Simulating Univariate and Multivariate Tukey $g$- and $h$ Distributions Based on the Method of Percentiles

Tzu-Chun Kou, Southern Illinois University Carbondale
Todd C. Headrick, Southern Illinois University Carbondale

Available at: https://works.bepress.com/todd_headrick/39/
Research Article

Simulating Univariate and Multivariate Tukey $g$-and-$h$ Distributions Based on the Method of Percentiles

Tzu Chun Kuo and Todd C. Headrick

Section on Statistics and Measurement, Department of EPSE, Southern Illinois University Carbondale, P.O. Box 4618, 222-J Wham Building, Carbondale, IL 62901-4618, USA

Correspondence should be addressed to Todd C. Headrick; headrick@siu.edu

Received 2 October 2013; Accepted 19 November 2013; Published 12 January 2014

Academic Editors: M. Campanino, X. Dang, and J. Villarroel

Copyright © 2014 T. C. Kuo and T. C. Headrick. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper derives closed-form solutions for the $g$-and-$h$ shape parameters associated with the Tukey family of distributions based on the method of percentiles (MOP). A proposed MOP univariate procedure is described and compared with the method of moments (MOM) in the context of distribution fitting and estimating skew and kurtosis functions. The MOP methodology is also extended from univariate to multivariate data generation. A procedure is described for simulating nonnormal distributions with specified Spearman correlations. The MOP procedure has an advantage over the MOM because it does not require numerical integration to compute intermediate correlations. Simulation results demonstrate that the proposed MOP procedure is superior to the MOM in terms of distribution fitting, estimation, relative bias, and relative error.

1. Introduction

The Tukey $g$-and-$h$ families of univariate and multivariate nonnormal distributions are commonly used for distribution fitting, modeling events, random variable generation, and other applied mathematical contexts such as operational risk, extreme oceanic wind speeds, common stock returns, and solar flare data. See [1–17].

The family of univariate $g$-and-$h$ distributions can be summarized as follows:

\[
Y = q(Z) = q_{g,h}(Z) = g^{-1}(e^{gZ} - 1)e^{hZ^2/2},
\]

(1)

\[
Y = q(Z) = q_{g,0}(Z) = \lim_{h \to 0} q_{g,h}(Z) = g^{-1}(e^{gZ} - 1),
\]

(2)

\[
Y = q(Z) = q_{0,h}(Z) = \lim_{g \to 0} q_{g,h}(Z) = Ze^{hZ^2/2},
\]

(3)

where $Z$ is an i.i.d. standard normal random variable with probability density function (pdf), $\phi(z)$, and cumulative distribution function (cdf), $\Phi(z)$. The transformations in (1)–(3) are strictly monotone increasing functions with real-valued constants $g$ and $h$ that produce distributions defined as (i) asymmetric $g$-and-$h$ ($g \neq 0$, $h > 0$), (ii) log-normal ($g \neq 0$, $h = 0$), and (iii) symmetric $h$ ($h \geq 0$), respectively. The constant $\pm g$ controls the skew of a distribution in terms of both direction and magnitude. Taking the negative of $g$ will change the direction of the skew but not its magnitude; that is, $q_{-g,h}(Z) = -q_{g,h}(-Z)$. The constant $h$ controls the tail-weight of a distribution where the function $e^{hZ^2/2}$ (i) preserves symmetry, (ii) is increasing for $Z \geq 0$ and $h \geq 0$, and (iii) produces increased tail-weight as the value of $h$ becomes larger. In summary, (1)–(3) are computationally efficient for the purpose of generating nonnormal distributions because they only require the specification of one or two shape parameters ($g, h$) and an algorithm that produces standard normal random deviates.

The values of $g$ and $h$ associated with (1)–(3) can be determined from either the method of moments (MOM), for example, [8, 10, 13], or the method of percentiles (MOP), for example, [9, 18]. However, these two methods have disadvantages. Specifically, in the context of the MOM, the estimates of conventional skew and kurtosis associated with heavy-tailed or -skewed distributions can be substantially biased, have high variance, or can be influenced by outliers; see, for example, [19–24]. In terms of the MOP, the primary
disadvantage associated with the $g$-and-$h$ procedure is the laborious “multiplicative iterative” approach needed to determine the $h$ constant for symmetric and asymmetric nonnormal $g$-and-$h$ distributions; see [9, p. 484–489].

On the other hand, the MOP approach described by Karian and Dudewicz [24–27] in the context of the generalized lambda distribution (GLD) has demonstrated to be an attractive and computationally efficient alternative to the MOM in terms of distribution fitting and computing the GLD shape parameters. Further, it has been demonstrated that the MOP is superior to the MOM over a broad range of combinations of skew and kurtosis for fitting GLDs to theoretical or empirical distributions (see [24]).

A procedure for simulating correlated $g$-and-$h$ distributions with a specified correlation structure based on (1)–(3) is described by Kowalchuk and Headrick [10] and Headrick [28]. Specifically, the $g$- and $h$-procedure makes use of the popular NORTA (NORMal To Anything) approach. That is, the procedure begins with generating multivariate standard normal deviates prior to transformation. However, one limitation arises because the Pearson correlation is not invariant under nonlinear strictly increasing transformations. This is a concern because the transformations in (1)–(3) have this characteristic, for example, [8, 28]. Thus, the initial multivariate normal correlation structure used in the NORTA approach will not be maintained subsequent to any of the transformations in (1)–(3). As such, the NORTA procedure must begin with the computation of an intermediate correlation (IC) matrix, which is different from the specified correlation matrix between the nonnormal distributions. The purpose of the IC matrix is to adjust for the non-normalization effect of a transformation such that the $g$-and-$h$ procedure can generate nonnormal distributions with a specified correlation matrix. Further, in the context of the $g$-and-$h$ procedure, NORTA also requires numerical integration techniques to solve for the IC matrix, see [28, p. 143], which can be computationally expensive.

