A method to generate multivariate data with the desired moments

<table>
<thead>
<tr>
<th>Journal:</th>
<th>Communications in Statistics - Simulation and Computation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscript ID:</td>
<td>LSSP-2007-0081.R1</td>
</tr>
<tr>
<td>Manuscript Type:</td>
<td>Original Paper</td>
</tr>
<tr>
<td>Date Submitted by the Author:</td>
<td>28-Apr-2008</td>
</tr>
<tr>
<td>Complete List of Authors:</td>
<td>Lyhagen, Johan; Division of Statistics, Department of Information Science</td>
</tr>
<tr>
<td>Keywords:</td>
<td>Bootstrap, Monte Carlo, Skewness</td>
</tr>
</tbody>
</table>

Abstract: We show how it is possible to generate multivariate data which has moments arbitrary close to the desired ones. They are generated as linear combinations of variables with known theoretical moments. It is shown how to derive the weights of the linear combinations in both the univariate and the multivariate setting. The use in bootstrapping is discussed and the method is exemplified with a Monte Carlo simulation where the importance of the ability of generating data with control of higher moments is shown.

Note: The following files were submitted by the author for peer review, but cannot be converted to PDF. You must view these files (e.g. movies) online.

revcssc.tex
1 Introduction

The generation of data is of crucial importance in many scientific fields. E.g. in statistics generated data is used to investigate the statistical properties of estimation methods and test statistics while in empirical research in e.g. econometrics or psychometrics computer intensive methods such as the bootstrap crucially depend on the generation of data. For a general treatment of generating data see Devroye (1986) and for the bootstrap Efron and Tibshirani (1993).

The properties of the data generated should mimic the desired ones and the question is how to generate data with these characteristics. It is very convenient to assume (multivariate) normality but as noted by many, e.g. Fleishman (1978), the distribution of real world data typically characterized by skewness and kurtosis which deviate from those of the normal distribution. Fleishman (1978) proposed to generate data according to

\[ Y = a + bX + cX^2 + dX^3 \]  

where \( X \) is standard normally distributed. The distribution of \( Y \) is generally unknown but it is sometimes possible to choose \( a, b, c \) and \( d \) in such a way that \( Y \) has the desired first four moments. Tadikamalla (1980) criticizes this method as there are some combinations of skewness and kurtosis which are impossible to reach and reviews methods that are better in this respect. These methods are more complicated than the one above and most likely the Fleishman method should be used when the desired skewness and kurtosis lie within the feasible area, and one of the methods reviewed by Tadikamalla(1980) otherwise. One advantage of the Fleishman method is that it is easily generalized to the
multivariate setting as done by Vale and Maurelli (1983). Define $x$ as

$$x = [1, X, X^2, X^3]^T$$

(2)

and the vector of weights

$$w = [a, b, c, d]^T$$

(3)

then

$$Y = w^T x$$

(4)

Assume without loss of generality $EY_1 = EY_2 = 0$ then the covariance of two variables achieved this way is,

$$\rho_{Y_1, Y_2} = E(Y_1 Y_2) = E(w^T x_1 x_2^T w_2) = w^T R w_2$$

(5)

where $R = E(x_1 x_2^T)$. The (third-degree polynomial) relationship between $\rho_{Y_1, Y_2}$ and the covariance of $X_1$ and $X_2$, expressed with the weights is

$$\rho_{Y_1, Y_2} = \rho_{X_1, X_2} (b_1 b_2 + 3b_1 d_2 + 3d_1 b_2 + 9d_1 d_2)$$

$$+ \rho^3_{X_1, X_2} (2c_1 c_2) + \rho^3_{X_1, X_2} (6d_1 d_2)$$

(6)

where $\rho_{X_1, X_2}$ is the covariance of $X_1$ and $X_2$. The proposed working order is to first decide the weights such that the first four moments are the desired and then solve (6) with respect to $\rho_{X_1, X_2}$ given the weights and $\rho_{Y_1, Y_2}$.

This method to generate multivariate data suffers from some drawbacks: i) It is a very tedious way if more than 2 variables are to be generated, ii) As observed by Tadikamalla (1980), not all combinations of skewness and kurtosis are
possible, \( iii \)) Some moments are ignored, e.g. \( EY_1^2Y_2 \) and \( EY_1^2Y_2^2 \) which might be of interest. In this paper a general method to generate multivariate data with arbitrary moments is proposed which does not suffer from the drawbacks outlined for the Vale and Maurelli method.

The paper is as follows: Section 2 presents the proposed method and Section 3 generalizes it to deal with the generation of multivariate data with arbitrary moments. How to choose between different competing distributions which the linear combination is built is the topic in Section 4. The use of this method in the bootstrap is discussed in Section 5 and includes a small Monte Carlo simulation. A conclusion ends the paper.

