2008

Parametric Probability Densities and Distribution Functions for Tukey *g*-and-*h* Transformations and their Use for Fitting Data

Todd C. Headrick
Southern Illinois University Carbondale, headrick@siu.edu

Rhonda K. Kowalchuk
Southern Illinois University Carbondale

Yanyan Sheng
Southern Illinois University Carbondale

Published in *Applied Mathematical Sciences*, Vol. 2 No. 9 (2008) at http://www.m-hikari.com/ams/ams-password-2008/ams-password9-12-2008/headrickAMS9-12-2008.pdf

Applied Mathematical Sciences, Vol. 2, 2008, no. 9, 449 - 462

Parametric Probability Densities and Distribution Functions for Tukey *g*-and-*h* Transformations and their Use for Fitting Data

Todd C. Headrick, Rhonda K. Kowalchuk and Yanyan Sheng

Section on Statistics and Measurement
Department EPSE, 222-J Wham Bldg
Southern Illinois University-Carbondale
Carbondale, IL 62901-4618, USA
headrick@siu.edu

Abstract

The family of g-and-h transformations are popular algorithms used for simulating non-normal distributions because of their simplicity and ease of execution. In general, two limitations associated with g-and-h transformations are that their probability density functions (pdfs) and cumulative distribution functions (cdfs) are unknown. In view of this, the g-and-h transformations' pdfs and cdfs are derived in general parametric form. Moments are also derived and it is subsequently shown how the g and h parameters can be determined for prespecified values of skew and kurtosis. Numerical examples and parametric plots of g-and-h pdfs and cdfs are provided to confirm and demonstrate the methodology. It is also shown how g-and-h distributions can be used in the context of distribution fitting using real data sets.

Mathematics Subject Classification: 65C05, 65C10, 65C60

Keywords: Distribution fitting, Moments, Monte Carlo, Non-normality, Random variable generation, Simulation, Statistical modeling

1 Introduction

The Tukey [18] g-and-h family of non-normal distributions (see also [5], [7], [8], and [15]) are often used in Monte Carlo or statistical modeling studies. A primary advantage that this family of distributions has is that it is based on elementary transformations of standard normal deviates. The g-and-h or the simpler g or the h classes of distributions have been used in statistical modeling of extreme events or simulation studies that have included such topics as: common stock returns [1] and [2], interest rate option pricing [4], portfolio management [17], stock market daily returns [16], extreme oceanic wind speeds [3], and regression, generalized additive models, or other applications of the general linear model [10], [11], [12], [13], [14], [19], [20], and [21].

In general, however, two problems associated with any given non-normal distribution generated by the g-and-h transformation are that its probability density function (pdf) and cumulative distribution function (cdf) are unknown [5]. As such, it may be difficult to determine a g-and-h distribution's tailweight or other measures of central tendency such as a mode or trimmed mean (TM). Another problem associated with this transformation is that it is cumbersome to fit a g-and-h distribution to a set of data [2] or theoretical pdf given their specified values of skew and kurtosis [8]. See, for example, the laborious procedure used for fitting a g-and-h distribution to the $\chi^2_{df=6}$ distribution (or other data sets) given in Hoaglin et al. [8].

In view of the above, the present aim is to derive the parametric forms of the pdfs and cdfs associated with the g-and-h family of distributions. In so doing, more heuristic methods for calculating percentage points, locating measures of central tendency e.g. modes, TMs, and fitting g-and-h pdfs to data will be available to the user as opposed to other previous suggested methods in [7] and [8]. In Section 2 we develop the notation for the g-and-h family of transformations and provide the derivations of the pdfs, cdfs, and various measures of central tendency associated with these transformations. Section 3 gives the equations to calculate moments and a method for obtaining values of g and h for prespecified values of skew and kurtosis. Section 4 gives examples of fitting g-and-h distributions to real-data to demonstrate the proposed methodology. *Mathematica* [22] 6.0 notebooks are available from the first author for implementing the procedures.

2 The g-and-h, g, and h distributions

The g-and-h family considered herein is based on three transformations to produce non-normal distributions with defined or undefined moments. These transformations are computationally efficient because they only require the knowledge of the g and h parameters and an algorithm that generates standard normal pseudo-random deviates. We begin the derivation of the parametric forms of the g-and-h family of pdfs and cdfs with the following definitions.

