us to easily compute probabilities, moments and related measures.*

*Keywords*: *Tukey's* $g-h$ *family of distributions, generalized error distribution, Lambert's function, Fourier transform.*

MSA 2010: 60E05, 62E15

## 1. Introduction

On many occasions, statistical data show asymmetry, indicating some kind of skewness. This is of the case of actuarial and financial data, which have characteristic asymmetrically distributed structures with extreme values yielding heavier tails. For example, the probability distributions of financial asset returns are not normally distributions, but usually have asymmetry and leptokurtosis. The most important and useful characteristic of the Tukey's $g-h$ family of distributions is that it covers most of the pearsonian family of distributions, and also can generate several known distributions, for example lognormal, Cauchy, Exponential, Chi-squared (see Martínez & Iglewicz (1984)). Tukey's $g-h$ family of distributions has been used in the context of statistical, simulation studies that include such topics as financial markets Badrinath & Chatterjee (1988), Mills (1995), and Badrinath & Chatterjee (1991) have used the $g$ and $h$ to model the return on a stock index, also the return on shares in several markets. Dutta & Babbel (2004) showed that the skewed and leptokurtic behavior of *LIBOR* was modeled effectively using the distribution $g-h$. Dutta & Babbel (2005) used $g$ and $h$ to model interest rates and options on interest rates, while Dutta & Perry


Copyright: the authors

(2007) used the $g-h$ to estimate operational risk; Tang & Wu (2006) studied the portfolio management. Jiménez & Arunachalam (2011) provided the explicit expressions of skewness and kurtosis for *VaR* and *CVaR* calculations. They propose the use of Tukey's classical $g$ and $h$ transformations applied to the normal distribution to capture these distributional features.

In this paper, we propose a generalization of Tukey's $g-h$ family of distributions, when the standard normal variate is replaced by a continuous random variable $U$ with mean 0 and variance 1. The attraction of this family of distribution is that from a symmetric variate with probability density function $(pdf)$, a large class of distributions can be generated with the parameters $g$ and $h$ which controls the skewness and the elongation of the tails. This new class of distribution allows us to models with large kurtosis measures and will useful in financial and other application in asymmetrical distributions.

The paper is organized as follows: Section 2 presents the Tukey's $g-h$ family of generalized distributions. Section 3 presents its statistical properties: *pdf*, cumulative distribution function $(cdf)$, expressions for the $n$th moment and quantile-based measures of skewness and kurtosis are derived. Section 4 introduces very briefly the $g$ generalized distribution and its moments. Section 5 explains the adjustment methodology based on real data, i.e., we demonstrate how the $g-h$ can be used to simulate or model combined data sets when only the mean, variance, skew, and kurtosis associated with the underlying individual data sets are available. Finally, conclusion are presented.

## 2. Tukey's $g-h$ family of generalized distributions

Tukey (1977) introduced a family of distributions by two nonlinear transformations called the $g-h$ distributions, which is defined by

$$Y = T_{g,h}(Z) = \frac{1}{g}\left(\exp\{gZ\}-1\right)\exp\{hZ^2/2\} \qquad \text{with } g \neq 0\ , h \in \mathbb{R} \tag{2.1}$$

where the distribution of $Z$ is standard normal. When these transformations are applied to a continuous random variable normalized $U$ , i.e., with mean 0 and variance 1, such that its *pdf* $f_U(\cdot)$ is symmetric about the origin and *cdf* $F_U(\cdot)$, the transformation $T_{g,h}(U)$ is obtained, which henceforth will be termed Tukey's $g-h$ generalized distribution:

$$Y = T_{g,h}(U) = \frac{1}{g}\left(\exp\{gU\}-1\right)\exp\{hU^2/2\} \qquad \text{with } g \neq 0\ , h \in \mathbb{R}. \tag{2.2}$$

The parameters $g$ and $h$ represent the skewness and the elongation of the tails of the Tukey's $g-h$ generalized distribution, respectively.

In this paper, for $h \neq 0$, we assume that the random variable $U$ has a Generalized Error Distribution of parameter $\alpha$, denoted $U \sim GED(\alpha)$, with *pdf* given by

$$f_U(u,\alpha) = \frac{1}{2\lambda\Gamma(\alpha+1)}\exp\left\{-\left|\frac{u}{\lambda}\right|^{\frac{1}{\alpha}}\right\}, \qquad u \in \mathbb{R}, 0 < \alpha \leq 1, \tag{2.3}$$

where $\lambda = \sqrt{\frac{\Gamma(\alpha)}{\Gamma(3\alpha)}}$ and $\Gamma(\cdot)$ is the gamma function, $\alpha$ is a tail-thickness parameter. When $\alpha = \frac{1}{2}$ then $U \sim N(0,1)$ and when $\alpha = 1$ then $U \sim$ Laplace $\left(0, \frac{\sqrt{2}}{2}\right)$, which are symmetric with standardized skewness of zero and standardized kurtosis of 3 and 6, respectively. Also, we present for $h = 0$ five special cases of the Tukey's $g-h$ distributions, when $U \sim GED\left(\frac{1}{2}\right)$, $U \sim GED(1)$, $U \sim$ Logistic $\left(0, \frac{\sqrt{3}}{\pi}\right)$, the hyperbolic secant (HyperSec) and the hyperbolic cosecant (HyperCsc).

When we assume $h=0$ in (2.2) the Tukey's $g-h$ generalized distribution reduces to

$$T_{g,0}(U)=\frac{1}{g}(\exp(gU)-1) \tag{2.4}$$

which is said to be *Tukey's g generalized distribution*. When $U \sim GED\left(\frac{1}{2}\right)$ its distribution also known as the family of lognormal distributions, because they have a lengthening of the tails than the standard normal distribution and they are skewed as well.

Similarly, when $g$ goes to 0 the Tukey's $g-h$ generalized distribution is given by

$$T_{0,h}(U)=U\exp\{hU^2/2\} \tag{2.5}$$

known as the *Tukey's h generalized distribution*. This distribution has the characteristic of being symmetrical but with tails heavier than the distribution of a random variable $U$ with increasing value of the parameter $h$.

If we wish to model an arbitrary random variable $X$ using the transformation given in (2.2), we introduce two new parameters, $A$ (location) and $B$ (scale) and propose the following model

$$X=A+BY \qquad \text{with} \qquad Y=T_{g,h}(U). \tag{2.6}$$

We must estimate four parameters that satisfy either of the following relationships:

$$x_p=A+By_p, \qquad \text{and} \qquad x_{1-p}=A-B\exp\{-gu_p\}y_p. \tag{2.7}$$

where $p>0.5$ and $x_p$ is the $p-$th quantile of the random variable $X$, such that

$$x_p=\inf\{x|P[X\leq x]>p\}=\sup\{x|P[X<x]\leq p\}.$$

Quantile $p-$value is the median, quartiles, eighth digit. Hoaglin et al. (1985) refer to them as the letter values, respectively, for the $M$ (median), $F$ (fourths), $E$ (eighths), etc. The estimation of parameters of Tukey's $g-h$ family of generalized distributions can be obtained using the method of moments Majumder & Ali (2008) or with the method of quantiles proposed by Hoaglin (1985).