2 The generation of data with arbitrary moments

The first and second (central) moments, i.e. mean and variance, of linear combinations are well known and repeatedly taught at basic statistics courses. Let \( X_i \) be a random variable, with sufficiently many moments existing, and \( a_i \) the associated weight and

\[
Y = a_0 + \sum_{i=1}^{p} a_i X_i
\]

The expectation is

\[
EY = a_0 + \sum_{i=1}^{p} a_i EX_i = a_0 + \sum_{i=1}^{p} a_i \mu_{X_i}
\]
Similarly, the variance is

\[ V(Y) = E \left( a_0 + \sum_{i=1}^{p} a_i X_i - a_0 - \sum_{i=1}^{p} a_i \mu_{X_i} \right)^2 \]

\[ = \sum_{i=1}^{p} \sum_{j=1}^{p} a_i a_j \text{Cov}(X_i, X_j) \quad (9) \]

In general, the \( k \)th central moment for \( Y \) is

\[ E(Y - \mu_Y)^k = \sum_{i_1=1}^{p} \sum_{i_2=1}^{p} \cdots \sum_{i_k=1}^{p} a_{i_1} a_{i_2} \cdots a_{i_k} \mu_{i_1 i_2 \cdots i_k} \quad (10) \]

where \( \mu_{i_1 i_2 \cdots i_k} = E \prod_{i=i_1}^{i_k} (X_i - \mu_X) \).

To generate \( Y \), values of \( a_i \) must be decided upon such that the desired moments are obtained. For notational convenience \( X \) denotes the set \( X_1, X_2, \ldots, X_p \).

Let \( \mathbf{m}_Y \) be a vector of desired moments and \( \mathbf{m}_X \) a vector of moments (which may be function of the parameters) of \( X \). \( \mathbf{m}_X \) may also contain the number 1 to be able to get the wanted mean of \( Y \). The relationship between the two vectors is

\[ \mathbf{m}_Y = A \mathbf{m}_X \quad (11) \]

where \( A \) is a matrix with elements that are functions of \( a_i \). The precise contents of \( \mathbf{m}_Y, \mathbf{m}_X \) and \( A \) depends on which moment’s one wants to mimic and the distribution of the \( X \). For example, if the first three moments of \( Y \) are of interest and \( X \) is a \( \chi^2 \) variable with \( \lambda \) degrees of freedom, then:

\[ \mathbf{m}_Y = \left[ \begin{array}{c} \mu_Y \\ E(Y - \mu_Y)^2 \\ E(Y - \mu_Y)^3 \end{array} \right]^T \quad (12) \]

\[ A = \left[ \begin{array}{ccc} a_0 & 0 & 0 \\ a_1 & 2a_1^2 & 8a_1^3 \end{array} \right]^T \quad (13) \]

\[ \mathbf{m}_X = \left[ \begin{array}{c} 1 \\ \lambda \end{array} \right]^T \quad (14) \]
where we used the moment structure of the $\chi^2$ distribution, i.e. $EX = \lambda$, $V(X) = 2\lambda$ and $E(X - \lambda)^3 = 8\lambda$. Note that we are interested in three moments and there are three unknowns to use, $a_0$, $a_1$ and $\lambda$. To derive the unknown parameters it is useful to minimize the squared deviations from the implied and the desired, i.e. let

$$f = (m_Y - Am_X)^T (m_Y - Am_X) \quad (15)$$

and then minimize $f$ with respect to the unknown parameters. This is simple to do with some numerical optimization technique commonly implemented in programming languages such as Gauss or MATLAB. If the minimum of $f$ is zero then the weights give a $Y$ with exactly the wanted moments. If $p$ is larger than necessary then an appropriate optimization method will give an indication of that as the covariance matrix would in that case not be of full rank, i.e. in one direction of the space spanned by the parameters the function is flat. If the minimum of $f$ is not zero then it is not possible to get the wanted moments with the chosen $p$ and distribution of $X$. Note that the minimum needs not be unique. Depending on the situation increasing $p$ or changing the distribution used of the $X$ is to be preferred. E.g. if a non-zero third central moment of $Y$ is wanted and $X$ is normally distributed then increasing $p$ with another normal will obviously not solve the problem as a linear combination of normal variates is also normal, hence the third central moment of $Y$ equals zero. If the minimum of $f$ is zero then deriving the variance of the moments of interest can be used to choose suitable distributions and $p$, see Section 4. If the minimum is not zero the mean squared error can be used instead of the variance.
If, for some reason, one wants to stick to a particular \( p \) and distribution of \( X \) then a better choice of function to minimize is

\[
f = (m_Y - Am_X)^T \Sigma^{-1} (m_Y - Am_X)
\]  

(16)

where a typical element of \( \Sigma \) is the covariance between the desired moments, see Cramer (1946) or Kendall and Stuart (1969) for specific formulas on the covariances. Minimizing this minimizes the Kullback-Leibler distance, see Kullback and Leibler (1951), between the desired distribution of \( Y \) and the implied, given by the weights and the distribution of \( X \), and results in a variable that is in that sense optimal for the chosen set of moments. A subsequent section discusses the generation of data for the bootstrap where it might not be possible to gain a minimum of zero.