Definition 2.1 Let Z be a random variable that has a standard normal distribution with pdf and cdf expressed as

$$f_Z(z) = (2\pi)^{-\frac{1}{2}} \exp\{-z^2/2\} \tag{1}$$

$$F_Z(z) = \Pr(Z \le z) = \int_{-\infty}^{z} (2\pi)^{-\frac{1}{2}} \exp\{-w^2/2\} dw, \quad -\infty < z < +\infty. \tag{2}$$

Let $z = (x, y)$ be the auxiliary variable that maps the parametric curves of (1) and (2) as

$$f : z \to \Re^2 := f_Z(z) = f_Z(x, y) = f_Z(z, f_Z(z)) \tag{3}$$

$$F : z \to \Re^2 := F_Z(z) = F_Z(x, y) = F_Z(z, F_Z(z)). \tag{4}$$

Definition 2.2 Let the analytical and empirical forms of the quantile function for g-and-h distributions be defined as

$$q(z) = q_{g,h}(z) = g^{-1}(\exp\{gz\} - 1) \exp\{hz^2/2\} \tag{5}$$

$$q(Z) = q_{g,h}(Z) = g^{-1}(\exp\{gZ\} - 1) \exp\{hZ^2/2\} \tag{6}$$

where $q_{g,h}(z)$ is said to be a strictly increasing monotonic function in z i.e. derivative $q'_{g,h}(z) > 0$, with parameters $g, h \in \Re$ subject to the conditions that $g \ne 0$ and $h > 0$. The parameter $\pm g$ controls the skew of a distribution in terms of both direction and magnitude. The parameter h controls the tail-weight or elongation of a distribution and is positively related with kurtosis.

Two subclasses of distributions based on (5) are the g and the h classes which are defined as

$$q(z) = q_{g,0}(z) = \lim_{h \to 0} q_{g,h}(z) = g^{-1}(\exp\{gz\} - 1) \tag{7}$$

$$q(z) = q_{0,h}(z) = \lim_{g \to 0} q_{g,h}(z) = z \exp\{hz^2/2\} \tag{8}$$

where (7) and (8) consist of asymmetric g and symmetric h distributions, respectively. By inspection of (8), it is straightforward to see that $q_{0,0}(z) = z$ and where skew and kurtosis are defined to be zero. We note that the explicit forms of the derivatives associated with (5), (7), and (8) are

$$q'(z) = q'_{g,h}(z) = \exp\{gz + (hz^2)/2\} + g^{-1}(\exp\{(hz^2)/2\}(\exp\{gz\} - 1))hz \tag{9}$$

$$q'(z) = q'_{g,0}(z) = \lim_{h \to 0} q'_{g,h}(z) = \exp\{gz\} \tag{10}$$

$$q'(z) = q'_{0,h}(z) = \lim_{g \to 0} q_{g,h}(z) = \exp\{(hz^2)/2\}(1 + hz^2). \tag{11}$$

Proposition 2.1 If the compositions $f \circ q$ and $F \circ q$ map the parametric curves of $f_{q(Z)}(q(z))$ and $F_{q(Z)}(q(z))$ where $q(z) = q(x, y)$ as

$$f \circ q : q(z) \to \Re^2 := f_{q(Z)}(q(z)) = f_{q(Z)}(q(x, y)) = f_{q(Z)}(q(z), \frac{f_Z(z)}{q'(z)}) \tag{12}$$

$$F \circ q : q(z) \to \Re^2 := F_{q(Z)}(q(z)) = F_{q(Z)}(q(x, y)) = F_{q(Z)}(q(z), F_Z(z)) \tag{13}$$

then $f_{q(Z)}(q(z), f_Z(z)/q'(z))$ and $F_{q(Z)}(q(z), F_Z(z))$ in (12) and (13) are the pdf and cdf associated with the quantile function $q(Z)$.
Proof. It is first shown that $f_{q(Z)}(q(z), f_Z(z)/q'(z))$ in (12) has the following properties:

Property 2.1 $\int_{-\infty}^{+\infty} f_{q(Z)}(q(z), f_Z(z)/q'(z))dz = 1$, and

Property 2.2 $f_{q(Z)}(q(z), f_Z(z)/q'(z)) \geq 0$, $-\infty < z < +\infty$.

To prove Property 2.1, let $y = f(x)$ be a function where $\int_{-\infty}^{+\infty} f(x)dx = \int_{-\infty}^{+\infty} ydx$. Thus, given that $x = q(z)$ and $y = f_Z(z)/q'(z)$ in $f_{q(Z)}(q(x, y))$ in equation (12) we have

$$\begin{aligned} \int_{-\infty}^{+\infty} f_{q(Z)}(q(z), f_Z(z)/q'(z))dz &= \int_{-\infty}^{+\infty} ydx = \int_{-\infty}^{+\infty} (f_Z(z)/q'(z))dq(z) \\ &= \int_{-\infty}^{+\infty} (f_Z(z)/q'(z))q'(z)dz \\ &= \int_{-\infty}^{+\infty} f_Z(z)dz = 1 \end{aligned}$$

