On simulating multivariate non-normal distributions from the generalized lambda distribution

Todd C. Headrick\textsuperscript{a,}*, Abdel Mugdadi\textsuperscript{b}

\textsuperscript{a}Section on Statistics and Measurement, Department of EPSE, Southern Illinois University-Carbondale, 222-J Wham Building, Mail Code 4618, Carbondale, IL 62901, USA
\textsuperscript{b}Department of Mathematics, Southern Illinois University-Carbondale, 281 Neckers Building, Mail Code 4408, Carbondale, IL 62901, USA

Received 17 June 2005; accepted 20 June 2005
Available online 19 July 2005

Abstract

The class of generalized lambda distributions (GLDs) is primarily used for modeling univariate real-world data. The GLD has not been as popular as some other methods for simulating observations from multivariate distributions because of computational difficulties. In view of this, the methodology and algorithms are presented for extending the GLD from univariate to multivariate data generation with an emphasis on reducing computational difficulties. Algorithms written in \textit{Mathematica} 5.1 and Fortran 77 are provided for implementing the procedure and are available from the authors. A numerical example is provided and a Monte Carlo simulation was conducted to confirm and demonstrate the methodology.

© 2005 Elsevier B.V. All rights reserved.

Keywords: Correlated data; Generalized lambda distribution; Moments; Simulation

1. Introduction

The generalized lambda distribution (GLD) is a class of distributions often used for parameter estimation, fitting distributions to data, or in simulation studies that primarily involve univariate data generation (Karian and Dudewicz, 2000; Karian et al., 1996; Ramberg and Schmeiser, 1972, 1974; Ramberg et al., 1979). The univariate GLD is attrac-
Table 1

Mathematica 5.1 source code that solves for the \( \lambda_i \) for standardized GLDs

\[
x_3 = -6;
x_4 = 2000;
\]

\[
\text{FindRoot}
\begin{align*}
\lambda_1 + \frac{1}{1 + \lambda_3} - \frac{1}{1 + \lambda_4} &= 0, \\
-2\beta \left[1 + \lambda_3, 1 + \lambda_4\right] + \frac{1}{1 + 2\lambda_3} + \frac{1}{1 + 2\lambda_4} - \left(\frac{1}{1 + \lambda_3} - \frac{1}{1 + \lambda_4}\right)^2 &= 1,
\end{align*}
\]

\[
\frac{1}{\lambda_2^2} \left(3\beta \left[1 + \lambda_3, 1 + 2\lambda_4\right] - 3\beta \left[1 + 2\lambda_3, 1 + \lambda_4\right] + \frac{1}{1 + 3\lambda_3} - \frac{1}{1 + 3\lambda_4}ight)
\]

\[
+ 2\left(\frac{1}{1 + \lambda_3} - \frac{1}{1 + \lambda_4}\right)^3 - 3 \left(\frac{1}{1 + \lambda_3} - \frac{1}{1 + \lambda_4}\right) \left(-2\beta \left[1 + \lambda_3, 1 + \lambda_4\right] + \frac{1}{1 + 2\lambda_3} + \frac{1}{1 + 2\lambda_4}\right)
\]

\[
+ \frac{1}{1 + 4\lambda_3} + \frac{1}{1 + 4\lambda_4} - 3 \left(\frac{1}{1 + \lambda_3} - \frac{1}{1 + \lambda_4}\right)^4 + 6 \left(\frac{1}{1 + \lambda_3} - \frac{1}{1 + \lambda_4}\right)^2 \left(-2\beta \left[1 + \lambda_3, 1 + \lambda_4\right] + \frac{1}{1 + 2\lambda_3} + \frac{1}{1 + 2\lambda_4}\right)
\]

\[
+ \frac{1}{1 + 2\lambda_3} + \frac{1}{1 + 2\lambda_4} - 4 \left(\frac{1}{1 + \lambda_3} - \frac{1}{1 + \lambda_4}\right) \left(3\beta \left[1 + \lambda_3, 1 + 2\lambda_4\right] - 3\beta \left[1 + 2\lambda_3, 1 + \lambda_4\right]
\]

\[
+ \frac{1}{1 + 3\lambda_3} - \frac{1}{1 + 3\lambda_4}\right) = x_4
\]

\[
\begin{bmatrix}
\lambda_1, 0.60, \lambda_2, -0.55, \lambda_3, -0.10, \lambda_4, -0.045, \text{AccuracyGoal} \rightarrow 16
\end{bmatrix}
\]