## 3. Statistical properties of the Tukey's $g-h$ family

In this section we discuss the statistical properties Tukey's $g-h$ family of generalized distributions.

### 3.1. *Density function*

In Jiménez (2004) using the inverse function theorem provides the following relation

$$\left(F_U^{-1}\right)'(F_U(u_p))=\frac{d}{dp}u_p=\frac{1}{F_U'(u_p)}=\frac{1}{f_U(u_p)} \tag{3.1}$$

where $p$ is the only number that satisfies $F_U(u_p)=p$ and $f_U(\cdot)$ is the *pdf* of the continuous random variable $U$. The *pdf* for the Tukey's $g-h$ generalized distribution is obtained by using the following result

$$t_{g,h}(y_p)=\frac{f_U(u_p)}{T_{g,h}'(u_p)} \qquad \text{whenever} \qquad |h|u_p\frac{e^{-gu_p}-1}{g}<1, \tag{3.2}$$

where $y_p$ and $u_p$ denote the $p-$th quantile of the transformation $Y=T_{g,h}(U)$ and the continuous random variable $U$, respectively. From equation (2.7) and using the expression (3.1) (Jiménez &

Martínez (2006)) obtained *the* pdf *for the random variable $X$* as follows:

$$f_X(x_p) = f_X(A + By_p) = \frac{1}{|B|} t_{g,h}(y_p). \tag{3.3}$$

The parameter $g$ controls the skewness with positive values of $g$ generate positive skewness and negative values generate negative skewness and $g = 0$ corresponds to symmetry.

### 3.2. *Cumulative distribution function*

We now proceed to find the *cdf* of the Tukey's $g - h$ family of generalized distributions, denote by $\mathrm{F}_{g,h}(y)$. The following equality can be easily verified :

$$\int_a^b t_{g,h}(u)\,du = \int_{T_{g,h}^{-1}(a)}^{T_{g,h}^{-1}(b)} f_U(v)\,du = F_U\left(T_{g,h}^{-1}(b)\right) - F_U\left(T_{g,h}^{-1}(a)\right), \tag{3.4}$$

where $T_{g,h}^{-1}(\cdot)$ is the inverse of the transformation given in (2.2) and $F_U(\cdot)$ is the *cdf* of the continuous random variable $U$.

There is no explicit form for the inverse of the transformation of $T_{g,h}(U)$. However we get the inverse transformation when $h = 0$ or $g = 0$ as given below,

- If $h = 0$ then $T_{g,0}(U)$ is given by (2.4) and

$$T_{g,0}^{-1}(y) = \frac{1}{g} \ln(1 + gy), \qquad gy > -1. \tag{3.5}$$

- If $g = 0$ then $T_{0,h}(U)$ is given by (2.5), it must be

$$hY^2 = h\left[T_{0,h}(U)\right]^2 = hU^2 \exp\left\{hU^2\right\}, \tag{3.6}$$

  the expression (3.6) is of the form $u = w\exp\{w\}$, where $w = W(z)$ is the Lambert's function. Then the solution of (3.6) is given by

$$hU^2 = W\left(hy^2\right) \qquad \Rightarrow \qquad T_{0,h}^{-1}(y) = \sqrt{\frac{1}{h} W\left(hy^2\right)}. \tag{3.7}$$

  The basic properties of the function $W(z)$ are given in Olver et al. (2010).

Though the inverse of the transformation of $T_{g,h}(U)$ cannot be evaluated analytically, it can be evaluated numerically.

### 3.3. *Measures of skewness and kurtosis*

Since the transformation given in (2.2) is simply a quantile-based distribution, we use quantile-based measures of skewness $(SK)$ and kurtosis $(KR)$. For $0.5 < p < 1$ the measure proposed by Hinkley

(1975) is given by[a]

$$SK_2(p) = \frac{UHS_p/LHS_p - 1}{UHS_p/LHS_p + 1} = \frac{\exp\{gu_p\}-1}{\exp\{gu_p\}+1} = \tanh\left\{\frac{g}{2}u_p\right\}, \tag{3.8}$$

where $UHS_p = x_p - x_{0.5}$ and $LHS_p = x_{0.5} - x_{1-p}$, denote the $p$-th *upper half-spread* and *lower half-spread*, respectively (Hoaglin et al. (1985)). Note that this expression only depends on the parameter $g$. For fixed $p$ one can have values of $SK_2(p)$ varying values of $g$ as is illustrated in Figure 1

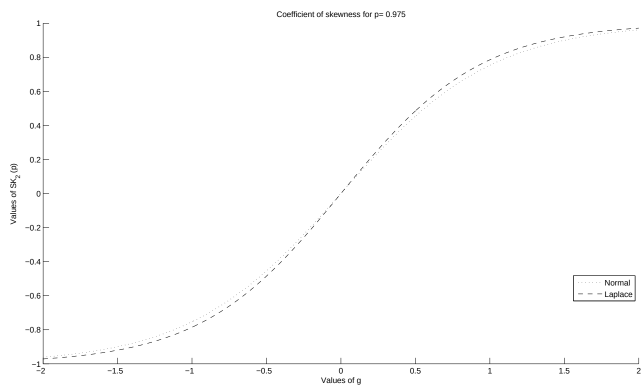



Fig. 1. Measure of skewness $SK_2(p)$

When $U \sim GED\left(\frac{1}{2}\right)$ we use the measure of skewness given in Groeneveld & Meeden (1984) to obtain

$$SK_3 = \frac{1-\exp\left\{-\frac{1}{2}\frac{g^2}{1-h}\right\}}{2\Phi\left(\frac{g}{\sqrt{1-h}}\right)-1} = \frac{1-\exp\left\{-\frac{1}{2}\frac{g^2}{1-h}\right\}}{\tanh\left\{\sqrt{\frac{2}{\pi}}\frac{g}{\sqrt{1-h}}\right\}}.$$

Here we use the expression given in Tocher (1964). Note that this last expression depends on two parameters $g,h$ which is zero when $g=0$. Also Groeneveld & Meeden (1984) present four properties that any reasonable coefficient of skewness must satisfy.

Furthermore, assuming that $U \sim GED\left(\frac{1}{2}\right)$ measure of kurtosis presented in Hogg (1974) we would read

$$KR_2(p;q) = \frac{\overline{U}_p - \overline{L}_p}{\overline{U}_q - \overline{L}_q},$$

where

$$\begin{aligned}\overline{U}_s - \overline{L}_s =& \frac{1}{s}\left[\mu_{g,h}\Phi(\delta_{2s}) + \frac{\Phi(\delta_{2s})-\Phi(\delta_{1s})}{(1-h)(\delta_{2s}-\delta_{1s})} - \mu_{g,h}\Phi(\delta_{2s}^*) + \frac{\Phi(\delta_{1s})-\Phi(\delta_{2s}^*)}{(1-h)\left(\delta_{1s}-\delta_{2s}^*\right)}\right] \\ =& \frac{\mu_{g,h}}{s}\left(\Phi(\delta_{2s})-\Phi(\delta_{2s}^*)\right) + \frac{1/s}{1-h}\left[\frac{\Phi(\delta_{2s})-\Phi(\delta_{1s})}{\delta_{2s}-\delta_{1s}} + \frac{\Phi(\delta_{1s})-\Phi(\delta_{2s}^*)}{\delta_{1s}-\delta_{2s}^*}\right],\end{aligned}$$