3 The multivariate case

Frequently it is of interest to generate a multivariate set of observations with a certain connection, commonly measured with covariances, but other measures such as third or fourth cross moments are possible too. The above is easily generalized to generate multivariate data. Now, for \( j = 1, \ldots, m \), we have

\[
Y_j = a_{0j} + \sum_{i=1}^n a_{ij} X_i
\]  

(17)

and let \( m_Y \) be a vector of all moments of interest for the multivariate distribution of \( Y \), i.e. also the cross-moments, and let \( m_X \) be as above. Then by minimizing \( f \) which now is a function of a larger set of moments and weights the optimal weights and moments are gained. Note that \( Y_j \) and \( Y_i \) may or may
not be linear combinations of the same components of $X$. This produces a very
general structure for the moments.

The method of Fleishman, and for that case also the multivariate version of
Vale and Maurelli, is a special case of the above proposed method for generating
multivariate data. To see this consider the bivariate case, remember that $X$
is multivariate normal and let $p = 3$. For ease of notation, for $i = 1, 2$, let

$$m_Y = E \left[ Y_i, Y_i^2, Y_i^3 \right]^T, \quad m_X = E \left[ 1, X_i, X_i^2, \ldots, X_i^{12} \right]^T$$

and

$$a_{yi} = \begin{bmatrix}
    a & a^2 & a^3 & a^4 \\
    2ab & 3a^2b & 4a^3b & \\
    3b^2a + 3b^2c & 3ab^2 + 3ab^2c & 4ab^3 + 12ab^2c + 4ab^3 \\
    0 & 2bc + 6ab + 3ac^2 + 3b^2c & 2bc^2 + 12abc + 4b^3c \\
    4bc^2d + 6b^2d + 3bc^2 & 4b^2c + 12b^2c + 4b^3c & 4c^3 \\
    0 & 3b^2 + 3c^2d & 12ab + 12abc + 4d^3a \\
    0 & 3cd^2 & 6d^2b^2 + 12d^2ac + 12c^2db + c^4 \\
    0 & d^3 & 12d^2b + 4d^2c + 4d^3a \\
    0 & 0 & 4d^3b + 6c^2d^2 \\
    0 & 0 & d^4
\end{bmatrix}$$

where $a = a_{0i}$, $b = a_{1i}$, $c = a_{2i}$ and $d = a_{3i}$ in (17), and where the index $i$ for
the weights is suppressed to save space. Then $m_Y$, $A$ and $m_X$ are as follows:
\[ \mathbf{m}_Y = \begin{bmatrix} \mathbf{m}_{Y_1}^T & \mathbf{m}_{Y_2}^T & \rho_{Y_1,Y_2} \end{bmatrix}^T \]  
(18)

\[ \mathbf{A} = \begin{bmatrix} a_{y_1} & 0 & 0 & 0 & 0 \\ 0 & a_{y_2} & 0 & 0 & 0 \\ 0 & 0 & b_1b_2 + 3b_1d_2 + d_1b_2 + 9d_1d_2 & 2c_1c_2 & 6d_1d_2 \end{bmatrix} \]  
(19)

\[ \mathbf{m}_X = \begin{bmatrix} \mathbf{m}_{X_1}^T & \mathbf{m}_{X_2}^T & \rho_{X_1,X_2} & \rho_{X_1,X_2}^2 & \rho_{X_1,X_2}^3 \end{bmatrix}^T \]  
(20)

where the zeros in \( \mathbf{A} \) are matrices of zeros of suitable size. Minimizing \( f \) would yield the same weights as the method of Vale and Maurelli but in one step and are in that way simpler and more intuitive.

## 4 Choosing amongst competing distributions

It is well known that there are an infinite number of distributions that give the same set of finite moments. Hence, choosing the distribution of the elements of \( X \) is non-trivial as the sampling properties depend on the choice. Assume there are for convenience \( L \) competing sets of \( X \) vectors which give \( Y \) with the same desired moments through the associated \( \mathbf{A}_l \), \( l = 1, \ldots, L \). It is desirable that the sampling variability of the moments of interest of \( Y \) is as small as possible. Let \( \Sigma \) be the covariance matrix of \( \mathbf{m}_Y \) then the best \( X \) would be the one which has the smallest \( \Sigma \) using some measure. One natural choice would be the determinant, \( |\Sigma| \), referred to as the generalized variance by Mardia et al (1983). It is also possible to use the trace of \( \Sigma \) i.e. the total variation.