In contradistinction, the Spearman correlation has comparative advantages over the Pearson correlation in the context of the $g$-and-$h$ procedure for computing ICs. Specifically, the transformations in (1)–(3) must be strictly monotone increasing functions to produce valid pdfs. Thus, subsequent to transformation, the rank order of $Y$, $R(Y)$, remains the same as the rank order of $Z$, $R(Z)$; that is, the Spearman correlation remains unchanged. Moreover, there is a straightforward equation that can be used to directly solve for all pairwise ICs, for example, [28, p. 114], and thus does not require the numerical integration techniques associated with the conventional product MOM $g$-and-$h$ procedure.

View of the above, the present aim is to obviate the problems associated with the MOM in the context of the family of $g$-and-$h$ distributions in (1)–(3) by characterizing these distributions through the MOP as described in Karian and Dudewicz [27, p. 172-173]. Specifically, the purpose of this paper is to obviate the laborious “multiplicative iterative” technique given in [9] and develop the methodology and a procedure for simulating $g$-and-$h$ distributions with specified medians ($\gamma_1$), interdecile ranges ($\gamma_2$), left-right tail-weight ratios ($\gamma_3$, a skew function), and tail-weight factors ($\gamma_4$, a kurtosis function). In terms of simulating multivariate $g$-and-$h$ distributions, the Spearman correlation will be used in lieu of the Pearson correlation. In summary, the advantages of the proposed MOP procedure are that (i) the MOP parameters ($\gamma_3$, $\gamma_4$) exist for any distribution, whether the mean exists or not, for example, [25]; (ii) there are less relative bias and less relative standard error when juxtaposed with the MOM procedure; (iii) there are closed-form solutions for the $g$ and $h$ constants; (iv) there is a straightforward equation to obtain pairwise ICs for the purpose of simulating correlated nonnormal $g$-and-$h$ distributions in (1)–(3).

The remainder of the paper is outlined as follows. In Section 2, a summary of the univariate $g$-and-$h$ distributions based on the MOM is provided. In Section 3, the requisite information associated with the MOP is provided. Further, the systems of equations for determining the closed-form solutions of the $g$ and $h$ constants are subsequently derived for simulating univariate nonnormal distributions with specified values of $\gamma_1$–$\gamma_4$. In Section 4, a comparison of the MOM and the MOP is provided by fitting the SPSS data from [29]. In Section 5, the methodology for simulating correlated nonnormal $g$-and-$h$ distributions with the specified Spearman correlation matrices is provided. In Section 6, the steps for implementing the proposed MOP procedure are described. A numerical example and results of a simulation are also provided to confirm the derivations and compare the proposed procedure with the MOM procedure. In Section 7, the results of the simulation are discussed.

2. Preliminaries for the Tukey Family of $g$-and-$h$ Distributions

2.1. The Tukey Family of $g$-and-$h$ Distributions Based on the MOM. The requirement that $q(Z)$ in (1)–(3) be a strictly monotone increasing function implies that an inverse function ($q^{-1}$) exists and thus $F_{q[a]}(z) = \Phi(z)$, where $F_{q[a]}(z)$ is the general form of the cdf for both the MOM and the MOP. Differentiating both sides with respect to $q(z)$ yields

$$
dF_{q[a]}(z)/dq(z) = f_{q[a]}(z),$$

where $f_{q[a]}(z)$ is the general form of the pdf for both the MOM and the MOP. Hence,

$$
dF_{q[a]}(z)/dq(z) = (dF_{q[a]}(z)/dz)/(dz/dq(z)) = \phi(z)/q'(z).$$

Whence, the pdf integrates to one because $\phi(z)$ is the standard normal pdf and will be nonnegative on its support $q(z)$ for $z \in (-\infty, +\infty)$ given that $h \geq 0$ and where $\lim_{z \to \pm\infty} \phi(z)/q'(z) = 0$ for the transformations in (1)–(3).

The $g$ and/or $h$ constants associated with (1)–(3) that determine the shape of a distribution are computed using a moment-matching approach that involves the conventional measures of the mean ($\alpha_1$), variance ($\alpha_2$), skew ($\alpha_3$), and kurtosis ($\alpha_4$). Specifically, the values of $g$ and $h$ in (1) are determined by simultaneously solving (6) and (7) for specified values of skew and kurtosis (e.g., [28, p. 125, Eqs. 5.16-5.17], and [10, (A1)–(A4)]):

$$\alpha_1 = \frac{\exp \{g^2 / (2 - 2h)\} - 1}{g(1 - h)^{1/2}},$$

(4)
\[ \alpha_2 = \left[ \frac{\left(1 - 2 \exp \left\{ \frac{g^2}{2 - 2h} \right\} - 1 \right)^{1/2}}{(1 - 2h)^{1/2}} + \frac{\left( \exp \left\{ \frac{g^2}{2 - 2h} \right\} - 1 \right)^2}{(h - 1)^2} \right] \times g^{-2}, \]

\[ \alpha_3 = \left[ \frac{3 \exp \left\{ \frac{g^2}{2 - 6h} \right\} + 9 \exp \left\{ \frac{g^2}{2 - 6h} \right\}}{(2 - 6h)^{1/2}} - 3 \exp \left\{ \frac{g^2}{2 - 3h} \right\} - 1 \right] \times(1 - 3h)^{-1/2}
- 3 \left( 1 - 2 \exp \left\{ \frac{g^2}{2 - 4h} \right\} + \exp \left\{ \frac{2g^2}{1 - 2h} \right\} \right) \times \left( \exp \left\{ \frac{g^2}{2 - 2h} \right\} - 1 \right) \times \left( (1 - 2h)^{1/2} (1 - h)^{1/2} \right)^{-1}
+ 2(\exp \left\{ \frac{g^2}{2 - 2h} \right\} - 1)^3 \left( (1 - 2h)^{1/2} (1 - h)^{1/2} \right)^{-1}
\times g^3 \left( \left( 1 - 2 \exp \left\{ \frac{g^2}{2 - 4h} \right\} + \exp \left\{ \frac{2g^2}{1 - 2h} \right\} \right) \times (1 - 2h)^{-1/2} + \frac{\left( \exp \left\{ \frac{g^2}{2 - 2h} \right\} - 1 \right)^2}{(h - 1)} \right)^{3/2 - 1} \times g^{-2} \right], \]