Solutions: \( \lambda_1 \rightarrow 0.568751, \lambda_2 \rightarrow -0.491954, \lambda_3 \rightarrow -0.248018, \lambda_4 \rightarrow -0.0476365 \)

tive because its pdf and inverse distribution function are known (Karian and Dudewicz, 2000, p. 9) and its associated algorithm for data generation can be implemented with relative ease. For example, the GLD has been used in studies that have included such topics or techniques as: data mining (Dudewicz and Karian, 1999), independent component analysis (Karvanen, 2003; Mutihac and Van Hulle, 2003), microarray research (Beasley et al., 2004), operations research (Ganeshan, 2001), option pricing (Corrado, 2001), psychometrics (Bradley, 1993; Bradley and Fleisher, 1994; Delaney and Vargha, 2000), and structural equation modeling (Reinartz et al., 2002).

The class of GLDs was first proposed by Ramberg and Schmeiser (1972, 1974) as a generalization of Tukey’s (1960) lambda distribution. For the univariate case, a GLD is summarized by the inverse distribution function (Ramberg and Schmeiser, 1974, Eq. (1))

\[
x = \lambda_1 + \left(p^{\lambda_3} - (1 - p)^{\lambda_4}\right)/\lambda_2.
\]  

If \( p \sim U[0, 1] \) then \( x \) is a GLD where \( \lambda_1 \) and \( \lambda_2 \) are its location and scale parameters and where \( \lambda_3 \) and \( \lambda_4 \) are the shape parameters that determine its skew and kurtosis.

Presented in Table 1 is Mathematica (Wolfram, 2003, version 5.1) source code that solves for the \( \lambda_i \) in (1) for standardized GLDs. The system in Table 1 is based on Ramberg and Schmeiser’s (1974) Eqs. (3)–(6). The example in Table 1 gives the \( \lambda_i \) for a standardized GLD.
Fig. 1. Theoretical and empirical analogs of two GLDs. Panels (A) and (B) depict pdfs with shape parameters of: $\lambda_3 = 0$, $\lambda_4 = 3$ and $\lambda_3 = -6$, $\lambda_4 = 2000$. The empirical GLDs in Panels (C) and (D) are based on samples sizes of $N = 100,000$. The estimates of $\lambda_3$ and $\lambda_4$ are: $\hat{\lambda}_3 = 0.007$, $\hat{\lambda}_4 = 3.007$ (Panel C) and $\hat{\lambda}_3 = -4.464$, $\hat{\lambda}_4 = 50.512$ (Panel D).

with skew $\lambda_3 = -6$ and kurtosis $\lambda_4 = 2000$ used by Corrado (2001) to simulate non-normal security price distributions.

Fig. 1 gives theoretical (Panels A and B) and empirical (Panels C and D) GLDs with $\lambda_3 = 0$ and $\lambda_4 = 3$ (from the unit normal distribution) and $\lambda_3 = -6$ and $\lambda_4 = 2000$. The GLDs in Panels A and B are valid pdfs based on the criteria for the $\lambda_i$ given in Ramberg and Schmeiser (1974) or Karian et al. (1996). However, note the large discrepancy between the theoretical and empirical values of $\lambda_4 = 2000$ and $\hat{\lambda}_4 = 50.512$ for the distributions in Panels B and D.

The GLD has been extended to the EGLD by Karian et al. (1996) and is a combination of the GLD and the generalized beta distribution (GBD). The EGLD is useful because this class of distributions covers all possible combinations of skew ($\lambda_3$) and kurtosis ($\lambda_4$) defined for a continuous pdf to exist as $\lambda_4 > \lambda_3^2 + 1$ (see, for example, Devroye (1986, p. 688)).

In terms of multivariate data generation, it has been demonstrated that the GLD and EGLD have computational difficulties associated with (a) having to take several steps to
overcome the problem of generating biased correlation coefficients (Bradley and Fleisher, 1994), (b) finding the solutions for an appropriate mixing matrix and its associated moments for the GBD (Karvanen, 2003), or (c) having access to (or reliance on) commercial software packages (e.g., IMSL) and ensuring the accuracy of numerical solutions to complicated integrals (Corrado, 2001). For example, Bradley and Fleisher (1994) summarized their attempt of extending the GLD to multivariate data generation as: “Using the methods described in this paper to achieve highly accurate simulations of multivariate studies having non-normal distributions may seem more trouble than it is worth” (p. 165).