[a] $SK_1$ and $KR_1$ are the standardized values for skewness and kurtosis, respectively.

where

$$\delta_{1s} = \sqrt{1-h}z_s, \qquad \delta_{2s} = \delta_{1s} + \frac{g}{\sqrt{1-h}}, \qquad \delta_{2s}^{*} = \delta_{1s} - \frac{g}{\sqrt{1-h}}$$

Making use of the measure for kurtosis in Crow & Siddiqui (1967) for $p > q > 0.5$ we have

$$KR_3(p;q) = \begin{cases} \frac{\sinh(gu_p)}{\sinh(gu_q)} \exp\left\{\frac{h}{2}\left(u_p^2 - u_q^2\right)\right\} & \text{if } g \neq 0, \\ \frac{u_p}{u_q} \exp\left\{\frac{h}{2}\left(u_p^2 - u_q^2\right)\right\} & \text{if } g = 0. \end{cases}$$

### 3.4. *Moments of the Tukey's $g-h$ family of generalized distribution*

The next two propositions spell out the moments of the Tukey's $g-h$ family of generalized distributions. The corresponding proofs are given Appendix A.

**Proposition 3.1.** *The $m-$th power the Tukey's $g-h$ family of generalized distribution is given by*

$$Y^m = T_{g,h}^m(U) = \frac{m}{g^{m-1}} \sum_{k=0}^{m-1} (-1)^k \binom{m-1}{k} T_{\widetilde{g},\widetilde{h}}(U), \qquad m \geq 1, \tag{3.9}$$

*where $\widetilde{g} = (m-k)g$ and $\widetilde{h} = mh$.*

**Proposition 3.2.**

*Let $U$ be a continuous random variable with* pdf $f_U(u)$ *and* cdf $F_U(u)$. *If $F_U'(u)$ is never zero, then $F_U^{-1}(u)$ is differentiable and satisfies*

$$\mu_n' = \mathbb{E}(U^n) = \int_{-\infty}^{\infty} w^n f_U(w)\, dw = \int_0^1 \left[F_U^{-1}(q)\right]^n dq, \tag{3.10}$$

*where $q$ is the unique value that satisfies $F_U(u_q) = q$.*

**Proposition 3.3.**

*Let $Y = T_{g,h}(U)$ be transformation given in* (2.2)*, then the $n-$th moments of the random variable $Y$ are given by*

$$\mu_n' = \begin{cases} \frac{2}{g^n} \sum\limits_{k=0}^{n} (-1)^k \binom{n}{k} \int\limits_0^{\infty} \cosh(\widetilde{g}u) \exp\left\{\frac{1}{2}\widetilde{h}u^2\right\} f_U(u)du & \text{if } g \neq 0, \\ \left[1 + (-1)^n\right] \int\limits_0^{\infty} u^n \exp\left\{\frac{1}{2}\widetilde{h}u^2\right\} f_U(u)du & \text{if } g = 0, \end{cases} \tag{3.11}$$

*where $\widetilde{g} = (n-k)g$ and $\widetilde{h} = nh$.*

**Proof.** Using the expression (3.10) when $g \neq 0$ we obtain

$$\mathbb{E}(Y^n) = \int_0^1 Y_q^n dq = \int_{-\infty}^{\infty} y^n t_{g,h}(y)\, dy = \int_{-\infty}^{\infty} y^n \frac{f_U\left(T_{g,h}^{-1}(y)\right)}{T_{g,h}'\left(T_{g,h}^{-1}(y)\right)} dy.$$

Making the following change of variable

$$u = T_{g,h}^{-1}(y) \qquad du = \frac{dy}{T_{g,h}'\left(T_{g,h}^{-1}(y)\right)}, \tag{3.12}$$

and using the expression (3.9) we have

$$\begin{aligned}\mathbb{E}\left(Y^n\right) =&\frac{n}{g^{n-1}}\sum_{k=0}^{n-1}(-1)^k\binom{n-1}{k}\int_{-\infty}^{\infty}T_{\widetilde{g},\widetilde{h}}(u)f_U(u)du\\ =&\frac{2}{g^n}\sum_{k=0}^{n}(-1)^k\binom{n}{k}\int_0^{\infty}\cosh\left(\widetilde{g}u\right)\exp\left\{\frac{1}{2}\widetilde{h}u^2\right\}f_U(u)du,\end{aligned}$$

where $\widetilde{g}=(n-k)g$ and $\widetilde{h}=nh$. In the latter term, we used that $f_U(u)$ is a function symmetrical about the origin. □

### 3.4.1. *Special cases of moments*

In general, when the continuous random variable $U$ is symmetrically distributed about the origin, then the moment generating function $(mgf)$ can be written as follows

$$M_U(t)=\mathbb{E}\left(e^{tU}\right)=2\int_0^{\infty}\cosh\left(tu\right)f_U(u)du, \tag{3.13}$$

and the characteristic function for the random variable $U$ is given by

$$\Psi_U(t)=\mathbb{E}\left(e^{itU}\right)=2\int_0^{\infty}\cos\left(tu\right)f_U(u)du, \tag{3.14}$$

where $i$ is the imaginary quantity whose value is equal to $\sqrt{-1}$. Since that $f_U(u)$ is an even function, then the Fourier integral representation of $f_U(u)$ may be written as

$$f_U(u)=\int_0^{\infty}A(t)\cos(ut)\,dt, \qquad \text{with} \qquad A(t)=\frac{1}{\pi}\Psi_U(t).$$

Using the Fourier frequency convolution theorem we can write

$$2\int_0^{\infty}\cos(gt)\,f_U(t)e^{-\frac{|h|}{2}t^2}dt=\mathfrak{F}\left[f_U(t)\exp\left\{-\frac{|h|}{2}t^2\right\}\right]=\frac{1}{\sqrt{2|h|\pi}}\exp\left\{-\frac{g^2}{2|h|}\right\}*\mathfrak{F}\left[f_U(t)\right],$$

where $*$ denotes convolution. The expression (3.13) allows us to obtain the moments of Tukey's $g-h$ distribution. However moments of some orders do not exist for a certain range of values of the parameter $h$, considering that we have the following cases:

(1) Supposing that $U\sim GED\left(\frac{1}{2}\right)$ and $h<\frac{1}{n}$, we have

$$\mathbb{E}\left(Y^n\right)=\begin{cases}\dfrac{1}{g^n\sqrt{1-nh}}\sum\limits_{k=0}^{n}(-1)^k\binom{n}{k}M_U\left(\frac{n-k}{\sqrt{1-nh}}g\right) & g\neq 0\\ \dfrac{1+(-1)^n}{2^{\frac{n}{2}}\left[1-\widetilde{h}\right]^{\frac{n+1}{2}}}\dfrac{\Gamma(n)}{\Gamma(n/2)} & g=0\end{cases} \tag{3.15}$$

where $M_U(t)$ is the $mgf$ of a standard normal random variable and $\Gamma(\cdot)$ is the Gamma function. This expression is consistent with those obtained by Martínez & Iglewicz (1984).