\[ \alpha_4 = \left[ \exp \left\{ \frac{8g^2}{(1 - 4h)} \right\} \times \left( 1 + 6 \exp \left\{ \frac{6g^2}{4h - 1} \right\} + \exp \left\{ \frac{8g^2}{4h - 1} \right\} \right) - 4 \exp \left\{ \frac{7g^2}{(8h - 2)} \right\} - 4 \exp \left\{ \frac{15g^2}{(8h - 2)} \right\} \times (1 - 4h)^{-1/2} \right]
- 4 \left( 3 \exp \left\{ \frac{g^2}{2 - 6h} \right\} + \exp \left\{ \frac{9g^2}{(2 - 6h)} \right\} \right) - 3 \exp \left\{ \frac{2g^2}{1 - 3h} \right\} - 1 \right)
\times \left( \exp \left\{ \frac{g^2}{2 - 2h} \right\} - 1 \right) \times \left( (1 - 2h)^{1/2} (1 - h)^{1/2} \right), \]

\[ - \frac{6 \exp \left\{ \frac{g^2}{2 - 2h} \right\} - 1}{(h - 1)^2} - 12 \left( 1 - 2 \exp \left\{ \frac{g^2}{2 - 4h} \right\} + \exp \left\{ \frac{2g^2}{1 - 2h} \right\} \right) \times \frac{\left( \exp \left\{ \frac{g^2}{2 - 2h} \right\} - 1 \right)^2}{(1 - 2h)^{1/2} (1 - h)} + 3 \left( 1 - 2 \exp \left\{ \frac{g^2}{2 - 4h} \right\} + \exp \left\{ \frac{2g^2}{1 - 2h} \right\} \right)^2 \times (2h - 1)^{-1} \times \left( \exp \left\{ \frac{g^2}{2 - 4h} \right\} + \exp \left\{ \frac{2g^2}{1 - 2h} \right\} \right) \times (1 - 2h)^{-1/2} + \left( \exp \left\{ \frac{g^2}{2 - 2h} \right\} - 1 \right)^2 \times (h - 1)^{-1} \right]^2. \]

The mean and the variance for a \( g \)-and-\( h \) distribution can be determined by evaluating (4) and (5) using the solutions of \( g \) and \( h \) obtained from (6) and (7). For any \( g \)-and-\( h \) distribution, the \( k \)th moment will exist if \( 0 \leq h < \frac{1}{k} \) (see [8]). As such, (4)–(7) exist if the first four moments are defined. That is, \( 0 \leq h < \frac{1}{4} \). Note that the equations of \( \alpha_1 \)–\( \alpha_4 \) for \( g \) distributions in (2) and \( h \) distributions in (3) can be obtained from [10, Eqs. (A5)–(A8)] and [10, Eqs. (A9)–(A12)], respectively.

3. The Percentile Based \( g \)-and-\( h \) Family of Distributions

3.1. General Considerations. The percentiles (\( \theta_p \)) associated with a conventional based \( g \)-and-\( h \) pdf can be obtained by making use of the standard normal cdf, \( \Phi(z) \). As such, we will define the following location, scale, and shape parameters as in [27, p. 172-173]

\[ \gamma_1 = \theta_{0.50}, \]
\[ \gamma_2 = \theta_{0.90} - \theta_{0.10}, \]
\[ \gamma_3 = \frac{\theta_{0.50} - \theta_{0.10}}{\theta_{0.90} - \theta_{0.50}}, \]
\[ \gamma_4 = \frac{\theta_{0.75} - \theta_{0.25}}{\gamma_2}, \]

where (8) are the (i) median, (ii) interdecile range, (iii) left-right tail-weight ratio (a skew function), and (iv) tail-weight
factor (a kurtosis function), respectively. The parameters in (8) are defined to have the restrictions
\[-\infty < \gamma_1 < +\infty, \quad \gamma_2 \geq 0, \quad \gamma_3 \geq 0, \quad 0 \leq \gamma_4 \leq 1, \quad (9)\]
and where a symmetric distribution implies that \( \gamma_3 = 1 \).