2. Purpose of the study

In view of the above, the primary purpose of this study is to provide the methodology and algorithms for simplifying the extension of the univariate GLD to multivariate data generation. Mathematica (Wolfram, 2003, version 5.1) notebooks are available from the authors for implementing the procedure. An alternative Fortran 77 program MULTGLD is also provided to obviate the reliance on commercial software (e.g., IMSL) for the numerical integration techniques that are required for computing intermediate correlation coefficients. MULTGLD is available from the authors as an executable program for the IBM PC or compatible platform.

3. Methodology

Let \( Z_1 \ldots Z_T \) be continuous random variables where \( Z_i \) and \( Z_j \) have univariate and bivariate pdfs defined as

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

Let the distribution functions associated with (2) and (3) be denoted as

\[
\begin{align*}
  \Phi(z_i) & = \int_{-\infty}^{z_i} (2\pi)^{-1/2} \exp \left\{ -\frac{u_i^2}{2} \right\} \, du_i, \\
  \Phi(z_j) & = \int_{-\infty}^{z_j} (2\pi)^{-1/2} \exp \left\{ -\frac{u_j^2}{2} \right\} \, du_j,
\end{align*}
\]
where $\Phi(z_i) \sim U_i [0, 1]$, $\Phi(z_j) \sim U_j [0, 1]$ with correlation $\rho_{\Phi(z_i), \Phi(z_j)} = (6/\pi) \sin^{-1} \left( \rho_{z_i z_j} / 2 \right)$ (Pearson, 1907).

Let $x_i (z_i, \lambda_{ik})$ and $x_j (z_j, \lambda_{jk})$ where $k = 1, \ldots, 4$ be standardized GLDs that take the form of (1) for the bivariate case as

$$x_i (z_i, \lambda_{ik}) = \lambda_{i1} + \left( (\Phi(z_i))^{\frac{\lambda_{i3}}{2}} - (1 - \Phi(z_i))^{\frac{\lambda_{i4}}{2}} \right) / \lambda_{i2},$$

(7)

$$x_j (z_j, \lambda_{jk}) = \lambda_{j1} + \left( (\Phi(z_j))^{\frac{\lambda_{j3}}{2}} - (1 - \Phi(z_j))^{\frac{\lambda_{j4}}{2}} \right) / \lambda_{j2},$$

(8)

It follows from (4), (7), and (8) that the correlation between $x_i (z_i, \lambda_{ik})$ and $x_j (z_j, \lambda_{jk})$ can be expressed as

$$\rho_{x_i x_j} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x_i (z_i, \lambda_{ik}) \ast x_j (z_j, \lambda_{jk})) \cdot f_{ij} \, dz_i \, dz_j$$

(9)

or alternatively as an algorithm of Reimann sums as

$$\rho_{x_i x_j} \simeq \sum \sum \left[ \left( \left( \lambda_{i1} + \left( \sum \left( f_i \Delta u_i, \left\{ u_i, u_{i\min}, z_i, \Delta u_i \right\} \right)^{\frac{\lambda_{i3}}{2}} - (1 - \sum \left( f_i \Delta u_i, \left\{ u_i, u_{i\min}, z_i, \Delta u_i \right\} \right)^{\frac{\lambda_{i4}}{2}} \right) / \lambda_{i2} \right) \ast \left( \lambda_{j1} + \left( \sum \left( f_j \Delta u_j, \left\{ u_j, u_{j\min}, z_j, \Delta u_j \right\} \right)^{\frac{\lambda_{j3}}{2}} - (1 - \sum \left( f_j \Delta u_j, \left\{ u_j, u_{j\min}, z_j, \Delta u_j \right\} \right)^{\frac{\lambda_{j4}}{2}} \right) / \lambda_{j2} \right) \right) \right] \cdot f_{ij} \Delta u_i \Delta u_j, \left\{ z_i, u_{i\min}, u_{i\max}, \Delta u_i \right\}, \left\{ z_j, u_{j\min}, u_{j\max}, \Delta u_j \right\}$$

(10)

where $u_i, u_j, z_i, z_j$ in (10) start with $u_i = u_{i\min}, u_j = u_{j\min}, z_i = u_{i\min}, z_j = u_{j\min}$ and use steps of $\Delta u_i, \Delta u_j, \Delta u_i, \Delta u_j$, respectively. We note that $\rho_{z_i z_j}$ in $f_{ij}$ in (4), (9), and (10) is referred to as the “intermediate correlation.”