which integrates to one because $f_Z(z)$ is the unit normal pdf. To prove Property 2.2, it is given by definition that $f_Z(z) \geq 0$ and $q'(z) > 0$. Hence, $f_{q(Z)}(q(z), f_Z(z)/q'(z)) \geq 0$ because $f_Z(z)/q'(z)$ will be nonnegative in the space of z for all $z \in (-\infty, +\infty)$ and where $\lim_{z \to \pm\infty} f_{q(Z)}(q(z), f_Z(z)/q'(z)) = 0$ because $\lim_{z \to \pm\infty} f_Z(z)/q'_{g,h}(z) = 0$, $\lim_{z \to \pm\infty} f_Z(z)/q'_{g,0}(z) = 0$, and $\lim_{z \to \pm\infty} f_Z(z)/q'_{0,h}(z) = 0$. □

A corollary to Proposition 2.1 is stated as follows
Corollary 2.1 The derivative of the cdf $F_{q(Z)}(q(z), F_Z(z))$ in (13) is the pdf

$f_{q(Z)}(q(z), f_Z(z)/q'(z))$ in (12).

Proof. It follows from $x = q(z)$ and $y = F_Z(z)$ in $F_{q(Z)}(q(x,y))$ in (13) that $dx = q'(z)dz$ and $dy = f_Z(z)dz$. Hence, using the parametric form of the derivative we have $y = dy/dx = f_Z(z)/q'(z)$ in (12). Whence, $F'_{q(Z)}(q(z), F_Z(z)) = F'_{q(Z)}(q(x, dy/dx)) = f_{q(Z)}(q(x,y)) = f_{q(Z)}(q(z), f_Z(z)/q'(z))$. Thus, $f_{q(Z)}(q(z), f_Z(z)/q'(z))$ in (12) and $F_{q(Z)}(q(z), F_Z(z))$ in (13) are the pdf and cdf associated with the empirical form of the quantile function $q(Z)$. □

In terms of measures of central tendency, the mode associated with (12) is located at $f_{q(Z)}(q(\tilde{z}), f_Z(\tilde{z})/q'(\tilde{z}))$, where $z = \tilde{z}$ is the critical number that solves $dy/dz = d(f_Z(z)/q'(z))/dz = 0$ and globally maximizes $y = f_Z(\tilde{z})/q'(\tilde{z})$ at $x = q(\tilde{z})$. We note that the pdf in (12) will have a global maximum because the standard normal density in (1) has a global maximum and the transformation $q(z)$ is a strictly increasing monotonic function by definition.

The median associated with $f_{q(Z)}(q(z), f_Z(z)/q'(z))$ in (12) is located at $q(z = 0) = 0$. This can be shown by letting $x_{0.50} = q(z)$ and $y_{0.50} = 0.50 = F_Z(z) = \Pr(Z \leq z)$ denote the coordinates in the cdf in (13) that are associated with the 50th percentile. In general, we must have $z = 0$ such that $y_{0.50} = 0.50 = F_Z(0) = \Pr(Z \leq 0)$ holds in (13) for the standard normal distribution. As such, the limit of the quantile function $q(z)$ locates the median at $\lim_{z \to 0} q(z) = 0$.

The mean and the 100γ percent symmetric TM can be obtained from using (12), the proof of Property 2.1, and from the definition of a TM as

$$E[q(z)] = \int_{-\infty}^{+\infty} q(z) f_Z(z) dz \tag{14}$$

$$\text{TM} = (1 - 2\gamma)^{-1} \int_{F_Z^{-1}(\gamma)}^{F_Z^{-1}(1-\gamma)} q(z) f_Z(z) dz \tag{15}$$

where $0 \leq h < 1$ in $q(z)$ for $E[q(z)]$ to exist in (14) and where $\gamma \in (0, 0.50)$ in (15). As $\gamma \to 0$ the TM will converge to the mean. Conversely, as $\gamma \to 0.50$ then the TM will converge to the median $q(z = 0) = 0$.