3.2. The Tukey Family of \( g \)-and-\( h \) Distributions Based on the MOP. The derivation of a percentile based system of \( g \)-and-\( h \) pdfs begins by substituting the standard normal distribution percentiles \( (z_p) \) into the quantile functions in (1)–(3) and making use of (8) gives
\[
\begin{align*}
\gamma_1 &= q(z_{0.50}), \\
\gamma_2 &= q(z_{0.90}) - q(z_{0.10}), \\
\gamma_3 &= \frac{q(z_{0.50}) - q(z_{0.10})}{q(z_{0.90}) - q(z_{0.50})}, \\
\gamma_4 &= \frac{q(z_{0.75}) - q(z_{0.25})}{\gamma_2},
\end{align*}
\]
where \( z_{0.50} = 0, \ z_{0.90} = 1.281 \cdots, \ z_{0.75} = 0.6744 \cdots \) from the standard normal distribution. Note from symmetry that \( z_{0.10} = -z_{0.90} \) and \( z_{0.25} = -z_{0.75} \). The explicit forms of (10) are
\[
\begin{align*}
\gamma_1 &= 0, \\
\gamma_2 &= \frac{(1/2) \{ \exp \left[ z_{0.90} \left( h(z_{0.75} - 2g) \right) \right] \left( \exp \left[ 2gz_{0.90} \right] \right) - 1 \right\}}{g}, \\
\gamma_3 &= \exp \left[ -gz_{0.90} \right], \\
\gamma_4 &= \frac{1}{2} \left[ \left[ \exp \left( \frac{(z_{0.75} - z_{0.90}) (h(z_{0.90} + z_{0.75}) - 2g)}{2} \right) \right] \right. \\
&\quad \left. \times \left( \exp \left[ 2gz_{0.75} \right] - 1 \right) \right) \left( \coth \left( gz_{0.90} \right) - 1 \right].
\end{align*}
\]
Simultaneously, solving for the coefficients in (13) and (14) gives the convenient closed-form expressions
\[
\begin{align*}
g &= \frac{\ln(\gamma_3)}{z_{0.90}}, \\
h &= \frac{2 \ln(\left[ \gamma_3 z_{0.75} / z_{0.90} \right]^{\gamma_3 z_{0.75}/z_{0.90}} - 1) / \left( \gamma_2^2 - 1 \right) \gamma_4}{z_{0.90}^2 - z_{0.75}^2}.
\end{align*}
\]
For a symmetric distribution \( (g = 0) \), the closed-form solution for \( h \) can be expressed as
\[
h = \frac{2 \ln(z_{0.75} / (\gamma_2 z_{0.90}))}{z_{0.90}^2 - z_{0.75}^2}.
\]
Estimates of \( \gamma_1 - \gamma_4 \) based on the percentiles in (10) for a sample of size \( n \) can be determined by finding the \( j \) and \( j + 1 \) integer values and their corresponding expected values of the order statistics \( E[q(Z)_{j:n}] \) and \( E[q(Z)_{j+1:n}] \), by making use of the following equation (see [30, 31]):
\[
E[q(Z)_{j:n}] = \frac{n!}{(j-1)! \left( n-j \right)!} \left[ \int_{z_{0.10}}^{\infty} \phi(z) \Phi(z)^{-q(\gamma_2 - 1)} \left( 1 - \Phi(z) \right)^{n-j} \, dz \right]
\]
where \( \Phi(z) = q(\gamma_2 - 1) \) and \( \phi(z) = \frac{1}{\gamma_2} \left( 1 + \left( \frac{z}{\gamma_2} \right) \right) \). From Table 1, the MOP has a more accurate data fit because the relative standard errors (RSEs) of \( \bar{\gamma}_{3,4} \) are much less than those of \( \bar{\gamma}_{3,4} \). To compare the accuracy of the data fitting for the MOP and the MOM, the Euclidean distances are calculated as \( d = \sqrt{\sum(O - E)^2} \), where \( O \) is the observed proportion in each interval and \( E = \) the expected proportion in each interval for both the MOM and the MOP. From Table 1, the MOP has a more accurate data fit because the Euclidean distance \( (d) \) of the MOM is less than that of the MOM.
Table 1: Percentiles, expected proportions, observed proportions, and the Euclidean distances ($d$) for the MOP and the MOM approximations associated with the SPSS customer database data ($n = 5,000$) from [29].

<table>
<thead>
<tr>
<th>Percentile</th>
<th>Expected prop.</th>
<th>Obs. prop. (MOM)</th>
<th>Obs. prop. (MOP)</th>
<th>$d$</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.10</td>
<td>0.1240</td>
<td>0.1000</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>0.15</td>
<td>0.1208</td>
<td>0.1658</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>0.25</td>
<td>0.2392</td>
<td>0.2340</td>
<td></td>
</tr>
<tr>
<td>75</td>
<td>0.25</td>
<td>0.2564</td>
<td>0.2354</td>
<td></td>
</tr>
<tr>
<td>90</td>
<td>0.15</td>
<td>0.1632</td>
<td>0.1646</td>
<td></td>
</tr>
</tbody>
</table>

$\bar{d} = 0.04196 \quad \bar{d} = 0.03053$

Table 2: Specified correlation matrix for the distributions in Figures 2(a)–2(d).

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>0.40</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>0.60</td>
<td>0.50</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>0.65</td>
<td>0.70</td>
<td>0.60</td>
</tr>
</tbody>
</table>

Table 3: Intermediate correlations for the MOM procedure.

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>0.418072</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>0.786146</td>
<td>0.646711</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>0.906185</td>
<td>0.796341</td>
<td>0.702012</td>
</tr>
</tbody>
</table>

Table 4: Intermediate correlations for the MOP procedure ($n = 25$).

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>0.431321</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>0.638650</td>
<td>0.536062</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>0.688961</td>
<td>0.738501</td>
<td>0.638650</td>
</tr>
</tbody>
</table>

Table 5: Intermediate correlations for the MOP procedure ($n = 750$).

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>0.416344</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>0.618734</td>
<td>0.518259</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>0.668342</td>
<td>0.717483</td>
<td>0.618734</td>
</tr>
</tbody>
</table>

Table 6: The Cholesky decompositions for the MOM procedure.

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_{11}$ = 1</td>
<td>$a_{12} = 0.418072$</td>
<td>$a_{13} = 0.786146$</td>
<td>$a_{14} = 0.736275$</td>
</tr>
<tr>
<td>0</td>
<td>$a_{22} = 0.908414$</td>
<td>$a_{23} = 0.350111$</td>
<td>$a_{24} = 0.537778$</td>
</tr>
<tr>
<td>$a_{31}$ = 0</td>
<td>0</td>
<td>$a_{33} = 0.509310$</td>
<td>$a_{34} = -0.127801$</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>$a_{44} = 0.390334$</td>
</tr>
</tbody>
</table>

Table 7: The Cholesky decompositions for the MOP procedure ($n = 25$).