For example, let \( X \) be log normally distributed and as in the example above the three first moments are of interest then the three matrices are:
\[
\mathbf{m}_Y = \begin{bmatrix}
\mu_Y & E(Y - \mu_Y)^2 & E(Y - \mu_Y)^3
\end{bmatrix}^T
\]  \hspace{1cm} (21)
\[
\mathbf{A} = \begin{bmatrix}
a_0 & a_1 & 0 & 0 \\
0 & 0 & a_1^2 & 0 \\
0 & 0 & 0 & a_1^3
\end{bmatrix}
\]  \hspace{1cm} (22)
\[
\mathbf{m}_X = \begin{bmatrix}
1 \\
\exp(\mu + \sigma^2/2) \\
\exp(2\mu + 2\sigma^2) - \exp(2\mu + \sigma^2) \\
\exp(3\mu + 3\sigma^2) - 3\exp(3\mu + 2\sigma^2) + 2\exp(3\mu + 2\sigma^2)
\end{bmatrix}
\]  \hspace{1cm} (23)

Assume the aim is to generate data with
\[
\mathbf{m}_Y = \begin{bmatrix}
\mu_Y & E(Y - \mu_Y)^2 & E(Y - \mu_Y)^3
\end{bmatrix}^T
\]  \hspace{1cm} (24)
and the choice is between using a linear combination of a \(\chi^2\) or a lognormal distributed variable. Minimizing \(f\), the derived linear combinations are
\[
Y_1 = -2.00 + 0.25X_1
\]  \hspace{1cm} (25)
\[
Y_2 = -3.10 + 0.21X_2
\]  \hspace{1cm} (26)

where \(X_1\) is \(\chi^2\) with 8 degrees of freedoms and \(X_2\) is lognormal with mean and variance parameters 2.65 and 0.099 respectively. The covariance matrices for
the moments are

\[
V(m_{Y_1}) = \begin{bmatrix}
1.00 & 1.00 & 1.50 \\
1.00 & 3.50 & 9.00 \\
1.50 & 9.00 & 36.00
\end{bmatrix}
\] (27)

\[
V(m_{Y_2}) = \begin{bmatrix}
1.00 & 1.00 & 1.83 \\
1.00 & 3.83 & 11.03 \\
1.83 & 11.03 & 50.26
\end{bmatrix}
\] (28)

with determinants \( |V(m_{Y_1})| = 28.13 \) and \( |V(m_{Y_2})| = 48.10 \). This implies that the sampling variability is considerably less when using the linear combination of the \( \chi^2 \) variable than using the lognormal, although these two linear combinations give the same first three moments. The method of Fleishman is a linear combination up to the third power of a standard normal variable and, assuming \( c \) not equal to zero, the third moment is a function of expectations of the ninth power of \( X \) and the variance of the skewness of the eighteenth! Considering kurtosis the case becomes even more extreme with expectations of powers up to the twenty-fourth.

5 Generating data for the Bootstrap

In many applications the skewness and kurtosis in the estimated residual signal deviation from the normal distribution. This can be dealt with using the standard bootstrap procedure but not always successfully. E.g. fit statistics in structural equation models are very sensitive to skewness and kurtosis as shown by Powell and Schafer (2001). Enders (2002) get better results when applying the Bollen-Stine bootstrap but for some non-normal situations the empirical size
is close to zero or even zero. Adjusting the bootstrap for skewness and kurtosis would most likely help.

The basic idea with the bootstrap is to mimic the small sample properties by a Monte Carlo based method which in small samples is better than asymptotic methods. It can be shown that under rather weak conditions the bootstrap has a higher convergence rate than standard asymptotic methods. There are parametric and non-parametric bootstraps. With the parametric a distribution is used to generate data while for the non-parametric a re-sample scheme is used. We will not deal with the parametric in this paper. Consider the case where you want to evaluate the sampling variability of the OLS estimator when the residuals are non-normally distributed, but do not want to rely on the asymptotic normality of the estimator. Start with estimating the model on data and derive the residuals. From the residuals, draw with replacements $T$ (where $T$ is the sample size) residuals. This is your first set of bootstrap residuals. Generate the new dependent variable with the bootstrap residuals. Given the bootstrap dependent variable and the fixed explanatory variables re-estimate the model and save the parameter estimate. Repeat this $B$ times and then you can use the $B$ parameter estimates to characterize the sampling variability. Of course, a higher $B$ implies that you would have better knowledge of the sampling variability. Unfortunately, there is only a total of $N = \binom{2T-1}{T}$ distinct bootstrap samples, hence, a $B$ larger than that would not help.