(2) When $U \sim GED(1)$ and $h < 0$, we have[b]

$$\mu_n' = \begin{cases} \frac{1}{g^n}\sqrt{\frac{\pi}{n|h|}}\left\{\sum_{k=0}^{n-1}(-1)^k\binom{n}{k}\left[\exp\left\{\frac{\alpha_{n,k}^2}{2}\right\}\Phi(-\alpha_{n,k})+ \right.\right. \\ \quad \left.\left. \exp\left\{\frac{\beta_{n,k}^2}{2}\right\}\Phi(\beta_{n,k})\right]+2(-1)^n e^{\frac{1}{n|h|}}\Phi\left(-\sqrt{\frac{2}{n|h|}}\right)\right\}, & g \neq 0, \\ \frac{1+(-1)^n}{2\sqrt{n|h|}}\left(\frac{\sqrt{2}}{n|h|}\right)^n e^{\frac{1}{n|h|}}\sum_{k=0}^{n}\binom{n}{k}(-1)^k\left[\Gamma\left(\frac{k+1}{2}\right)\right. \\ \quad \left. -\int_0^{\frac{1}{n|h|}} u^{\frac{1}{2}(k-1)}e^{-u}\,du\right], & g = 0, \end{cases} \tag{3.16}$$

where $\alpha_{n,k}$ and $\beta_{n,k}$ are the larger and smaller roots respectively, of the quadratic equation

$$n|h|r^2 - 2(n-k)\sqrt{n|h|}gr + (n-k)^2 g^2 - 2 = 0. \tag{3.17}$$

Expression (3.16) was wrongly calculated in Klein & Fischer (2002).

From the preceding equations we obtain the expected value $\mu$ for $g \neq 0$:

(1) Assumes that $U \sim GED\left(\frac{1}{2}\right)$. Using the expression (3.15) with $n = 1$ for calculated $\mathbb{E}[Y]$, then we must

$$\mathbb{E}[Y] = \frac{1}{g\sqrt{1-h}}\left(e^{\frac{1}{2}\frac{g^2}{1-h}} - 1\right). \tag{3.18}$$

(2) Assuming in the expression (2.6) that the variable $U \sim GED(1)$, $h < 0$ and using the expression (3.16) with $n = 1$, we obtain

$$\mu_{g,h}^L = \frac{1}{g}\sqrt{\frac{\pi}{|h|}}\left[e^{\frac{1}{2}\alpha_{1,0}^2}\Phi(-\alpha_{1,0}) + e^{\frac{1}{2}\beta_{1,0}^2}\Phi(\beta_{1,0}) - 2e^{\frac{1}{|h|}}\Phi\left(-\sqrt{\frac{2}{|h|}}\right)\right], \tag{3.19}$$

where $\alpha_{1,0}$ and $\beta_{1,0}$ be the larger and smaller roots of the quadratic equation given in (3.17), respectively.

## 4. The *g* generalized distribution

The $g$ generalized distribution given by equation (2.4) is a nonlinear transform of a continuous random variable $U$ and is parameterized by $g$. This subfamily contains distributions whose skewness increases when the value of the parameter $g$ increases. This subfamily of distributions to help them get to have great importance in the statistical analysis to be a suitable means to study skewed distributions. Its distributional form includes only the parameter $g$ which fixes the amount and direction of skewness.

[b]Appendix B contains the respective proof of this expression.

Now, we give below an empirical rule for a random variable $X$ which can be expressed as (2.6) with $Y = T_{g,0}(U)$,

$$\frac{x_p - \theta}{x_{0.5} - \theta} = \frac{\theta - x_{0.5}}{\theta - x_{1-p}} \qquad \text{for all } \; p > 0.5. \tag{4.1}$$

In particular, the expression (4.1) is satisfied if

$$\theta = A - \operatorname{sgn}(g)\,\frac{B}{|g|}, \tag{4.2}$$

where $\operatorname{sgn}(\cdot)$ denote the signum function. The constant $\theta$ relates to the location and scale parameters, known as "threshold parameter" and was given by Hoaglin et al. (1985). Taking $h = 0$ in expression (3.2) and replacing the expression (3.5) we get that

$$t_{g,0}(y) = \frac{1}{1+gy} f_U\left(\frac{\ln(1+gy)}{g}\right), \qquad gy > -1. \tag{4.3}$$

Moreover, if we solve for the variable $y$ in equation (2.7) by substituting the expression given in (4.3), we obtain

$$t_{g,0}\left(\frac{x-A}{B}\right) = f_U\left(\frac{1}{g}\ln\left(1+\frac{x-A}{B/g}\right)\right)\left[1+\frac{x-A}{B/g}\right]^{-1} \qquad \frac{x-A}{B/g} > -1$$

Since $g \in \mathbb{R}$ then

$$t_{g,0}\left(\frac{x-A}{B}\right) = \begin{cases} \frac{B}{g}\,\dfrac{f_U\left(\frac{1}{g}\left(\ln(x-\theta) - \ln\left(\frac{B}{g}\right)\right)\right)}{x-\theta} & \text{if } g > 0 \\ \frac{B}{|g|}\,\dfrac{f_U\left(\frac{1}{|g|}\left(\ln\left(\frac{B}{|g|}\right) - \ln(\theta - x)\right)\right)}{\theta - x} & \text{if } g < 0 \end{cases} \tag{4.4}$$

where $\frac{B}{|g|} > 0$, for simplicity and without loss of generality we assume $g > 0$ and we replace the expression (4.2) and if we use the result given in (3.3), which relates the *pdf* of $X$ and $Y = T_{g,h}(Z)$ on the quantiles, we can rewrite (4.4) as follows

$$f_X(x) = \frac{1}{g(x-\theta)} f_U\left(\frac{1}{g}\left(\ln(x-\theta) - \mu^*\right)\right) \qquad x > \theta, \tag{4.5}$$

where $\mu^* = \ln\left(\frac{B}{g}\right)$. We say that the random variable $X$ has a log-symmetric distribution with threshold parameter $\theta$, scale parameter $\mu^*$ and shape parameter $g$, denoted by $X \sim LS\big(\mu^*, g, \theta\big)$. If $\theta = 0$ we denote by $X \sim LS\big(\mu^*, g\big)$. The *cdf* of the random variable $X$ given by

$$F_X(x) = F_U\left(\frac{1}{g}\left(\ln(x-\theta) - \mu^*\right)\right), \qquad x > \theta. \tag{4.6}$$

Expression (4.5) allows us to obtain the following *pdf* associated with the Tukey's $g$ function.

### 4.1. *Special cases*

(1) If $U \sim GED\left(\frac{1}{2}\right)$ and $g \neq 0$, we have that

$$f_X(x) = \frac{1}{\sqrt{2\pi}g(x-\theta)} \exp\left\{-\frac{1}{2}\left(\frac{\ln(x-\theta)-\mu^*}{g}\right)^2\right\}, \tag{4.7}$$

where $\mu^* = \ln(\mu_X - \theta) - \frac{1}{2}g^2$ and $x > \theta$. Note that when $\theta = 0$ the last expression coincides with the *pdf* of the classic Log Normal random variable. In this case, we say that $X$ is Log-Normal distributed with three parameters $\mu_X$, $g$ and $\theta$. Many practical applications of this distribution are discussed in the literature, for example, Aitchison & Brown (1963) and Crow & Shimizu (1988).