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_{11}$ = 1</td>
<td>$a_{12} = 0.431321$</td>
<td>$a_{13} = 0.638650$</td>
<td>$a_{14} = 0.688961$</td>
</tr>
<tr>
<td>0</td>
<td>$a_{22} = 0.902198$</td>
<td>$a_{23} = 0.288848$</td>
<td>$a_{24} = 0.489180$</td>
</tr>
<tr>
<td>$a_{31}$ = 0</td>
<td>0</td>
<td>$a_{33} = 0.713227$</td>
<td>$a_{34} = 0.080404$</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>$a_{44} = 0.528745$</td>
</tr>
</tbody>
</table>
5. The Spearman Correlations for the System of the MOP

We assume that the variates \( Y_i = q(Z_i) \) and \( Y_j = q(Z_j) \) in (1)–(3) produce valid pdfs and are thus increasing monotonic transformations in \( Z_i \) and \( Z_j \). This implies that the rank orders of \( Y_i \) and \( Y_j \) are identical and thus will have rank correlations of \( \rho_{R(Y_i),R(Y_j)} = \rho_{R(Z_i),R(Z_j)} = 1 \).

Given these assumptions, suppose it is desired to simulate a \( T \)-variate distribution from the quantile functions in (1)–(3) with a specified \( T \times T \) Spearman correlation matrix and where each distribution has specified \( \gamma_3 \) and \( \gamma_4 \). Let \( Z_i \) and \( Z_j \) have univariate normal pdfs \( \phi(z_i) \) and \( \phi(z_j) \), the Pearson correlation \( \rho_{z_i,z_j} \), and standard normal bivariate density of

\[
\begin{align*}
\hat{f}_{ij} := f_{z_i,z_j}(z_i,z_j,\rho_{z_i,z_j}) = & \left( \frac{2\pi}{1-\rho_{z_i,z_j}^2} \right)^{-\frac{1}{2}} \\
& \times \exp \left\{ -\left( 2(1-\rho_{z_i,z_j}^2) \right)^{-\frac{1}{2}} \left( z_i^2 - 2\rho_{z_i,z_j}z_i z_j + z_j^2 \right) \right\}.
\end{align*}
\]

The correlation \( \rho_{R(Y_i),R(Y_j)} \) can be obtained from the derivation of \( \rho_{R(Z_i),R(Z_j)} \) given in [32]. That is, because (1)–(3) are strictly increasing monotonic transformations, \( \rho_{R(Y_i),R(Y_j)} = \rho_{R(Z_i),R(Z_j)} \) and thus we have from Headrick [28, Eq. 4.34]

\[
\rho_{R(Y_i),R(Y_j)} = \left( \frac{6}{\pi} \right) \left[ \frac{n-2}{n+1} \sin^{-1} \left( \frac{\rho_{z_i,z_j}}{2} \right) + \frac{1}{n+1} \sin^{-1} \left( \rho_{z_i,z_j} \right) \right],
\]

(21)

where \( \rho_{R(Y_i),R(Y_j)} \) is the specified Spearman correlation and \( \rho_{z_i,z_j} \) is the intermediate correlation (IC). For a specified value of \( \rho_{R(Y_i),R(Y_j)} \) and a finite sample size \( n \), the value of the IC, \( \rho_{z_i,z_j} \), can be obtained by numerically solving (21).

6. The Procedure for Simulation and Monte Carlo Study

To implement the method for simulating \( g \) and \( h \) distributions with specified \( \gamma_3 \) and \( \gamma_4 \) and the Spearman correlations, we suggest the following six steps.

1. Specify the values of \( \gamma_3 \) and \( \gamma_4 \) for the \( T \) transformations of the forms in (1)–(3), that is, \( q_1(Z_1), \ldots, q_T(Z_T) \) and obtain the constants of \( g \) and/or \( h \) by solving (15)-(16) using the specified values of \( \gamma_3 \) and \( \gamma_4 \) for each distribution. Specify a \( T \times T \) matrix of the Spearman correlations between \( q_i(Z_j) \) and \( q_k(Z_k) \), where \( j < k \in \{1,2,\ldots,T\} \).

2. Compute the ICs \( \rho_{z_i,z_j} \) by substituting the solutions of the constants from Step 1 into (21) and then solve for \( \rho_{z_i,z_j} \). Repeat this step separately for all \( T(T-1)/2 \) pairwise combinations of correlations.

3. Assemble the ICs into a \( T \times T \) matrix and decompose this matrix using a Cholesky factorization. Note that this step requires the IC matrix to be positive definite.

4. Use the results of the Cholesky factorization from Step 3 to generate \( T \) standard normal random variables \( Z_1, \ldots, Z_T \) correlated at the intermediate levels as follows:

\[
Z_1 = a_{11}V_1
\]

\[
Z_2 = a_{12}V_1 + a_{22}V_2
\]

\[
\vdots
\]

\[
Z_j = a_{1j}V_1 + a_{2j}V_2 + \cdots + a_{jj}V_j
\]

(22)

\[
\vdots
\]

\[
Z_T = a_{1T}V_1 + a_{2T}V_2 + \cdots + a_{iT}V_i + \cdots + a_{TT}V_T,
\]

where \( V_1, \ldots, V_T \) are independent standard normal random variables and \( a_{ij} \) represents the element in the \( i \)th row and \( j \)th column of the matrix associated with the Cholesky factorization performed in Step 3.

5. Substitute \( Z_1, \ldots, Z_T \) from step 4 into T equations of the form in (1)–(3), as noted in Step 1, to generate the \( g \) and/or \( h \) distributions with the specified values of \( \gamma_3 \) and \( \gamma_4 \) and the Spearman correlations.