3 Moments, skew, kurtosis, and calculating values of *g* and *h*

Using equation (14) more generally, the moments for g-and-h distributions can be determined from

$$E[q(z)^k] = \int_{-\infty}^{+\infty} q(z)^k f_Z(z)dz \tag{16}$$

where $0 \leq h < 1/k$ for the k-th moment to exist. Given that the first four moments are defined, the measures of skew α_1 and kurtosis α_2 can subsequently be obtained from [9]

$$\alpha_1 = (E[q(z)^3] - 3E[q(z)^2]E[q(z)] + 2(E[q(z)])^3)/(E[q(z)^2] - (E[q(z)])^2)^{\frac{3}{2}} \tag{17}$$

$$\begin{aligned} \alpha_2 = {} & (E[q(z)^4] - 4E[q(z)^3]E[q(z)] - 3(E[q(z)^2])^2 + 12E[q(z)^2] \times \\ & (E[q(z)])^2 - 6(E[q(z)])^4)/(E[q(z)^2] - (E[q(z)])^2)^2. \end{aligned} \tag{18}$$

Using (16), (17), and (18), the formulae for the first four moments, skew, and kurtosis for g-and-h distributions are

$$E[q_{g,h}(z)] = (\exp\{g^2/(2-2h)\} - 1)/(g(1-h)^{\frac{1}{2}}) \tag{19}$$

$$\begin{aligned} E[q_{g,h}(z)^2] = {} & (1 - 2\exp\{g^2/(2-4h)\} + \\ & \exp\{2g^2/(1-2h)\})/(g^2(1-2h)^{\frac{1}{2}}) \end{aligned} \tag{20}$$

$$\begin{aligned} E[q_{g,h}(z)^3] = {} & (3\exp\{g^2/(2-6h)\} + \exp\{9g^2/(2-6h)\} - \\ & 3\exp\{2g^2/(1-3h)\} - 1)/(g^3(1-3h)^{\frac{1}{2}}) \end{aligned} \tag{21}$$

$$\begin{aligned} E[q_{g,h}(z)^4] = {} & (\exp\{8g^2/(1-4h)\}(1 + 6\exp\{6g^2/(4h-1)\} + \\ & \exp\{8g^2/(4h-1)\} - 4\exp\{7g^2/(8h-2)\} - \\ & 4\exp\{15g^2/(8h-2)\}))/(g^4(1-4h)^{\frac{1}{2}}) \end{aligned} \tag{22}$$

$$\begin{aligned} \alpha_1(g,h) = {} & [(3\exp\{g^2/(2-6h)\} + \exp\{9g^2/(2-6h)\} - \\ & 3\exp\{2g^2/(1-3h)\} - 1)/(1-3h)^{\frac{1}{2}} - 3(1 - 2\exp\{g^2/ \\ & (2-4h)\} + \exp\{2g^2/(1-2h)\})(\exp\{g^2/(2-2h)\} - 1)/ \\ & ((1-2h)^{\frac{1}{2}}(1-h)^{\frac{1}{2}}) + 2(\exp\{g^2/(2-2h)\} - 1)^3/(1-h)^{\frac{3}{2}}]/ \\ & [g^3(((1 - 2\exp\{g^2/(2-4h)\} + \exp\{2g^2/(1-2h)\})/ \\ & (1-2h)^{\frac{1}{2}} + (\exp\{g^2/(2-2h)\} - 1)^2/(h-1))/g^2)^{\frac{3}{2}}] \end{aligned} \tag{23}$$

$$\alpha_2(g,h) = [\exp\{8g^2/(1-4h)\}(1+6\exp\{6g^2/(4h-1)\} + \exp\{8g^2/(4h-1)\} - 4\exp\{7g^2/(8h-2)\} - 4\exp\{15g^2/(8h-2)\})/(1-4h)^{\frac{1}{2}} - 4(3\exp\{g^2/(2-6h)\} + \exp\{9g^2/(2-6h)\} - 3\exp\{2g^2/(1-3h)\} - 1)(\exp\{g^2/(2-2h)\} - 1)/((1-3h)^{\frac{1}{2}}(1-h)^{\frac{1}{2}}) - 6(\exp\{g^2/(2-2h)\} - 1)^4/(h-1)^2 - 12(1-2\exp\{g^2/(4h-2)\} + \exp\{2g^2/(2h-1)\})(\exp\{g^2/(2-2h)\} - 1)^2/((1-2h)^{\frac{1}{2}}(h-1)) + 3(1-2\exp\{g^2/(4h-2)\} + \exp\{2g^2/(2h-1)\})^2/(2h-1)]/[(1-2\exp\{g^2/(4h-2)\} + \exp\{2g^2/(2h-1)\})/(2h-1)^{\frac{1}{2}} + (\exp\{g^2/(2-2h)\} - 1)^2/(h-1)]^2. \tag{24}$$