Tables 2 and 3 give Mathematica (Wolfram, 2003, version 5.1) and Fortran 77 source code for numerically solving (9) and (10), respectively. The numerical examples in these tables are associated with the distributions depicted in Fig. 1 and have a specified correlation. The results in Table 2 (Table 3) indicate that an intermediate correlation of approximately $\hat{\rho}_{z_{12}} = 0.606984$ ($\hat{\rho}_{z_{12}} = 0.609231$) is required such that $x_1$ and $x_2$ have their specified correlation. It should be noted that these GLDs and correlation ($\rho_{x_{12}} = 1/2$) were also used in the study by Corrado (2001). However, Corrado (2001, p. 230) indicated that the required intermediate correlation based on (9) should be approximately $\hat{\rho}_{z_{12}} = 0.793$. 
Table 2
*Mathematica* 5.1 source code for estimating the intermediate correlation \( \hat{\rho}_{c_{12}} \) using (9)

<table>
<thead>
<tr>
<th>Specified Correlation: ( \rho_{x_1 x_2} = 1/2 )</th>
<th>Estimated Intermediate Correlation (see below): ( \hat{\rho}<em>{c</em>{12}} = 0.606984 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \hat{\lambda}_{21} = 1.27 )</td>
<td>( \hat{\lambda}_{22} = 2.34 )</td>
</tr>
<tr>
<td>( \hat{\lambda}_{11} = 0.0 )</td>
<td>( \hat{\lambda}_{12} = -0.51 )</td>
</tr>
<tr>
<td>( \hat{\lambda}_{13} = -0.51 )</td>
<td>( \hat{\lambda}_{14} = -0.47 )</td>
</tr>
<tr>
<td>( \hat{\lambda}_{21} = 0.51 )</td>
<td>( \hat{\lambda}_{22} = -1.23 )</td>
</tr>
<tr>
<td>( \hat{\lambda}_{23} = -0.47 )</td>
<td>( \hat{\lambda}_{24} = -0.49 )</td>
</tr>
</tbody>
</table>

\[
\phi_1 = \int_{-\infty}^{0} \left( \frac{1}{\sqrt{2\pi}} \right) e^{-u^2/2} du_1;
\]
\[
\phi_2 = \int_{-\infty}^{0} \left( \frac{1}{\sqrt{2\pi}} \right) e^{-u^2/2} du_2;
\]
\[
x_1 = \hat{\lambda}_{12} = \frac{\phi_1^{13} - (1 - \Phi_1)^{14}}{\hat{\lambda}_{14}};
\]
\[
x_2 = \hat{\lambda}_{22} = \frac{\phi_2^{23} - (1 - \Phi_2)^{24}}{\hat{\lambda}_{24}};
\]
\[
f_{12} = e^{\int_{\hat{\lambda}_{12}}^{\hat{\lambda}_{22}} \frac{1}{2(1-(\hat{\lambda}_{12})^2)^2} \left( \hat{\lambda}_{12} - 2(\hat{\lambda}_{12}) \hat{\lambda}_{22} \right) \frac{1}{2(1-(\hat{\lambda}_{22})^2)^2}};
\]

int = NIntegrate[(x1 x2 f12, {x1, -5, 5}, {x2, -5, 5}, Method -> MultiDimensional]
Solution: int = 0.4999997493143916