In a system of $m$ equations it is possible to bootstrap "pairs" to save the dependence between the $m$ residuals. Also in this case there would still be $N$ distinct bootstrap samples. To increase $N$ there are two possible solutions.
The first is that for the bootstrap residual for the $i$th equation draw with replacement from the all residuals of the $i$th equation. The second is to draw with replacement from all residuals. The number of distinct bootstrap samples are $N^m$ and \( \binom{2mT - 1}{mT} \) respectively. In both cases the bootstrap destroys the dependence between the $m$ bootstrapped residuals. In the first case the marginal distribution for the $i$th bootstrapped residual is valid but not in the second.

If the multivariate normality assumption is made the solution is simple and well known. Let $\mathbf{\Omega}$ be the covariance matrix of order $m \times m$ and use the decomposition $\mathbf{\Omega} = \mathbf{P}\mathbf{P}^T$ on the $T \times m$ data matrix $\mathbf{X}$ then each element of $\mathbf{Y} = \mathbf{X}\mathbf{P}^{-T}$ is iid. The bootstrap samples can now be made as $\mathbf{X}^* = \mathbf{Y}^*\mathbf{P}^T$ where every element of $\mathbf{Y}^*$ is drawn from $\mathbf{Y}$ with replacement. This gives bootstrap data which exactly mimics the first two moments even if the data are not multivariate normally distributed. The method outlined in the previous sections may be used if higher moments than two are of interest.

In the case when the marginal distributions are valid the weights have to be chosen such that the univariate moments are not changed but the sample multivariate moments are the sample ones. To be more precise, the bootstrap $\mathbf{Y}^*$ can be drawn with replacement from the corresponding column of the data matrix $\mathbf{X}$ Then weights, $\mathbf{P}$, have to be chosen such that each row of $\mathbf{X}^* = \mathbf{Y}^*\mathbf{P}^T$ have the expected univariate and multivariate moments you want them to have, i.e. the one estimated from the original sample $\mathbf{X}$ For the other case we need to transform $\mathbf{X}$ into $\mathbf{Y}$ such that each row has moments which obeys

$$
E \left( Y_i - \mu_{Y_i} \right)^{k_i} \left( Y_j - \mu_{Y_j} \right)^{k_j} = E \left( Y_i - \mu_{Y_i} \right)^{k_i} E \left( Y_j - \mu_{Y_j} \right)^{k_j} = \ldots
$$
\[ E \left( Y_j - \mu_{Y_j} \right)^{k_j} \text{ and } EY_i = EY_j \text{ up to a certain order of interest } (k_i + k_j \leq k). \]

This can be done by not deciding the elements of \( m_Y \) a priori but minimizing \( f \) with the restriction that the elements obey the above mentioned moment restrictions. When \( Y \) is formed one can continue as in the first case.

It is also possible to derive all weights in the same minimization by minimizing

\[
f = (m_Y - Am_X)^T (m_Y - Am_X) + (m_X - Bm_Y)^T (m_X - Bm_Y) \tag{29}
\]

At minimum \( m_Y = Am_X \) hence substitution gives

\[
f = (m_Y - Am_X)^T (m_Y - Am_X) + (m_X - BAm_X)^T (m_X - BAm_X) \tag{30}
\]

where the second part is zero iff \( BA = I \). If \( f = 0 \) then it is possible to bootstrap data with exactly the moments requested.

Consider a bivariate example with moment matrices

\[
m_Y = \begin{bmatrix} 1 & \mu_{Y_1} & V(Y_1) & E \left( Y_1 - \mu_{Y_1} \right)^3 & \mu_{Y_2} & V(Y_2) \\ E \left( Y_2 - \mu_{Y_2} \right)^3 & \rho_{Y_1,Y_2} & \mu_{1Y_1,2Y_2} & \mu_{2Y_1,1Y_2} \end{bmatrix}^T \tag{31}
\]

and

\[
m_X = \begin{bmatrix} 1 & \mu_{X_1} & V(X_1) & E \left( X_1 - \mu_{X_1} \right)^3 & \mu_{X_2} & V(X_2) \\ E \left( X_2 - \mu_{X_2} \right)^3 & \rho_{X_1,X_2} & \mu_{1X_1,2X_2} & \mu_{2X_1,1X_2} \end{bmatrix}^T \tag{32}
\]

where \( \mu_{1Y_1,2Y_2} \equiv EY_1 \left( Y_2 - \mu_{Y_2} \right)^2 \) and \( \mu_{2Y_1,1Y_2} \equiv E \left( Y_1 - \mu_{Y_1} \right)^2 Y_2 \). The linear combinations of interest are

\[
Y_1 = a_{10} + a_{11} X_1 + a_{12} X_2 \tag{33}
\]

\[
Y_2 = a_{20} + a_{21} X_1 + a_{22} X_2 \tag{34}
\]
In the bootstrap exercise, the elements in $m_x$ must be estimated from data. To get $Y_1$ and $Y_2$ which are identically distributed up to order three and obey the moments restrictions outlined above the following restrictions are set

$$m_Y = \begin{bmatrix} 1 & \mu_Y & E(Y - \mu_Y)^2 & E(Y - \mu_Y)^3 \\ 0 & \mu_Y E(Y - \mu_Y)^2 & \mu_Y E(Y - \mu_Y)^3 \end{bmatrix}^T$$

(35)

while the estimated moments are used in $m_x$. The arguments in the minimization are $a_{10}, a_{11}, a_{12}, a_{20}, a_{21}, a_{22}, \mu_Y, E(Y - \mu_Y)^2$, and $E(Y - \mu_Y)^3$, i.e. a total of nine which is the same as the number of elements in $m_x$, hence, it might be possible to get an exact solution. In the appendix the relationship between the moments which are to be used to form $A$ are shown.