Subsequently using (19) through (24), the moments, skew and kurtosis for g distributions reduce to

$$E[q_{g,0}(z)] = (\exp\{g^2/2\} - 1)/g \tag{25}$$

$$E[q_{g,0}(z)^2] = (1 - 2\exp\{g^2/2\} + \exp\{2g^2\})/g^2 \tag{26}$$

$$E[q_{g,0}(z)^3] = (3\exp\{g^2/2\} + \exp\{9g^2/2\} - 3\exp\{2g^2\} - 1)/g^3 \tag{27}$$

$$E[q_{g,0}(z)^4] = (1 - 4\exp\{g^2/2\} + 6\exp\{2g^2\} - 4\exp\{9g^2/2\} + \exp\{8g^2\})/g^4 \tag{28}$$

$$\alpha_1(g) = (3\exp\{2g^2\} + \exp\{3g^2\} - 4)^{\frac{1}{2}} \tag{29}$$

$$\alpha_2(g) = 3\exp\{2g^2\} + 2\exp\{3g^2\} + \exp\{4g^2\} - 6. \tag{30}$$

Analogously, the moments, skew, and kurtosis for the subclass of h distributions are

$$E[q_{0,h}(z)] = 0 \tag{31}$$

$$E[q_{0,h}(z)^2] = 1/(1-2h)^{\frac{3}{2}} \tag{32}$$

$$E[q_{0,h}(z)^3] = 0 \tag{33}$$

$$E[q_{0,h}(z)^4] = 3/(1-4h)^{\frac{5}{2}} \tag{34}$$

$$\alpha_1(h) = 0 \tag{35}$$

$$\alpha_2(h) = 3(1-2h)^3(1/(1-4h)^{\frac{5}{2}} + 1/(2h-1)^3). \tag{36}$$

To demonstrate the use of the methodology above, presented in Figure 1 are asymmetric and symmetric pdfs and cdfs from the g-and-h family. The values and graphs in Figure 1 were obtained using various *Mathematica* [22] functions. More specifically, the values of g and h for the asymmetric pdfs were determined by setting equations (23) and (24) to the values of $\alpha_1(g,h)$ and $\alpha_2(g,h)$ given in Figure 1, e.g. $\alpha_1(g,h) = 1$ and $\alpha_2(g,h) = 3$, and then simultaneously solved by invoking the function FindRoot. Similarly, for the symmetric distribution, (36) was set equal to $\alpha_2(h) = 10$ and then solved for h.

The graphs of the pdfs and cdfs were obtained using (12) and (13) and the graphing function ParametricPlot. The heights of the pdfs were obtained by computing the value of $\tilde{z}$ that maximizes $y = f_Z(\tilde{z})/q'(\tilde{z})$ in (12) using the function FindMaximum and the modes were then determined by evaluating $x = q(\tilde{z})$ given $\tilde{z}$. The critical values that yielded the probabilities of obtaining values of $q(z)$ in the upper 5% of the tail regions were determined by solving $\sigma q(z) + \mu - \delta = 0$ for z, where δ is the critical value, using FindRoot and then evaluating the unit normal cdf in (13) using the Erf function.

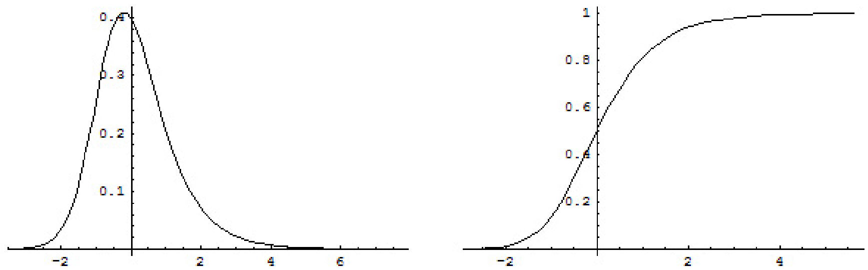
To demonstrate empirically that the solved values of g and h yield the specified values of skew and kurtosis, single samples of size $n = 2,000,000$ were drawn using the empirical forms of the g-and-h and the h quantile functions for each distribution. The sample statistics computed on the data associated with the three distributions depicted in Figure 1 were (a) $\hat{\alpha}_1 = 1.01$ and $\hat{\alpha}_2 = 3.04$, (b) $\hat{\alpha}_1 = 4.02$ and $\hat{\alpha}_2 = 39.95$, and (c) $\hat{\alpha}_1 = 0.02$ and $\hat{\alpha}_2 = 9.93$ which are all close to their respective parameter.