(2) When $U \sim GED(1)$ and $0 < g < \frac{\sqrt{2}}{n}$, the resulting distribution is given by the *pdf*

$$f_X(x) = \frac{\beta}{2(\varepsilon-\theta)} \begin{cases} \left(\frac{x-\theta}{\varepsilon-\theta}\right)^{\beta-1}, & \theta < x < \varepsilon \\ \left(\frac{\varepsilon-\theta}{x-\theta}\right)^{\beta+1}, & x \geq \varepsilon, \end{cases} \tag{4.8}$$

where $\beta = \frac{\sqrt{2}}{g}$ and $(\varepsilon - \theta) = (\mu_X - \theta)\left(1 - \frac{1}{\beta^2}\right)$. Note again that this expression coincides with the *pdf* of log-Laplace with three parameters $\mu_X$, $g$ and $\theta$.

(3) If $U \sim \text{Logistic}\left(0, \lambda^{-1}\right)$, $0 < g < \frac{\lambda}{n}$ and $\lambda = \frac{\pi}{\sqrt{3}}$, then the *pdf* of $X$ can be expressed as

$$f_X(x) = \frac{\pi}{\varepsilon-\theta}\left[\frac{\pi}{\alpha}\frac{x-\theta}{\varepsilon-\theta}\right]^{\alpha-1}\left[1+\frac{\pi}{\alpha}\frac{x-\theta}{\varepsilon-\theta}\right]^{-2\alpha}, \tag{4.9}$$

where $\alpha = \frac{\lambda}{g}$ y $(\varepsilon - \theta) = (\mu_X - \theta)\sin\left(\sqrt{3}g\right)$. Note that this expression coincides with the *pdf* of three parameters Log-Logistic ($\mu_X$, $g$ and $\theta$).

Taking the expectation of the linear transformation given in equation (2.6) we obtain

$$\mathbb{E}(X-\theta) = \frac{B}{g}\mathbb{E}\left[e^{gU}\right] = \frac{B}{g}M_U(g) \qquad \Rightarrow \qquad B = g\frac{\mathbb{E}(X)-\theta}{M_U(g)},$$

where $M_U(g)$ is the $mgf$ of the random variable $U$. The $n$th moment of the random variable $X$ could be obtained using the formula

$$\mathbb{E}\left[(X-\mathbb{E}[X])^n\right] = \mu_n(X) = \exp\{n\mu^*\}\sum_{k=0}^{n}(-1)^k\binom{n}{k}M_U(\tilde{g})M_U^k(g),$$

note that these expressions do not depend on the parameter $\theta$. Thus, the standardized values for skewness and kurtosis corresponding to linear transformation given by equation (2.6) with $Y = T_{g,0}(U)$ can be expressed as

$$SK_1(X) = \frac{M_U(3g) - 3M_U(2g)M_U(g) + 2M_U^3(g)}{\left[M_U(2g) - M_U^2(g)\right]^{\frac{3}{2}}}, \tag{4.10}$$

$$KR_1(X) = \frac{M_U(4g) - 4M_U(3g)M_U(g) + 6M_U(2g)M_U^2(g) - 3M_U^4(g)}{\left[M_U(2g) - M_U^2(g)\right]^2}. \tag{4.11}$$

Note that the above expressions depend only on the parameter $g$.

The $n$-th moment of the random variable $X-\theta$ is given by

$$\mathbb{E}\left[(X-\theta)^n\right]=\left(\frac{B}{g}\right)^n M_U(ng). \tag{4.12}$$

When we rewrite the expression (4.12) and use properties of the $mgf$, we obtain

$$\mathbb{E}\left(e^{n\ln(X-\theta)}\right)=M_V(n)=e^{n\ln\left(\frac{B}{g}\right)}M_U(ng)=M_{\ln\left(\frac{B}{g}\right)+gU}(n), \tag{4.13}$$

where $V=\ln(X-\theta)$, then $\mathbb{E}(V)=\mu_V=\ln\left(\frac{B}{g}\right)$ and $Var(V)=\sigma_V^2=g^2$. When the relation (4.1) is satisfied, then $h=0$ and if we assume that $\theta>x_{\min}$, we can conclude that the value of $g$ is estimated by $g=\operatorname{sgn}(SK_1(X))\,\sigma_V$. Here $SK_1(X)$ denote the coefficient of skewness from the variable we want to approximate. The scale parameter is estimated by $B=g\exp\{\mathbb{E}(V)\}$.

### 4.2. *Approximations*

We first assume the value of $\theta$ to be negligibly small in (4.12) to obtain

$$\mathbb{E}(X^n)=\left(\frac{B}{g}\right)^n M_U(ng). \tag{4.14}$$

The above expression allows to obtain the various moments about the origin of the random variable $X$, when the distribution of $U$ includes the normal, hyperbolic secant, hyperbolic cosecant, Logistic and Laplace, which are all symmetric with standarized skewness of zero.

In (4.14) if we let $U\sim GED\left(\frac{1}{2}\right)$ and $g>0$, we obtain

$$\mathbb{E}(X^n)=\left(\frac{B}{g}\right)^n\exp\left\{\frac{1}{2}n^2g^2\right\}=\exp\left\{n\ln\left(\frac{B}{g}\right)+\frac{1}{2}n^2g^2\right\}. \tag{4.15}$$

This expression coincides with the $mgf$ of a Normal random variable with parameters $\mu=\ln\left(\frac{B}{g}\right)$ and $\sigma=g$. By the uniqueness of the $mgf$, we conclude that $V=\ln(X)\sim N\left(\ln\left(\frac{B}{g}\right),g\right)$, i.e., $V$ is a Lognormal random variable with parameters $\mu=\ln\left(\frac{B}{g}\right)$ and $\sigma=g$.

Similarly, we show that the relation between the random variables $X$ and $U$ presented in Table 1, for the selected set of well known symmetrical distributions.

| **Distribution of the r.v. U** | **Parameters** | | | **Distribution of the r.v. V** | **Parameters** | |
|---|---|---|---|---|---|---|
| | $\mu, a$ | $\sigma, b$ | $g\neq 0$ | | $\mu, a$ | $\sigma, b$ |
| Laplace | 0 | $\frac{\sqrt{2}}{2}$ | $0<g<\frac{\sqrt{2}}{n}$ | Log-Laplace | $\ln\left(\frac{B}{g}\right)$ | $\frac{\sqrt{2}}{2}\lvert g\rvert$ |
| Logistic | 0 | $\frac{\sqrt{3}}{\pi}$ | $0<g<\frac{\pi}{\sqrt{3}n}$ | Loglogistic | $\ln\left(\frac{B}{g}\right)$ | $\frac{\sqrt{3}}{\pi}\lvert g\rvert$ |
| Normal | 0 | 1 | $g>0$ | Lognormal | $\ln\left(\frac{B}{g}\right)$ | $g$ |
| HyperSec | 0 | $\frac{2}{\pi}$ | $0<g<\frac{\pi}{2n}$ | LoghyperSec | $\ln\left(\frac{B}{g}\right)$ | $\frac{2}{\pi}\lvert g\rvert$ |
| HyperCsc | 0 | $\frac{\sqrt{2}}{\pi}$ | $0<g<\frac{\pi}{\sqrt{2}n}$ | LoghyperCsc | $\ln\left(\frac{B}{g}\right)$ | $\frac{\sqrt{2}}{\pi}\lvert g\rvert$ |