5.1 A small Monte Carlo simulation

To exemplify the usefulness of the ability to mimic higher order moments, a small Monte Carlo simulation is carried out. For each observation $N = 2, 5, 10$ bivariate normally variates are generated, squared and summed, i.e.

$$X_{it} = \sum_{j=1}^{N} Z_{ijt}^2, \quad t = 1, \ldots, T, \quad i = 1, 2$$

(36)

where $[Z_{1jt}, Z_{2jt}]^T$ is the standard bivariate normal with correlation $\rho = 0, 0.5, 0.7$. A total of $T = 10, 20$ observations are generated. As the variables are squared generating variables with negative correlations is the same thing as generating variables with a positive correlation, hence, only the positive ones are used in the study. This distribution is discussed in Krishnamoorty and Parthasarathy (1951) and is a multivariate generalization of the $\chi^2$ distribution. It can be seen as a special case of the Wishart distribution, see Wishart (1928). First we
bootstrap data according to the method where we use the Cholesky decom-
position of the covariance matrix to derive data. This implies that the first two
moments are correct. Then \( T \) observations are drawn with replacement from
the \( 2T \) "bootstrap residuals" and the observations are then transformed back.
This is repeated 1001 times and the 2.5th and the 97.5th percentiles are cal-
culated from the 1001 bootstrap samples. Each of these replicates is repeated
2000 times. The empirical percentiles are calculated as the percentages below
the mean of the 1000 estimated percentiles, i.e. we answer the question: You
want the 2.5/97.5th percentile but what do you get in practice? This method,
denoted 2nd moments, is compared to the one proposed in this paper where
we also added \( E(X_1 - \mu_{X_1})^3 \) and \( E(X_2 - \mu_{X_2})^3 \) to the set of first and second
moments matched in the 2nd moments procedure, this is denoted 3rd moments.
The estimators used for the various moments are the ordinary unbiased ones.

The results, which are shown in Table (1), show that almost all the time
taking the third moments into account improves the percentiles. The general
impression of the impact of changing \( \rho \) is that moving away from zero makes the
distribution more skewed and, hence, taking care of the third moments improves
the percentiles but it still has a negative impact on the estimates. Increasing
\( N \) makes the distribution more symmetric when \( \rho = 0 \), and for other values of
\( \rho \) the skewness converges to a negative number with \( N \) with the corresponding
improvement of using the third moments. Increasing the sample size from \( T =
10 \) to 20 does not make the percentiles better when only taking care of the 2nd
moments, contrary, mostly it makes the results worse. The opposite is true when
also correcting for the third moments. The overall result shows that taking care
of higher moments may be of great importance and the result is linked to the existence of skewness different from zero.

6 Conclusion

This paper presents a general method for the generation of multivariate data with arbitrary moments. The method is shown to have the method presented by Fleishman and in multivariate form by Vale and Maurelli as a special case. It is also explained why that method has a tendency to generate data with higher moments with large variance and an example is shown how one can reduce the uncertainty of the higher moments. In the bootstrap there is a general need for the generation of data which mimics the higher moments well and how this is done is also shown. A Monte Carlo simulation shows the advantages of using the proposed method where we choose to mimic up to the third moments compared to when only considering the second moments which is usually the case.

Acknowledgments

I wish to express my thanks to Jonas Andersson, Rolf Larsson and an anonymous referee for helpful comments and to The Bank of Sweden Tercentenary Foundation and The Jan Wallander and Tom Hedelius Foundation for financial support.
References


of Educational and Behavioral Statistics, 26, 105-132.


Appendix: Relationship between moments

Notation: \( EX' = \mu'_X \), \( E(\mathbf{X} - \mu'_1) = \mu_r \), \( EX'_1 X'_2 = \mu_{r'2} \), \( E(\mathbf{X}_1 - \mu'_{10})^T (\mathbf{X}_2 - \mu'_{01}) = \mu_{r'1} \) and when the moments is for \( Y \) we use \( \lambda \) instead of \( \mu \).