4 Fitting g-and-h distributions to data

Presented in Figure 2 are g-and-h pdfs superimposed on histograms of circumference measures (in centimeters) taken from the neck, chest, hip, and ankle of $n = 252$ adult males (http://lib.stat.cmu.edu/datasets/bodyfat. Inspection of Figure 2 indicates that the g-and-h pdfs provide good approximations to the empirical data. We note that to fit the g-and-h distributions to the data,

$\mu = 0.135$
$\sigma = 1.145$
$\alpha_1(g, h) = 1$
$\alpha_2(g, h) = 3$

$g = 0.244596$
$h = 0.053356$

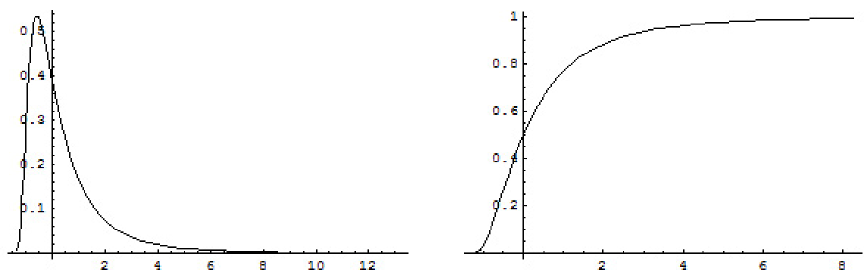


Height=0.409
Mode = -0.205

$\Pr\{q(Z) \leq q(z) = 2.177\} = 0.95$

$\mu = 0.474$
$\sigma = 1.674$
$\alpha_1(g, h) = 4$
$\alpha_2(g, h) = 40$

$g = 0.787142$
$h = 0.016356$

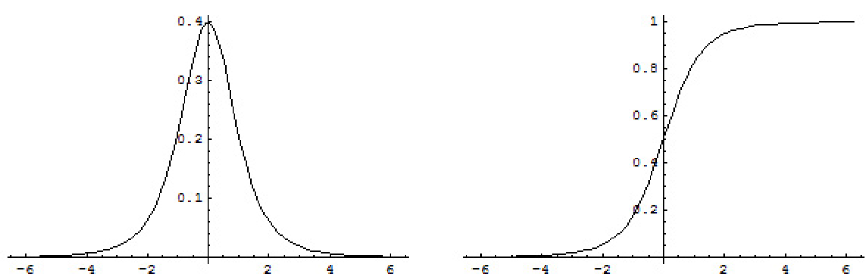


Height=0.534
Mode = -0.562

$\Pr\{q(Z) \leq q(z) = 3.442\} = 0.95$

$\mu = 0$
$\sigma = 1.346$
$\alpha_1(h) = 0$
$\alpha_2(h) = 10$

$g = 0$
$h = 0.163554$



Height=0.399
Mode = 0.0

$\Pr\{q(Z) \leq q(z) = 2.052\} = 0.95$

Figure 1: Examples of g and h parameters and their associated pdfs and cdfs.

the following linear transformation had to be imposed on $q(z)$: $Aq(z) + B$ where $A = s/\sigma$, $B = m - A\mu$, and where the values of the means (m, μ) and standard deviations (s, σ) for the data and g-and-h pdfs are given in Figure 2, respectively.

One way of determining how well a g-and-h pdf models a set of data is to compute a chi-square goodness of fit statistic. For example, listed in Table 1 are the cumulative percentages and class intervals based on the g-and-h pdf for the chest data in Panel B of Figure 2. The asymptotic value of $p = 0.153$ indicates that the g-and-h pdf provides a good fit to the data. We note that the degrees of freedom for this test were computed as [6] $df = 5 = 10$ (class intervals)-4 (parameter estimates)-1 (sample size). Further, the g-and-h TMs given in Table 2 also indicate a good fit as the TMs are all within the 95% bootstrap confidence intervals based on the data. These confidence intervals are based on 25,000 bootstrap samples.

Cumulative %	g-and-h class intervals	Observed Freq	Expected Freq
5	< 88.70	12	12.60
10	88.70 – 90.89	13	12.60
15	90.98 – 92.47	13	12.60
30	92.47 – 95.98	35	37.80
50	95.98 – 99.96	56	50.40
70	99.96 – 104.40	49	50.40
85	104.40 – 109.28	39	37.80
90	109.28 – 111.83	9	12.60
95	111.83 – 115.90	13	12.60
100	> 115.90	13	12.60
$\chi^2 = 2.015$	$\Pr\{\chi^2_5 \leq 2.015\} = 0.153$	$n = 252$	

Table 1: Observed and expected frequencies and chi-square test based on the g-and-h approximation to the chest data in Panel B of Figure 2.