To demonstrate the previous steps and evaluate the proposed procedure, a comparison between the MOP and the MOM procedures is subsequently described. Specifically, the distributions in Figure 2 are used as a basis for a comparison using the specified correlation matrices in Table 2. Tables 3, 4, and 5 give the solved IC matrices for the MOM and the MOP procedures with samples of sizes 25 and 750, respectively. Note that the ICs for the MOM were computed by using the Mathematica source code as in [10, Table 1] and the ICs for the MOP are based on (21). Tables 6, 7, and 8 give the results of the Cholesky decompositions on the IC matrices, which are then used to create \( Z_1, \ldots, Z_T \) with the specified ICs by making use of the formulae given in (22) of step 4 with \( T = 4 \). The values of \( Z_1, \ldots, Z_T \) are subsequently substituted into equations of the form in (1)–(3) to produce \( q_1(Z_1), \ldots, q_T(Z_T) \) for both procedures.

In terms of the simulation, a Fortran algorithm was written for both methods to generate 25,000 independent sample.
Table 9: Skew ($\alpha_3$) and kurtosis ($\alpha_4$) results for the MOM procedure ($n = 25$).

<table>
<thead>
<tr>
<th>Dist.</th>
<th>Parameter</th>
<th>Estimate</th>
<th>95% bootstrap CI</th>
<th>Standard error</th>
<th>Relative bias %</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$\alpha_3 = 0$</td>
<td>0.0034</td>
<td>$-0.0035, 0.0065$</td>
<td>0.0036</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>$\alpha_4 = 0$</td>
<td>$-0.1785$</td>
<td>$-0.1896, -0.1673$</td>
<td>0.0056</td>
<td>$-5.93$</td>
</tr>
<tr>
<td>2</td>
<td>$\alpha_3 = 0$</td>
<td>$-0.0009$</td>
<td>$-0.0158, 0.0142$</td>
<td>0.0075</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>$\alpha_4 = 25$</td>
<td>1.297</td>
<td>$1.2668, 1.3246$</td>
<td>0.0151</td>
<td>$-94.81$</td>
</tr>
<tr>
<td>3</td>
<td>$\alpha_3 = 10$</td>
<td>1.763</td>
<td>$1.7466, 1.7778$</td>
<td>0.0082</td>
<td>$-82.37$</td>
</tr>
<tr>
<td></td>
<td>$\alpha_4 = 1000$</td>
<td>3.471</td>
<td>$3.3925, 3.5469$</td>
<td>0.0392</td>
<td>$-99.65$</td>
</tr>
<tr>
<td>4</td>
<td>$\alpha_3 = 3$</td>
<td>1.436</td>
<td>$1.4227, 1.4487$</td>
<td>0.0066</td>
<td>$-52.13$</td>
</tr>
<tr>
<td></td>
<td>$\alpha_4 = 21$</td>
<td>2.068</td>
<td>$2.0184, 2.1203$</td>
<td>0.0270</td>
<td>$-90.15$</td>
</tr>
</tbody>
</table>

Table 10: Skew ($\alpha_3$) and kurtosis ($\alpha_4$) results for the MOM procedure ($n = 750$).

<table>
<thead>
<tr>
<th>Dist.</th>
<th>Parameter</th>
<th>Estimate</th>
<th>95% bootstrap CI</th>
<th>Standard error</th>
<th>Relative bias %</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$\alpha_3 = 0$</td>
<td>$-0.0003$</td>
<td>$-0.0017, 0.0013$</td>
<td>0.0007</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>$\alpha_4 = 0$</td>
<td>$-0.0149$</td>
<td>$-0.0175, -0.0123$</td>
<td>0.0013</td>
<td>—</td>
</tr>
<tr>
<td>2</td>
<td>$\alpha_3 = 0$</td>
<td>0.0043</td>
<td>$-0.0043, 0.0141$</td>
<td>0.0045</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>$\alpha_4 = 25$</td>
<td>6.231</td>
<td>$6.1678, 6.2958$</td>
<td>0.0313</td>
<td>$-75.08$</td>
</tr>
<tr>
<td>3</td>
<td>$\alpha_3 = 10$</td>
<td>4.325</td>
<td>$4.3022, 4.3491$</td>
<td>0.0121</td>
<td>$-56.75$</td>
</tr>
<tr>
<td></td>
<td>$\alpha_4 = 1000$</td>
<td>30.82</td>
<td>$30.4145, 31.1905$</td>
<td>0.2079</td>
<td>$-96.92$</td>
</tr>
<tr>
<td>4</td>
<td>$\alpha_3 = 3$</td>
<td>2.524</td>
<td>$2.5145, 2.5335$</td>
<td>0.0051</td>
<td>$-15.87$</td>
</tr>
<tr>
<td></td>
<td>$\alpha_4 = 21$</td>
<td>10.58</td>
<td>$10.4758, 10.6732$</td>
<td>0.0506</td>
<td>$-49.62$</td>
</tr>
</tbody>
</table>

estimates for the specified parameters of (i) conventional skew ($\alpha_3$), kurtosis ($\alpha_4$), and the Pearson correlation; (ii) left-right tail-weight ratio ($\gamma_3$), tail-weight factor ($\gamma_4$), and the Spearman correlation. All estimates were based on sample sizes of $n = 25$ and $n = 750$. The formulae used for computing estimates of $\alpha_3, \alpha_4$ were based on Fisher's $k$-statistics, that is, the formulae currently used by most commercial software packages such as SAS, SPSS, and Minitab, for computing indices of skew and kurtosis (where $\alpha_4 = 0$ for the standard normal distribution). The formulae used for computing estimates of $\gamma_3, \gamma_4$ were based on (10). Note that the estimates of percentiles were based on (17). The estimate for the Pearson correlations were based on the usual formula for the Pearson product-moment of correlation statistic and the estimate for the Spearman correlations were computed based on usual formula for the Spearman rank of correlation statistic. The estimates for the Pearson and Spearman correlations were both transformed using Fisher’s $z$ transformation. Bias-corrected accelerated bootstrapped average (median) estimates, confidence intervals (CIs), and standard errors were subsequently obtained for the estimates associated with the parameters using 10,000 resamples via the commercial software package Spotfire S+ [33]. The bootstrap results for the estimates of the medians and CIs associated with the Pearson and Spearman correlations were transformed back to their original matrices (i.e., estimates for the Pearson and Spearman correlations). Further, if a parameter ($P$) was outside its associated bootstrap CI, then an index of relative bias (RB) was computed for the estimate ($E$) as $RB = \left(\frac{(E - P)}{P}\right)\times 100$. Note that the small amount of bias associated with any bootstrap CI containing a parameter was considered negligible and thus not reported. The results of the simulation are reported in Tables 9, 10, 11, and 12 and are discussed in the next section.