Table 1. Parameters of the *pdf* of the random variable $V=\ln(X)$

Published by Atlantis Press
Copyright: the authors

## 5. An Illustration

We consider now data concerning the circumference measures (centimeters) taken from the ankle, chest, hip, neck and of 252 adult men. The data have been previously analyzed in Headrick (2010) and are available for download at *http://lib.stat.cmu.edu/datasets/bodyfat*. The following table presents the statistics for these data.

| Variable | Mean | St. Dev. | SK1 | KR1 | JB test |
|---|---|---|---|---|---|
| Ankle | 23.1024 | 1.6949 | 2.2417 | 14.6858 | 1631.8565 |
| Chest | 100.8242 | 8.4305 | 0.6775 | 3.9441 | 28.4092 |
| Hip | 99.9048 | 7.1641 | 1.4882 | 10.3002 | 647.4181 |
| Neck | 37.9921 | 2.4309 | 0.5493 | 5.6422 | 85.2964 |

Table 2. Summary Descriptive Statistics

By using the test proposed by Jarque & Bera (1987), the statistics in Table 2 clearly indicate that the distribution of each of the variables can not be normal random variable. When $U \sim GED\left(\frac{1}{2}\right)$, the $g$ and $h$ parameter estimates result in a fitted distribution matching the sample moments

| Variable | $A$ | $B$ | $g$ | $h$ | Mean | St. Dev. | SK1 | KR1 |
|---|---|---|---|---|---|---|---|---|
| Ankle | 22.7282 | 1.2843 | 0.5125 | 0.0376 | 23.1016 | 1.6915 | 2.2417 | 14.6858 |
| Chest | 99.9523 | 8.0301 | 0.2117 | 0.0082 | 100.8225 | 8.4138 | 0.6775 | 3.9441 |
| Hip | 98.9181 | 5.7427 | 0.2933 | 0.0846 | 99.9028 | 7.1498 | 1.4882 | 10.3003 |
| Neck | 37.8553 | 2.0760 | 0.1143 | 0.0871 | 37.9918 | 2.4261 | 0.5493 | 5.6422 |

Table 3. Estimation results

When $U \sim GED(1)$, the $g$ and $h$ parameter estimates result in a fitted distribution matching the sample moments

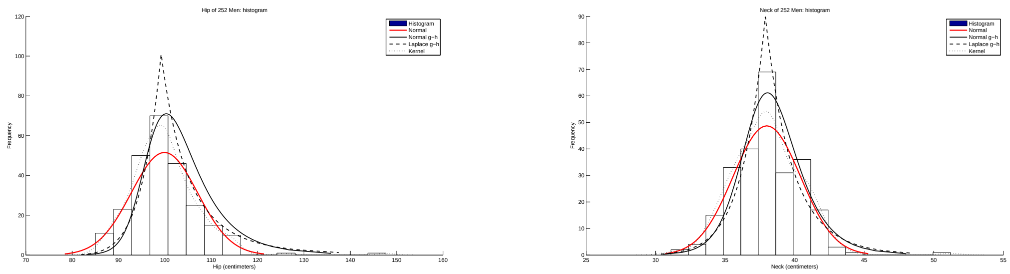
| Variable | $A$ | $B$ | $g$ | $h$ | Mean | St. Dev. | SK1 | KR1 |
|---|---|---|---|---|---|---|---|---|
| Ankle | 22.8330 | 1.5613 | 0.3349 | -0.0273 | 23.0878 | 1.6915 | 2.2417 | 14.6858 |
| Chest | 100.0895 | 9.6635 | 0.1771 | -0.0721 | 100.8069 | 8.4137 | 0.6775 | 3.9441 |
| Hip | 99.1886 | 6.9025 | 0.2040 | -0.0098 | 99.8867 | 7.1498 | 1.4882 | 10.3001 |
| Neck | 37.8884 | 2.4850 | 0.0856 | -0.0122 | 37.9914 | 2.4261 | 0.5493 | 5.6422 |

Table 4. Estimation results

Inspection of these tables indicates that both the Normal $g-h$ and Laplace $g-h$ pdfs provide good approximations to the empirical data.

Figures 2 for Hip and Neck, respectively, shows such a histogram and the pdfs indicates that the two transformations will produce similar approximations for this particular set of sample statistics.

Since the value of $h$ for variable Chest when $U \sim GED\left(\frac{1}{2}\right)$ is very small, we assume this parameter equal to zero, to illustrate the process of adjusting using Tukey's $g$ generalized family of distributions, we assume zero to approximate g by Tukey's generalized.




(a) (b)

Fig. 2. (a) Hip vs. Normal Distribution and estimated *pdf*'s Tukey's $g-h$. (b) Neck vs. Normal Distribution and estimated *pdf*'s Tukey's $g-h$

To pursue elongation in these data, we first verify whether if it satisfies the condition given in (4.1). The value of $\theta$ turns out to be $-66.5955$. Letting the parameter $h$ equal to zero, the mean and standard deviation of the variable $Z$ are 5.1193 and 0.04969, respectively. The expression (2.6) reduces to

$$X = \frac{B}{g}\exp\{gU\} + \theta; \tag{5.1}$$

where

$$g = 0.04969 \qquad \text{and} \qquad B = 8.3088.$$

Figure 3 shows such a histogram and it is evident that the data have a slight degree of skewness to the left, leptokurtic and do not follow the normal distribution.

As shown in Figure 3, there is a marked difference between the empirical distribution of the data (represented by the histogram) and the normal distribution. Tukey's $g-h$ family of generalized distributions better approximates the empirical.

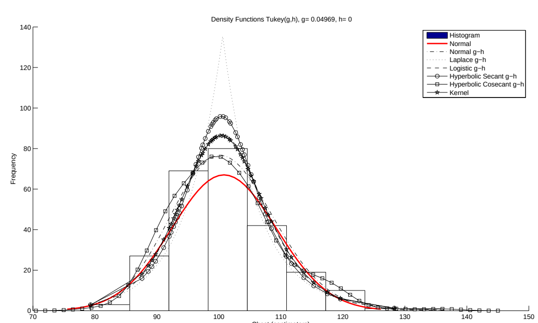



Fig. 3. Chest vs. Normal Distribution and estimated *pdf*'s Tukey's $g-h$

In order to determine how the fitted distribution agrees with fitted date, we use the methodology described by Hoaglin et al. (1985) to determine the sample quantiles of the form $p = 2^{-k}, k = 1,2,\ldots,8$. In Table 5 we present these quantile $p-$values along with their estimates, calculated using (5.1) by varying the variable $U$.