We have two variables which are linear combinations of two variables

\[
Y_1 = a_{10} + a_{11}X_1 + a_{12}X_2 \\
Y_2 = a_{20} + a_{21}X_1 + a_{22}X_2
\]

The moments we want are

\[
m_Y = \begin{bmatrix}
1 & \lambda_{10} & \lambda_{11} & \lambda_{12} \\
\lambda_{03} & \lambda_{20} & \lambda_{21} & \lambda_{22}
\end{bmatrix}^T
\]

and the moments \( X_1 \) and \( X_2 \) are

\[
m_X = \begin{bmatrix}
1 & \mu'_{10} & \mu_{20} & \lambda_{01} & \mu_{02} \\
\mu_{03} & \mu_{11} & \mu_{21} & \lambda_{12}
\end{bmatrix}^T
\]

Below we show the relationships between the moments which are used to form \( m \).

1.

\[
1 = 1
\]

2.

\[
\lambda'_{10} = a_{10} + a_{11}\mu'_{10} + a_{12}\mu_{01}
\]
\( \lambda_{20} = E(Y_1 - X_{10}^2) \)
\[
= E\left( a_{10} + a_{11}X_1 + a_{12}X_2 - a_{10} - a_{11}\mu_{10} - a_{12}\mu_{01} \right)^2 \\
= E\left( a_{11} (X_1 - \mu_{10}) + a_{12} (X_2 - \mu_{01}) \right)^2 \\
= E\left( a_{11}^2 (X_1 - \mu_{10})^2 + 2a_{11}a_{12} (X_1 - \mu_{10}) (X_2 - \mu_{01}) + a_{12}^2 (X_2 - \mu_{01})^2 \right) \\
= a_{11}^2\mu_{20} + 2a_{11}a_{12}\mu_{11} + a_{12}^2\mu_{02} 
\]

\( \lambda_{30} = E(Y_1 - X_{10}^3) \)
\[
= E\left( a_{11} (X_1 - \mu_{10}) + a_{12} (X_2 - \mu_{01}) \right)^3 \\
= E\left[ a_{11}^3 (X_1 - \mu_{10})^3 + a_{11}^2a_{12} (X_1 - \mu_{10})^2 (X_2 - \mu_{01}) + a_{11}a_{12}^2 (X_1 - \mu_{10}) (X_2 - \mu_{01})^2 + a_{12}^3 (X_2 - \mu_{01})^3 \right] \\
= a_{11}^3\mu_{30} + 3a_{11}^2a_{12}\mu_{21} + 3a_{11}a_{12}^2\mu_{12} + a_{12}^3\mu_{03} 
\]

\( \lambda_{11} = E(Y_1 - X_{10}) (Y_2 - \mu_{01}) \)
\[
= E\left[ (a_{11} (X_1 - \mu_{10}) + a_{12} (X_2 - \mu_{01})) (a_{21} (X_1 - \mu_{10}) + a_{22} (X_2 - \mu_{01})) \right] \\
= E\left[ a_{11}a_{21} (X_1 - \mu_{10}) (X_1 - \mu_{10}) + a_{11}a_{22} (X_1 - \mu_{10}) (X_2 - \mu_{01}) + a_{12}a_{21} (X_2 - \mu_{01}) (X_1 - \mu_{10}) + a_{12}a_{22} (X_2 - \mu_{01}) (X_2 - \mu_{01}) \right] \\
= a_{11}a_{21}\mu_{20} + a_{11}a_{22}\mu_{11} + a_{12}a_{21}\mu_{11} + a_{12}a_{22}\mu_{02} \\
= a_{11}a_{21}\mu_{20} + (a_{11}a_{22} + a_{12}a_{21}) \mu_{11} + a_{12}a_{22}\mu_{02} 
\]
6.

\[
\lambda_{21'} = E \left( Y_1 - \lambda_{10}' \right)^2 \lambda_{22}
\]

\[
= E \left( Y_1^2 - 2\lambda_{10}' Y_1 + \lambda_{10}'^2 \right)
\]

\[
= \lambda_{22}' - 2\lambda_{10}' \lambda_{11} + \lambda_{10}'^2 \lambda_{01}'
\]

As \( m_X \) contains various central moments and means \( \lambda_{21'} \) has to be rewritten as a function of those quantities. From Cramer (1946), p. 263, we have the result:

\[
\lambda_{11} = \lambda_{11}' - \lambda_{10}' \lambda_{01}'
\]

and in Cook (1951), where \( k \) is the cumulant,

\[
\lambda_{21}' = k_{12} + k_{20} \lambda_{01} + 2k_{11} \lambda_{10} + k_{10}^2 \lambda_{01}
\]

\[
= \lambda_{21} + \lambda_{20} \lambda_{01} + 2\lambda_{11} \lambda_{10} + \lambda_{10}^2 \lambda_{01}
\]