DATA	PDF
$m = 37.992$	$\mu = 0.065$
$s = 2.426$	$\sigma = 1.172$
$\alpha_1 = 0.549$	$g = 0.113318$
$\alpha_2 = 2.642$	$h = 0.088872$
$m = 100.824$	$\mu = 0.108$
$s = 8.414$	$\sigma = 1.052$
$\alpha_1 = 0.677$	$g = 0.209937$
$\alpha_2 = 2.642$	$h = 0.010783$
$m = 99.905$	$\mu = 0.172$
$s = 7.150$	$\sigma = 1.248$
$\alpha_1 = 1.488$	$g = 0.293304$
$\alpha_2 = 7.300$	$h = 0.085829$
$m = 23.102$	$\mu = 0.292$
$s = 1.692$	$\sigma = 1.321$
$\alpha_1 = 2.242$	$g = 0.512894$
$\alpha_2 = 11.686$	$h = 0.038701$

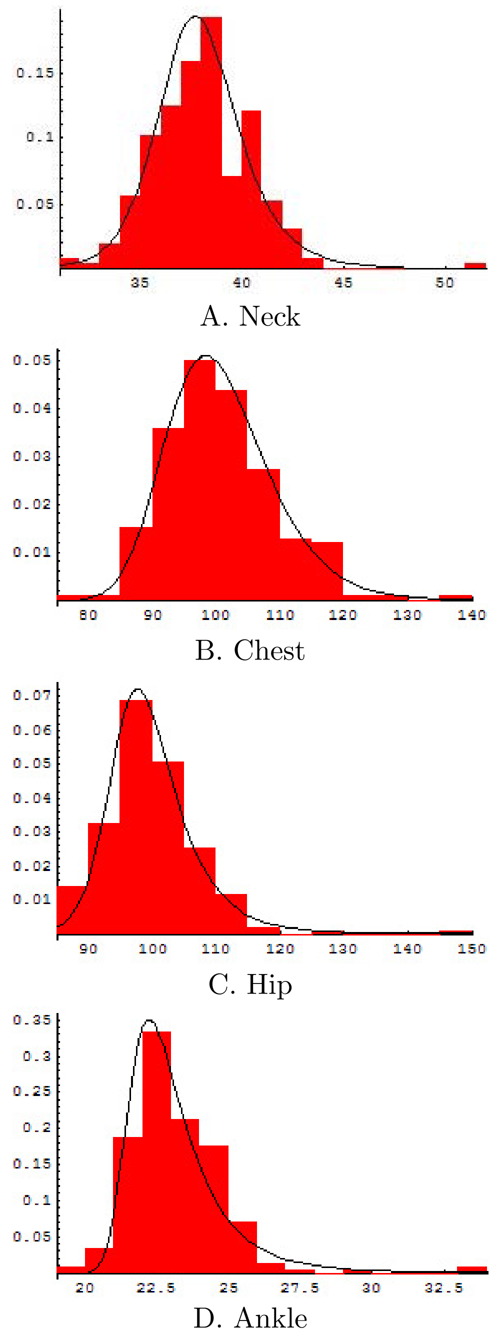


Figure 2: Examples of g-and-h pdfs' approximations to empirical pdfs using measures of circumference (in centimeters) taken from $n = 252$ men. The g-and-h pdfs were scaled using $Aq(z) + B$, where $A = s/\sigma$ and $B = m - A\mu$.

Empirical Distribution	20% TM	g-and-h TM
Neck	37.929 (37.756, 38.100)	37.899
Chest	100.128 (99.541, 100.753)	99.825
Hip	99.328 (98.908, 99.780)	99.020
Ankle	22.914 (22.798, 23.007)	22.800

Table 2: Examples of g-and-h trimmed means (TMs) based on the data in Figure 2. Each TM is based on a sample size of $n = 152$ and has a 95% bootstrap confidence interval enclosed in parentheses.

5 Comments

The ability to compute the values of g and h for prespecified values of skew and kurtosis will often times obviate the need to use the method described in Hoaglin et al. [8] such as the case for the approximation of the $\chi^2_{df=6}$ distribution. More specifically, the values of g and h for this example can be easily obtained using the method described above in Section 3. That is, setting $\alpha_1(g,h) = (8/df)^{\frac{1}{2}}$ and $\alpha_2(g,h) = 12/df$, for $df = 6$, in (23) and (24) and then solving yields $g = 0.404565$ and $h = -0.031731$. This direct approach is much more efficient than having to take the numerous steps described in [8] which also yield estimates that have less precision i.e. $g = 0.406$ and $h = -0.033$. Further, we note that the values of skew and kurtosis for this distribution will not yield a valid g-and-h pdf because h is negative.