In view of the disparate intermediate correlations above, we verified that the intermediate correlations given in Tables 2 and 3 to be empirically correct. More specifically, standard normal data of sample sizes of \( N = 1,000,000 \) were drawn using *Minitab* (2000, version 13.3) and correlated at the intermediate levels (\( \hat{\rho}_{c_{12}} \)) given in Tables 2, 3, and in the Corrado (2001, p. 230) study. Using the values of \( \hat{\lambda}_{ik} \) and \( \hat{\lambda}_{jk} \) from Table 2 in (7) and (8), the resulting GLDs had correlations of \( \hat{\rho}_{x_{12}} = 0.499 \) (\( \hat{\rho}_{c_{12}} = 0.606984 \)), \( \hat{\rho}_{x_{12}} = 0.501 \) (\( \hat{\rho}_{c_{12}} = 0.609231 \)), and \( \hat{\rho}_{x_{12}} = 0.652 \) (\( \hat{\rho}_{c_{12}} = 0.793 \)). Thus, the value of \( \hat{\rho}_{c_{12}} = 0.793 \) computed by Corrado (2001, p. 230) is obviously in error. It should also be noted that Corrado (2001) did not provide empirical evidence (e.g., Monte Carlo results) to support that the GLDs had either their respective specified correlations or moments.

4. Numerical example

Suppose it is desired to generate four standardized GLDs with the specified values of skew (\( z_3 \)), kurtosis (\( z_4 \)), and the correlation matrix given in Tables 4 and 5. Table 6 gives the required values of the \( \hat{\lambda}_{ik} \) solved from the system of equations in Table 1. Given these specifications, presented in Table 7 are the estimates of the intermediate correlations (\( \hat{\rho}_{c_{12}} \)) solved by both (9) and (10). The intermediate correlations solved by (10) used partitioning steps of \( \Delta u_i = \Delta u_j = 0.01 \) and are the values enclosed in parentheses.
Table 3 Fortran 77 source code for estimating the intermediate correlation \( \rho_{z_1z_2} \) using (10)

```
REAL C11,C12,C13,C14,C21,C22,C23,C24,PI,RHO,DELTA,
&     U1MIN,U1MAX,U2MIN,U2MAX,Z1,Z2,INT
C*******************************************************************************
C SPECIFIED CORRELATION: \( \rho_{z_1z_2} = 1/2 = \text{INT} = 0.4999997085564828 \)
C ESTIMATED INTERMEDIATE CORRELATION: \( \hat{\rho}_{z_1z_2} = \text{RHO} = 0.609231 \)
C SPECIFICATIONS: \( \lambda_k = C_k, \hat{\lambda}_k = C_k, k = 1, \ldots, 4 \) (SEE TABLE 2)
C U1MIN = U2MIN = -5.0; U1MAX = U2MAX = 5.0; \( \Delta u_1 = \Delta u_2 = \text{DELTA} = 0.01 \)
C*******************************************************************************
PI = ATAN(1.0)*4.0
INT = 0.0
DO 10 Z1 = U1MIN,U2MAX,DELTA
DO 20 Z2 = U2MIN,U2MAX,DELTA
INT = INT+SUBTOT (C11,C12,C13,C14,C21,C22,C23,C24,PI,RHO,
&     U1MIN,U2MIN,DELTA,Z1,Z2)*DELTA*DELTA
20 CONTINUE
10 CONTINUE
END
C*******************************************************************************
FUNCTION SUBTOT (C11,C12,C13,C14,C21,C22,C23,C24,PI,RHO,
&     U1MIN,U2MIN,DELTA,Z1,Z2)
REAL CON,SUBTOT,X1,X2
X1 = C11+(PHI1 (PI,U1MIN,Z1,DELTA)*C13-
&     (1-PHI1 (PI,U1MIN,Z1,DELTA))*C14)/C12
X2 = C21+(PHI2 (PI,U2MIN,Z2,DELTA)*C23-
&     (1-PHI2 (PI,U2MIN,Z2,DELTA))*C24)/C22
CON = ((X1*X2)*((2*PI)*(SQRT(1-RHO**2)))*((-1)))*
&     EXP((-1/(2* (1-RHO**2))*(Z1**2)-2*RHO*(Z1*Z2)+(Z2**2))))
SUBTOT = CON
RETURN
END
C*******************************************************************************
FUNCTION PHI1 (PI,U1MIN,Z1,DELTA)
REAL PHI1,SUM1,U1
SUM1 = 0.0
DO 30 U1 = U1MIN,Z1,DELTA
SUM1 = SUM1+1.0/(SQRT(2.0*PI))*EXP (- (U1**2)/2.0)*DELTA
30 CONTINUE
PHI1 = SUM1
RETURN
END
C*******************************************************************************
FUNCTION PHI2 (PI,U2MIN,Z2,DELTA)
REAL PHI2,SUM2,U2
SUM2 = 0.0
DO 40 U2 = U2MIN,Z2,DELTA
SUM2 = SUM2+1.0/(SQRT (2.0*PI))*EXP (- (U2**2)/2.0)*DELTA
40 CONTINUE
PHI2 = SUM2
RETURN
END
```
Table 4
Specified moments \((m_i)\) for the GLDs \((x_i)\)