7. Discussion and Conclusion

One of the advantages that the MOP has over the MOM is that the MOP can be much less biased when sampling is from distributions with more severe departures from normality. Moreover, inspection of the simulation results in Tables 9 and 12 clearly indicates this result. That is, the superiority of the MOP estimates ($\gamma_3, \gamma_4$) have over their corresponding MOM counterparts ($\alpha_3, \alpha_4$). For example, with samples size of $n = 25$, the estimates of skew and kurtosis for Distribution 3 (heavy-skewed and heavy-tailed) were, on average, 82.37% and 99.65% below their associated population parameters, whereas the estimates of $\gamma_3$ and $\gamma_4$ were 3.02% and 11.06% over their respective parameters. It is also evident from comparing Tables 9 and 12 that $\gamma_3$ and $\gamma_4$ are more efficient estimators as their relative standard errors $RSE = \left(\frac{\text{standard error}}{\text{estimate}}\right)\times 100$ are considerably smaller than those of the MOM estimators of skew and kurtosis. For example, with samples size of $n = 25$, in terms of Distribution 3, inspection of Tables 9 and 12 indicates RSE measures of $RSE(\gamma_3) = 0.47\%$ and $RSE(\gamma_4) = 1.13\%$ compared with RSE ($\alpha_3$) = 0.36% and RSE ($\alpha_4$) = 0.20%. This demonstrates that $\gamma_3$ and $\gamma_4$ have more precision because they have less variance around their estimates.

Presented in Tables 13, 14, 15, and 16 are the results associated with the conventional Pearson and the Spearman correlations. Overall inspection of these tables indicates that the Spearman correlation is superior to the Pearson correlation.
### Table 11: Left-right tail-weight ratio ($\gamma_3$) and tail-weight factor ($\gamma_4$) results for the MOP procedure ($n = 25$).

<table>
<thead>
<tr>
<th>Dist</th>
<th>Parameter</th>
<th>Estimate</th>
<th>95% bootstrap CI</th>
<th>Standard error</th>
<th>Relative bias %</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$\gamma_3 = 1$</td>
<td>0.9978</td>
<td>0.9908, 1.0040</td>
<td>0.0034</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>$\gamma_4 = 0.526307$</td>
<td>0.5294</td>
<td>0.5279, 0.5310</td>
<td>0.0008</td>
<td>0.5877</td>
</tr>
<tr>
<td>2</td>
<td>$\gamma_3 = 1$</td>
<td>1.001</td>
<td>0.9940, 1.0087</td>
<td>0.0039</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>$\gamma_4 = 0.469319$</td>
<td>0.4762</td>
<td>0.4745, 0.4778</td>
<td>0.0008</td>
<td>1.4662</td>
</tr>
<tr>
<td>3</td>
<td>$\gamma_3 = 0.387801$</td>
<td>0.3995</td>
<td>0.3967, 0.4024</td>
<td>0.0014</td>
<td>3.02</td>
</tr>
<tr>
<td></td>
<td>$\gamma_4 = 0.440929$</td>
<td>0.4895</td>
<td>0.4876, 0.4913</td>
<td>0.0010</td>
<td>11.06</td>
</tr>
<tr>
<td>4</td>
<td>$\gamma_3 = 0.432409$</td>
<td>0.4452</td>
<td>0.4418, 0.4484</td>
<td>0.0017</td>
<td>2.96</td>
</tr>
<tr>
<td></td>
<td>$\gamma_4 = 0.477822$</td>
<td>0.4894</td>
<td>0.4876, 0.4913</td>
<td>0.0009</td>
<td>2.42</td>
</tr>
</tbody>
</table>

### Table 12: Left-right tail-weight ratio ($\gamma_3$) and tail-weight factor ($\gamma_4$) results for the MOP procedure ($n = 750$).

<table>
<thead>
<tr>
<th>Dist</th>
<th>Parameter</th>
<th>Estimate</th>
<th>95% bootstrap CI</th>
<th>Standard error</th>
<th>Relative bias %</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$\gamma_3 = 1$</td>
<td>1</td>
<td>0.9992, 1.0014</td>
<td>0.0005</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>$\gamma_4 = 0.526307$</td>
<td>0.5264</td>
<td>0.5261, 0.5267</td>
<td>0.0002</td>
<td>—</td>
</tr>
<tr>
<td>2</td>
<td>$\gamma_3 = 1$</td>
<td>0.9994</td>
<td>0.9980, 1.0009</td>
<td>0.0007</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>$\gamma_4 = 0.469319$</td>
<td>0.4696</td>
<td>0.4693, 0.4699</td>
<td>0.0002</td>
<td>—</td>
</tr>
<tr>
<td>3</td>
<td>$\gamma_3 = 0.387802$</td>
<td>0.3882</td>
<td>0.3876, 0.3887</td>
<td>0.0003</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>$\gamma_4 = 0.440929$</td>
<td>0.4746</td>
<td>0.4742, 0.4749</td>
<td>0.0002</td>
<td>7.64</td>
</tr>
<tr>
<td>4</td>
<td>$\gamma_3 = 0.432409$</td>
<td>0.4326</td>
<td>0.4321, 0.4332</td>
<td>0.0003</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>$\gamma_4 = 0.477822$</td>
<td>0.4784</td>
<td>0.4780, 0.4788</td>
<td>0.0002</td>
<td>—</td>
</tr>
</tbody>
</table>