| $p$ | $X^{(1)}$ | $X^{(2)}$ | $X^{(3)}$ | $X^{(4)}$ | $X^{(5)}$ |
|---|---|---|---|---|---|
| $\frac{1}{256}$ | 83.4 | 81.0038 | 76.7796 | 78.7795 | 77.6844 |
| $\frac{1}{128}$ | 85.1 | 82.3667 | 79.1386 | 80.8365 | 80.0255 |
| $\frac{1}{64}$ | 86.7 | 84.0673 | 82.1685 | 83.2546 | 82.7684 |
| $\frac{1}{32}$ | 88.2 | 85.6874 | 84.8543 | 85.4107 | 85.1965 |
| $\frac{1}{16}$ | 89.2 | 88.2914 | 88.7569 | 88.5973 | 88.7254 |
| $\frac{1}{8}$ | 92.1 | 91.3309 | 92.6691 | 91.9394 | 92.2918 |
| $\frac{1}{4}$ | 94.2 | 95.0540 | 96.5452 | 95.6102 | 95.9715 |
| $\frac{1}{2}$ | 99.6 | 100.5735 | 100.5883 | 100.5753 | 100.5787 |
| $\frac{3}{4}$ | 105.3 | 106.2579 | 104.6864 | 105.6713 | 105.2917 |
| $\frac{7}{8}$ | 110.1 | 110.2843 | 108.7677 | 109.5978 | 109.1998 |
| $\frac{15}{16}$ | 115.3 | 113.6521 | 112.9997 | 113.2513 | 113.0733 |
| $\frac{31}{32}$ | 118.5 | 116.5300 | 117.2418 | 116.7229 | 116.9047 |
| $\frac{63}{64}$ | 119.8 | 119.0230 | 121.4055 | 120.0352 | 120.6461 |
| $\frac{127}{128}$ | 121.6 | 121.1539 | 125.3393 | 123.1144 | 124.1705 |
| $\frac{255}{256}$ | 128.3 | 122.8980 | 128.8264 | 125.8167 | 127.2879 |

Table 5. Observed and estimated values by the expression (5.1) for the heights of Australian athletes

The columns of Table 5 provide the following information:

$X^{(1)}$ : Sample quantiles.
$X^{(2)}$ : Values obtained using equation (5.1) with $U \sim GED\left(\frac{1}{2}\right)$.
$X^{(3)}$ : Values obtained using equation (5.1) with $U \sim GED(1)$.
$X^{(4)}$ : Values obtained using equation (5.1) with $U \sim \text{Logistic}\left(0, \frac{\sqrt{3}}{\pi}\right)$.
$X^{(5)}$ : Values obtained using equation (5.1) with $U \sim \text{sech}\left(0, \frac{2}{\pi}\right)$.

Note that these adjustments are satisfactory for the four distributions used in the expression (5.1). Table 6 summarize the statistical results for the *pdf* of each estimated $g-h$.

| Fitted distribution | Mean | Stan. Dev. | SK1 | KR1 |
|---|---|---|---|---|
| Normal $g-h$ | 100.8222 | 8.3189 | 0.1426 | 2.9492 |
| Laplace $g-h$ | 100.8194 | 8.2724 | 0.3109 | 5.3597 |
| Logistic $g-h$ | 100.8207 | 8.2891 | 0.2087 | 3.8893 |
| HyperSec $g-h$ | 100.8199 | 8.2811 | 0.2534 | 4.5306 |

Table 6. Results for the estimation of Chest taken from 252 men.

These results indicate the importance of selecting a distribution on the $g-h$ transformation, when $U \sim \text{Logistic}\left(0, \frac{\sqrt{3}}{\pi}\right)$ the sample moments are closer to the theoretical moments.

## 6. Conclusion

This paper presents a generalization of the well-known Tukey's $g-h$ family of distributions for fitting skewed data. We calculate explictly the *cdf* and *pdf*, and also the set of regularity properties obtained with respect to the expected values and variances. We also present a simulation procedure to estimate the value of the paramater $g$, that is, the standard deviation of the random variable $\ln(X-\theta)$, when the parameter $h$ goes to zero. The proposed generalization is also used to generated a large class distributions from a symmetric density of the parameters $g$ and $h$ which controls the skewness and the elongation of the tails, respectively.

## References

Aitchison, J. & Brown, J. (1963), *The Lognormal Distribution*, Cambridge University Press, United Kingdom.

Badrinath, S. & Chatterjee, S. (1988), 'On measuring skewness and elongation in common stock return distributions: The case of the market index', *The Journal of Business* **61**(4), 451–472.

Badrinath, S. & Chatterjee, S. (1991), 'A data analytic look at skeness and elongation in common stock return distributions', *Journal of Business & Economic Statistics* **9**(2), 223–233.

Crow, E. L. & Shimizu, K. (1988), *Lognormal distributions: Theory and applications*, Statistics, textbooks and monographs, CRC Press, New York.

Crow, E. L. & Siddiqui, M. M. (1967), 'Robust estimation of location', *Journal of the American Statistical Association* **62**(318), 353–389.

Dutta, K. K. & Babbel, D. F. (2004), 'On measuring skewness and kurtosis in short rate distributions: The case of the us dollar london inter bank offer rates', *Working paper* $02-25$, *Wharton School Financial Institutions Center* .

Dutta, K. K. & Babbel, D. F. (2005), 'Extracting probabilistic information from the prices of interest rate options: Tests of distributional assumptions', *Journal of Business* **78**(3), 841–870.

Dutta, K. K. & Perry, J. (2007), 'A tale of tails: An empirical analysis of loss distribution models for estimating operational risk capital', *Federal Reserve Bank of Boston, Working Paper No. 06-13* .

Groeneveld, R. A. & Meeden, G. (1984), 'Measuring skewness and kurtosis', *Journal of the Royal Statistical Society. Series D (The Statistician)* **33**(4), 391–399.

Headrick, T. C. (2010), *Statistical Simulation: Power Method Polynomials and Other Transformations*, Taylor & Francis Group, LLC, Chapman & Hall/CRC Press, Boca Raton, FL, USA.

Hinkley, D. V. (1975), 'On power transformations to symmetry', *Biometrika* **62**(1), 101–111.

Hoaglin, D. C. (1985), 'Summarizing shape numerically: the $g-$and$-h$ distributions', *In: Hoaglin, D. C., Mosteller, F., Tukey, J. W. (Eds.), Exploring Data Tables, Trends, and Shapes.* pp. 461–513. John Wiley & Sons.

Hoaglin, D. C., Mosteller, F. & Tukey, J. W. (1985), *Exploring Data Tables, Trends, and Shapes*, John Wiley & Sons, New York.

Hogg, R. V. (1974), 'Adaptive robust procedures: A partial review and some suggestions for future applications and theory', *Journal of the American Statistical Association* **69**(348), 909–923.

Jarque, C. M. & Bera, A. K. (1987), 'A test for normality of observations and regression residuals', *International Statistical Review* **55**(2), 163–172.

Jiménez, J. A. (2004), Aproximaciones de las funciones de riesgo del tiempo de sobrevivencia mediante la distribución $g-h$ de tukey, Especialista en actuaría, Facultad de Ciencias. Departamento de Matemáticas. Universidad Nacional de Colombia. Sede Bogotá.

Jiménez, J. A. & Arunachalam, V. (2011), 'Using tukey's $g$ and $h$ family of distributions to calculate value at risk and conditional value at risk', *Journal of Risk* **13**(4), 95 – 116.

Jiménez, J. A. & Martínez, J. (2006), 'An estimation of the parameter tukey's $g$ distribution', *Colombian Journal of Statistics* **29**(1), 1 – 16.

Klein, I. & Fischer, M. (2002), '$gh-$transformation of symmetrical distributions', *In: Mittnik, Stefan, and Klein, Ingo (Eds.), Contribution to Modern Econometrics* pp. 119–134. Kluwer Academic Publishers.