<table>
<thead>
<tr>
<th>GLD</th>
<th>(m_1)</th>
<th>(m_2)</th>
<th>(m_3)</th>
<th>(m_4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(x_1)</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>(x_2)</td>
<td>0</td>
<td>1</td>
<td>1/2</td>
<td>4</td>
</tr>
<tr>
<td>(x_3)</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>(x_4)</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>12</td>
</tr>
</tbody>
</table>

Table 5
Specified correlation structure \((\rho_{x_i,x_j})\) for the GLDs

<table>
<thead>
<tr>
<th>GLD</th>
<th>(x_1)</th>
<th>(x_2)</th>
<th>(x_3)</th>
<th>(x_4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(x_1)</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(x_2)</td>
<td>.30</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(x_3)</td>
<td>.40</td>
<td>.60</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>(x_4)</td>
<td>.50</td>
<td>.70</td>
<td>.80</td>
<td>1</td>
</tr>
</tbody>
</table>

The data generation procedure begins by factoring (e.g., a Cholesky factorization) the intermediate correlation matrix in Table 7 (based on either equation 9 or 10). The entries in the factored matrix are then used to produce standard normal deviates that are correlated at their intermediate levels \(\rho_{\pm z_i z_j}\). To reduce computer run-time, these standard normal deviates are subsequently transformed to uniform deviates by the following approximation to the integrals of the form in (5) from Bagby (1995)

\[
\Phi (\pm z_i) = \frac{1}{2} + \frac{1}{2} \left\{ 1 - \frac{1}{30} \left[ 7e^{-\left(\pm z_i\right)^2/2} + 16e^{-\left(\pm z_i\right)^2\left(2-\sqrt{2}\right)} + \left(7 + \frac{1}{4}\pi (\pm z_i)^2\right)e^{-\left(\pm z_i\right)^2}\right]\right\}^{1/2} + \epsilon (\pm z_i),
\]

where the absolute error is \(\epsilon (\pm z_i) < 3.04 \times 10^{-5}\). The uniform deviates generated from equations of the form in (11), having correlations of \(\rho_{\Phi(z_i),\Phi(z_j)} = (6/\pi) \sin^{-1}\left(\rho_{z_i z_j}/2\right)\), are then transformed by equations of the form in (7) to produce GLDs with their specified shapes and correlations \(\rho_{x_i,x_j}\).

5. Monte Carlo simulation

To evaluate the proposed procedure, the specified parameters of moments and correlations from the numerical example listed in Tables 4 and 5 were simulated using an algorithm coded in Fortran 77. The algorithm employed the use of subroutines UNI1 and NORMB1 from RANGEN (Blair, 1987) to generate pseudo-random uniform and standard normal
Table 6
Solutions of $\hat{\lambda}_{ik}$ based on the system of equations in Table 1 for the specified GLDs ($x_i$)

<table>
<thead>
<tr>
<th>GLD</th>
<th>$\hat{\lambda}_{i1}$</th>
<th>$\hat{\lambda}_{i2}$</th>
<th>$\hat{\lambda}_{i3}$</th>
<th>$\hat{\lambda}_{i4}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_1$</td>
<td>0.000000</td>
<td>0.197451</td>
<td>0.134912</td>
<td>0.134912</td>
</tr>
<tr>
<td>$x_2$</td>
<td>−0.291041</td>
<td>0.0603855</td>
<td>0.0258673</td>
<td>0.0447025</td>
</tr>
<tr>
<td>$x_3$</td>
<td>−0.378520</td>
<td>−0.0561934</td>
<td>−0.0187381</td>
<td>−0.038800</td>
</tr>
<tr>
<td>$x_4$</td>
<td>−0.579659</td>
<td>−0.142276</td>
<td>−0.0272811</td>
<td>−0.0995192</td>
</tr>
</tbody>
</table>

Table 7
Estimates of intermediate correlations $\left(\hat{\rho}_{zi,zj}\right)$ calculated by (9) and (10) for the correlation structure in Table 5

<table>
<thead>
<tr>
<th>$z_1$</th>
<th>$z_2$</th>
<th>$z_3$</th>
<th>$z_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$z_1$</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$z_2$</td>
<td>0.302245 (0.302177)</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>$z_3$</td>
<td>0.409561 (0.409249)</td>
<td>0.610044 (0.608950)</td>
<td>1</td>
</tr>
<tr>
<td>$z_4$</td>
<td>0.536005 (0.534750)</td>
<td>0.732598 (0.730390)</td>
<td>0.823539 (0.820815)</td>
</tr>
</tbody>
</table>

The estimates computed by (10) are enclosed in parentheses.