It is also worthy to point out that the inequality given in [8] for determining where monotonicity fails for g-and-h distributions is not correct. Specifically, for the $g = 0.406$ and $h = -0.033$ distribution, Hoaglin et al. [8] submit that this g-and-h distribution loses its monotoncity at $z^2 > -1/h$ or $|z| > 5.505$ which would be correct if the distribution was a symmetric h distribution i.e. if $g = 0$. Rather, the correct values of z are determined by equating (9), not (11), to be equal to zero. As such, using the values of $g = 0.404565$ and $h = -0.031731$ from above and solving we get the (correct) values of $z = -3.692$ and $z = 12.822$ and thus $q(z = -3.692) = -1.544$ and $q(z = 12.822) = 32.406$ are the points where monotonicity fails.

References

[1] S. G. Badrinath and S. Chatterjee, On measuring skewness and elongation

in common stock return distributions: The case of the market index, *Journal of Business,* **61** (1988), 451-472.

[2] S. G. Badrinath and S. Chatterjee, A data-analytic look at skewness and elongation in common-stock return distributions, *Journal of Business & Economic Statistics,* **9** (1991), 223-233.

[3] D. J. Dupuis and C. A. Field, Large windspeeds, modeling and outlier detection, *Journal of Agricultural, Biological & Environmental Statistics,* **9** (2004), 105-121.

[4] K. K. Dutta and D. F. Babbel, Extracting probabilistic information from the prices of interest rate options: Tests of distributional assumptions, *Journal of Business,* **78** (2005), 841-870.

[5] C. A. Field and M. G. Genton, The multivariate g-and-h distribution, *Technometrics,* **48** (2006), 104-111.

[6] T. C. Headrick, Y. Sheng, and F. A. Hodis, Numerical computing and graphics for the power method transformation using Mathematica, *Journal of Statistical Software,* **19** (2007), 1-17.

[7] D. C. Hoaglin, g-and-h distributions, In S. Kotz and N. L. Johnson (Eds.), *Encyclopedia of Statistical Sciences, Vol. 3, pp. 298-301,* Wiley, New York, 1983.

[8] D. C. Hoaglin, Summarizing shape numerically; The g-and-h distributions. In D. C. Hoaglin, F. Mosteller, and J. W. Tukey (Eds.), *Exploring data, tables, trends, and shapes, pp. 461-511,* Wiley, New York, 1985.

[9] M. Kendall and A. Stuart, *The Advanced Theory of Statistics,* Macmillan, New York, 1977.

[10] H. J. Keselman, R. K. Kowalchuk, and L. M. Lix, Robust nonorthogonal analyses revisited: An update based on trimmed means, *Psychometrika,* **63** (1998), 145-163.

[11] H. J. Keselman, L. M. Lix, and R. K. Kowalchuk, Multiple comparison procedures for trimmed means, *Psychological Methods,* **3** (1998), 123-141.

[12] H. J. Keselman, R. R. Wilcox, R. K. Kowalchuk, and S. Olejnik, Comparing trimmed or least squares means of two independent skewed populations, *Biometrical Journal,* **44** (2002), 478-489.

[13] H. J. Keselman, R. R. Wilcox, J. Taylor, and R. K. Kowalchuk, Tests for mean equality that do not require homogeneity of variances: Do they really work?, *Communications in Statistics: Simulation and Computation,* **29** (2000), 875-895.

[14] R. K. Kowalchuk, H. J. Keselman, R. R. Wilcox, and J. Algina, Multiple comparison procedures, trimmed means and transformed statistics, *Journal of Modern Applied Statistical Methods,* **5** (2006), 44-65.

[15] J. Martinez, and B. Inglewicz, Some properties of the Tukey g and h family of distributions, *Communications in Statistics: Theory and Methods,* **13** (1984), 353-369.

[16] T. C. Mills, Modelling skewness and kurtosis in the London Stock Exchange FT-SE index return distributions, *The Statistician,* **44** (1995), 323-332.

[17] X. Tang, and X. Wu, A new method for the decomposition of portfolio VaR, *Journal of Systems Science and Information,* **4** (2006), 721-727.

[18] J. W. Tukey, Modern techniques in data analysis, NSF-sponsored regional research conference at Southern Massachusetts University, 1977, North Dartmouth, MA.

[19] R. R. Wilcox, Detecting nonlinear associations, plus comments on testing hypotheses about the correlation coefficient, *Journal of Educational and Behavioral Statistics,* **26**, (2001) 73-83.

[20] R. R. Wilcox, Inferences about the components of a generalized additive model, *Journal of Modern Applied Statistical Methods,* **5** (2006), 309-316.

[21] R. R. Wilcox, H. J. Keselman, and R. K. Kowalchuk, Can tests for treatment group equality be Improved? The bootstrap and trimmed means conjecture, *British Journal of Mathematical and Statistical Psychology,* **51** (1998), 123-134.

[22] S. Wolfram, *The Mathematica Book,* Wolfram Media, Inc, 2003.

Received: August 8, 2007