Majumder, M. M. A. & Ali, M. M. (2008), 'A comparison of methods of estimation of parameters of tukey's *gh* family of distributions', *Pakistan Journal of Statistics* **24**(2), 135–144.

Martínez, J. & Iglewicz, B. (1984), 'Some properties of the tukey $g$ and $h$ family of distributions', *Communications in Statistics - Theory and Methods* **13**(3), 353–369.

Mills, T. C. (1995), 'Modelling skewness and kurtosis in the london stock exchange $ft-se$ index return distributions', *The Statistician* **44**(3), 323–332.

Oberhettinger, F. (1973), *Fourier transforms of distributions and their inverses: a collection of tables*, Academic Press, New York.

Olver, F. W. J., Lozier, D. W., Boisvert, R. F. & Clark, C. W. (2010), *Handbook of Mathematical Functions*, Cambridge University Press, New York.

Tang, X. & Wu, X. (2006), 'A new method for the decomposition of portfolio *var*', *Journal of Systems Science and Information* **4**(4), 721–727.

Tocher, K. D. (1964), *The art of simulation*, Electrical engineering series, D. Van Nostrand Company, Inc, Princeton, NJ.

Tukey, J. W. (1977), Modern techniques in data analysis, Nsp-sponsored regional research conference at southeastern massachesetts university, North Dartmouth, Massachesetts.

## Appendix A: Proof of propositions 3.1 and 3.2

**Proof.** (Proposition 3.1)

We consider the $m-$th power of the expression (2.2),

$$\begin{aligned} Y^m =& \frac{1}{g^m}\sum_{k=0}^{m}\binom{m}{k}(-1)^k \exp\left\{\widetilde{g}U + \frac{1}{2}\widetilde{h}U^2\right\} \\ =& \frac{m}{g^{m-1}}\left[\sum_{k=0}^{m-1}\binom{m-1}{k}\frac{(-1)^k}{\widetilde{g}}\exp\left\{\widetilde{g}U + \frac{1}{2}\widetilde{h}U^2\right\} + \frac{(-1)^m}{mg}e^{\frac{1}{2}\widetilde{h}U^2}\right], \end{aligned}$$

where $\widetilde{g} = (m-k)g$ and $\widetilde{h} = mh$, since $(-1)^m = -\sum_{k=0}^{m-1}\binom{m}{k}(-1)^k$, then

$$\begin{aligned} Y^m =& \frac{m}{g^{m-1}}\sum_{k=0}^{m-1}\binom{m-1}{k}\frac{(-1)^k}{\widetilde{g}}\left[\exp\left\{\widetilde{g}U + \frac{1}{2}\widetilde{h}U^2\right\} - e^{\frac{1}{2}\widetilde{h}U^2}\right] \\ =& \frac{m}{g^{m-1}}\sum_{k=0}^{m-1}\binom{m-1}{k}\frac{(-1)^k}{\widetilde{g}}\left[\exp(\widetilde{g}U) - 1\right]\exp(\widetilde{h}U^2/2) \end{aligned}$$

which is the required result. □

**Proof.** (Proposition 3.2)

Suppose that $u_q$ is the smallest number satisfying $F_U(u_q) = q$ ie $q$-th quantile of $U$, making the change of variable

$$w = u_q = F_U^{-1}(q) \qquad\qquad dw = du_q = \frac{dq}{F_U'(u_q)},$$

here we use the expression given in (3.1), since $F_U'(w) = f_U(w)$, and

$$\lim_{u\to-\infty} F_U(u) = 0 \qquad\qquad \lim_{u\to\infty} F_U(u) = 1,$$

moreover given that $f_U(w)$ is a function with domain the real line and counterdomain the infinite interval $[0,\infty)$, we solve for $dq$ and we obtain

$$\int_0^1 \left[F_U^{-1}(q)\right]^n dq = \int_{-\infty}^{\infty} w^n f_U(w)\, dw. \qquad \square$$

**Appendix B: Proof of formula given in** (3.16)

In this Appendix, we present the calculation details of the equation given in (3.16), using the Table $I$ of Fourier transforms (Oberhettinger (1973), of expression (79)) after some calculations and simplifying, we get

$$2\int_0^{\infty} \cos(\widetilde{g}t)\, f_U(t) \exp\left\{-\frac{|\widetilde{h}|}{2}t^2\right\} dt = \sqrt{\frac{\pi}{n|h|}}\left[\exp\left\{\left(\frac{\sqrt{2}-i\widetilde{g}}{\sqrt{2n|h|}}\right)^2\right\}\Phi\left(\frac{i\widetilde{g}-\sqrt{2}}{\sqrt{n|h|}}\right)\right.$$
$$\left. + \exp\left\{\left(\frac{\sqrt{2}+i\widetilde{g}}{\sqrt{2n|h|}}\right)^2\right\}\Phi\left(-\frac{i\widetilde{g}+\sqrt{2}}{\sqrt{n|h|}}\right)\right],$$

where $i$ is the imaginary quantity and $\Phi(\cdot)$ is the *cdf* of a standard normal variable, then

$$2\int_0^{\infty} \cosh(\widetilde{g}t)\, f_U(t) e^{-\frac{|\widetilde{h}|}{2}t^2} dt = \sqrt{\frac{\pi}{n|h|}}\left[\exp\left\{\left(\frac{\sqrt{2}+\widetilde{g}}{\sqrt{2n|h|}}\right)^2\right\}\Phi\left(-\frac{\widetilde{g}+\sqrt{2}}{\sqrt{n|h|}}\right)\right.$$
$$\left. + \exp\left\{\left(\frac{\sqrt{2}-\widetilde{g}}{\sqrt{2n|h|}}\right)^2\right\}\Phi\left(\frac{\sqrt{2}-\widetilde{g}}{\sqrt{n|h|}}\right)\right].$$

Substituting the above expression in (3.11) and simplifying we get,

$$\mu_n' = \frac{1}{g^n}\sqrt{\frac{\pi}{n|h|}}\sum_{k=0}^{n}(-1)^k\binom{n}{k}\left[\exp\left\{\frac{1}{2}\left(\frac{\widetilde{g}+\sqrt{2}}{\sqrt{n|h|}}\right)^2\right\}\Phi\left(-\frac{\widetilde{g}+\sqrt{2}}{\sqrt{n|h|}}\right)\right.$$
$$\left. + \exp\left\{\frac{1}{2}\left(\frac{\widetilde{g}-\sqrt{2}}{\sqrt{n|h|}}\right)^2\right\}\Phi\left(\frac{\widetilde{g}-\sqrt{2}}{\sqrt{n|h|}}\right)\right].$$

When $g = 0$ and $h < 0$, we have

$$\mu_n' = \frac{1+(-1)^n}{2\sqrt{n\,|h|}}\left(\frac{\sqrt{2}}{n\,|h|}\right)^n e^{\frac{1}{n|h|}}\sum_{k=0}^{n}\binom{n}{k}(-1)^k\left[\Gamma\left(\frac{k+1}{2}\right) - \int_0^{\frac{1}{n|h|}} u^{\frac{1}{2}(k-1)}e^{-u}\,du\right].$$


Copyright: the authors