Table 8
Empirical estimates of the specified moments in Table 4 using single draws of size $N = 1,000,000$

<table>
<thead>
<tr>
<th>GLD</th>
<th>$\hat{\mu}_1$</th>
<th>$\hat{\mu}_2$</th>
<th>$\hat{\mu}_3$</th>
<th>$\hat{\mu}_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_1$</td>
<td>−0.000267</td>
<td>1.000267</td>
<td>0.000381</td>
<td>2.996471</td>
</tr>
<tr>
<td>$x_2$</td>
<td>0.000061</td>
<td>1.000147</td>
<td>0.503183</td>
<td>4.004508</td>
</tr>
<tr>
<td>$x_3$</td>
<td>0.000273</td>
<td>0.999419</td>
<td>0.999521</td>
<td>6.003119</td>
</tr>
<tr>
<td>$x_4$</td>
<td>−0.001382</td>
<td>1.000213</td>
<td>2.004192</td>
<td>12.004460</td>
</tr>
</tbody>
</table>

deviates. Independent sample sizes of $N = 10$ and $N = 100$ were generated for simulating the specified moments and correlations. The averaging procedure used to compute the estimates of moments ($\hat{\mu}_i$) and correlations ($\hat{\rho}_{xi,xj}$) were based on 50,000 replications and thus $N \times 50,000$ random deviates. Single draws of size of $N = 1,000,000$ were also generated and the same sample statistics were computed on these data.

The overall average estimates of the moments and correlations obtained from the simulation (based on $N = 10$ and $N = 100$) were in close agreement with all specified parameters. However, while the long-run behavior of these estimates is important, the procedure used to evaluate this behavior may generate biased sample estimates yet still converge in close proximity to their associated population parameters. In view of this concern, presented in Tables 8 and 9 are the estimates of the parameters based on the single draws of $N = 1,000,000$. 
Table 9
Empirical estimates of the specified correlations in Table 5 using single draws of size $N = 1,000,000$

<table>
<thead>
<tr>
<th>GLD</th>
<th>$x_1$</th>
<th>$x_2$</th>
<th>$x_3$</th>
<th>$x_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_1$</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$x_2$</td>
<td>0.299014 (0.298945)</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$x_3$</td>
<td>0.400203 (0.399893)</td>
<td>0.599711 (0.598611)</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>$x_4$</td>
<td>0.500512 (0.499352)</td>
<td>0.701577 (0.699411)</td>
<td>0.801004 (0.798179)</td>
<td>1</td>
</tr>
</tbody>
</table>

The estimates based on (10) are enclosed in parentheses.

Inspection of Tables 8 and 9 indicates that these estimates were also in close agreement to their respective population parameters.

6. Concluding remarks

Monte Carlo and simulation techniques are widely used in statistical research. Because real-world data sets can often be radically non-normal (see, for example, Micceri, 1989; Sawilowsky and Blair, 1992), it is essential that statisticians have a variety of techniques available for univariate or multivariate non-normal data generation. The GLD has not been as popular as some other competing methods e.g., the power method transformations (see Fleishman, 1978; Headrick, 2002; Headrick and Beasley, 2004; Headrick and Sawilowsky, 1999; Vale and Maurelli, 1983) for generating multivariate non-normal distributions. More specifically, in terms of the power method’s extension from univariate to multivariate data generation Kotz et al. (2000) noted: “this [power] method does provide a way of generating bivariate non-normal random variables. Simple extensions of other univariate methods are not available yet” (p. 37). However, it is our belief that the extension of the GLD from univariate to multivariate data generation, as presented in this paper, makes the GLD a viable competitor to the power method because of its simplicity and ease of execution.

References

Blair, R.C., 1987. RANGEN. IBM, Boca Raton, FL.