### Table 13: The Pearson correlation (corr.) results for the MOM procedure ($n = 25$).

<table>
<thead>
<tr>
<th>Corr. parameter</th>
<th>Estimate</th>
<th>95% bootstrap CI</th>
<th>Standard error</th>
<th>Relative bias %</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.40</td>
<td>0.4188</td>
<td>0.4158, 0.4212</td>
<td>0.0016</td>
<td>4.69</td>
</tr>
<tr>
<td>0.60</td>
<td>0.7092</td>
<td>0.7077, 0.7108</td>
<td>0.0016</td>
<td>18.19</td>
</tr>
<tr>
<td>0.65</td>
<td>0.6879</td>
<td>0.6862, 0.6893</td>
<td>0.0015</td>
<td>5.83</td>
</tr>
<tr>
<td>0.50</td>
<td>0.5738</td>
<td>0.5718, 0.5760</td>
<td>0.0016</td>
<td>14.76</td>
</tr>
<tr>
<td>0.70</td>
<td>0.7347</td>
<td>0.7331, 0.7364</td>
<td>0.0018</td>
<td>4.96</td>
</tr>
<tr>
<td>0.60</td>
<td>0.6580</td>
<td>0.6558, 0.6605</td>
<td>0.0021</td>
<td>9.66</td>
</tr>
</tbody>
</table>

### Table 14: The Pearson correlation (corr.) results for the MOM procedure ($n = 750$).

<table>
<thead>
<tr>
<th>Corr. parameter</th>
<th>Estimate</th>
<th>95% bootstrap CI</th>
<th>Standard error</th>
<th>Relative bias %</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.40</td>
<td>0.4017</td>
<td>0.4012, 0.4022</td>
<td>0.0003</td>
<td>0.43</td>
</tr>
<tr>
<td>0.60</td>
<td>0.6337</td>
<td>0.6330, 0.6342</td>
<td>0.0006</td>
<td>5.61</td>
</tr>
<tr>
<td>0.65</td>
<td>0.6544</td>
<td>0.6541, 0.6548</td>
<td>0.0003</td>
<td>0.68</td>
</tr>
<tr>
<td>0.50</td>
<td>0.5193</td>
<td>0.5187, 0.5200</td>
<td>0.0005</td>
<td>3.86</td>
</tr>
<tr>
<td>0.70</td>
<td>0.7047</td>
<td>0.7042, 0.7051</td>
<td>0.0005</td>
<td>0.67</td>
</tr>
<tr>
<td>0.60</td>
<td>0.6173</td>
<td>0.6164, 0.6181</td>
<td>0.0007</td>
<td>2.88</td>
</tr>
</tbody>
</table>

### Table 15: The Spearman correlation (corr.) results for the MOP procedure ($n = 25$).

<table>
<thead>
<tr>
<th>Corr. parameter</th>
<th>Estimate</th>
<th>95% bootstrap CI</th>
<th>Standard error</th>
<th>Relative bias %</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.40</td>
<td>0.4138</td>
<td>0.4115, 0.4169</td>
<td>0.0017</td>
<td>3.45</td>
</tr>
<tr>
<td>0.60</td>
<td>0.6139</td>
<td>0.6115, 0.6162</td>
<td>0.0020</td>
<td>2.31</td>
</tr>
<tr>
<td>0.65</td>
<td>0.6660</td>
<td>0.6646, 0.6692</td>
<td>0.0019</td>
<td>2.46</td>
</tr>
<tr>
<td>0.50</td>
<td>0.5129</td>
<td>0.5108, 0.5161</td>
<td>0.0017</td>
<td>2.57</td>
</tr>
<tr>
<td>0.70</td>
<td>0.7156</td>
<td>0.7131, 0.7169</td>
<td>0.0019</td>
<td>2.23</td>
</tr>
<tr>
<td>0.60</td>
<td>0.6137</td>
<td>0.6115, 0.6169</td>
<td>0.0020</td>
<td>2.29</td>
</tr>
</tbody>
</table>
### Table 16: The Spearman correlation (corr.) results for the MOP procedure (n = 750).

<table>
<thead>
<tr>
<th>Corr. parameter</th>
<th>Estimate</th>
<th>95% bootstrap CI</th>
<th>Standard error</th>
<th>Relative bias %</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.40</td>
<td>0.39995</td>
<td>0.3996, 0.4005</td>
<td>0.0003</td>
<td>—</td>
</tr>
<tr>
<td>0.60</td>
<td>0.6002</td>
<td>0.5998, 0.6006</td>
<td>0.0003</td>
<td>—</td>
</tr>
<tr>
<td>0.65</td>
<td>0.6502</td>
<td>0.6499, 0.6506</td>
<td>0.0003</td>
<td>—</td>
</tr>
<tr>
<td>0.50</td>
<td>0.5004</td>
<td>0.4999, 0.5008</td>
<td>0.0003</td>
<td>—</td>
</tr>
<tr>
<td>0.70</td>
<td>0.7005</td>
<td>0.7002, 0.7008</td>
<td>0.0003</td>
<td>0.072</td>
</tr>
<tr>
<td>0.60</td>
<td>0.6000</td>
<td>0.5997, 0.6004</td>
<td>0.0003</td>
<td>—</td>
</tr>
</tbody>
</table>

### References


Impact Factor 1.730
28 Days Fast Track Peer Review
All Subject Areas of Science
Submit at http://www.tswj.com