Substituting these two results we have

\[
\lambda_{21'} = \lambda_{21} + \lambda_{20} \lambda_{01}' + 2\lambda_{11} \lambda_{10}' + \lambda_{10}^2 \lambda_{01}' - 2\lambda_{10}' \left( \lambda_{11} + \lambda_{10}' \lambda_{01}' \right) + \lambda_{10}^2 \lambda_{01}'
\]

\[
= \lambda_{21} + \lambda_{20} \lambda_{01}'
\]

From the previous results we know the expressions for \( \lambda_{20} \) and \( \lambda_{01}' \), and we only need to evaluate \( \lambda_{21} \). This one is slightly more complicated but we use the result (which is a generalization of Cramer, 1946, p 298) \( \alpha_{123} = \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} a_{1i} a_{2j} a_{3k} \beta_{ijk}, \beta_{ijk} = EZ_i Z_j Z_k, \ Z_1 = (X_1 - \mu_{10}') \) and

\[
Z_2 = (X_2 - \mu_{01}') , \ i.e. \ we \ have \ three \ variables, \ denoted \ by \ 1 \ to \ 3, \ that \ are
\]
linear combinations of \( n \) variables.

\[
\lambda_{21} = \sum_{i=1}^{2} \sum_{j=1}^{2} \sum_{k=1}^{2} \alpha_{11}a_{11}a_{2k}\beta_{ijk} = a_{11}\alpha_{11}a_{21}\beta_{111} + a_{11}\alpha_{11}a_{22}\beta_{112} + a_{11}\alpha_{12}a_{21}\beta_{211} + a_{11}\alpha_{12}a_{22}\beta_{212} + a_{12}\alpha_{11}a_{21}\beta_{121} + a_{12}\alpha_{11}a_{22}\beta_{122} + a_{12}\alpha_{12}a_{21}\beta_{221} + a_{12}\alpha_{12}a_{22}\beta_{222}
\]

\[
= (a_{11}\alpha_{11}a_{21} + a_{12}\alpha_{11}a_{21}) \beta_{111} + (a_{11}\alpha_{11}a_{22} + a_{11}\alpha_{12}a_{21} + a_{12}\alpha_{11}a_{22} + a_{12}\alpha_{12}a_{21}) \beta_{112} + (a_{11}\alpha_{12}a_{22} + a_{12}\alpha_{12}a_{22}) \beta_{122} + (a_{11}\alpha_{11}a_{21} + a_{12}\alpha_{11}a_{21}) \mu_{30} + (a_{11}\alpha_{11}a_{22} + a_{11}\alpha_{12}a_{21} + a_{12}\alpha_{11}a_{22} + a_{12}\alpha_{12}a_{21}) \mu_{21} + (a_{11}\alpha_{12}a_{22} + a_{12}\alpha_{12}a_{22}) \mu_{12}
\]

Hence,
\[
\lambda_{21} = \left( a_{11}a_{11}a_{21} + a_{12}a_{11}a_{21} \right) \mu_{30} \\
+ \left( a_{11}a_{11}a_{22} + a_{11}a_{12}a_{21} + a_{12}a_{11}a_{22} + a_{12}a_{12}a_{21} \right) \mu_{21} \\
+ \left( a_{11}a_{12}a_{22} + a_{12}a_{12}a_{22} \right) \mu_{12} \\
+ \left( a_{12}^2 \mu_{20} + 2a_{11}a_{12} \mu_{11} + a_{22}^2 \mu_{02} \right) \left( a_{20} + a_{21} \mu'_{10} + a_{22} \mu'_{01} \right) \\
= \left( a_{11}a_{11}a_{21} + a_{12}a_{11}a_{21} \right) \mu_{30} \\
+ \left( a_{11}a_{11}a_{22} + a_{11}a_{12}a_{21} + a_{12}a_{11}a_{22} + a_{12}a_{12}a_{21} \right) \mu_{21} \\
+ \left( a_{11}a_{12}a_{22} + a_{12}a_{12}a_{22} \right) \mu_{12} + a_{11}^2 \mu_{20} a_{20} + a_{11}^2 \mu_{20} a_{21} \mu'_{10} \\
+ a_{12}^2 \mu_{20} a_{22} \mu'_{01} + 2a_{11}a_{12} \mu_{11} a_{20} \\
+ 2a_{11}a_{12} \mu_{11} a_{21} \mu'_{10} + 2a_{11}a_{12} \mu_{11} a_{22} \mu'_{01} \\
+ a_{12}^2 \mu_{02} a_{20} + a_{12}^2 \mu_{02} a_{21} \mu'_{10} + a_{12}^2 \mu_{02} a_{22} \mu'_{01}
\]
<table>
<thead>
<tr>
<th>N</th>
<th>T</th>
<th>ρ</th>
<th>μ</th>
<th>ρ³</th>
<th>2nd moments</th>
<th>3rd moments</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>P0.025</td>
<td>P0.975</td>
</tr>
<tr>
<td>2</td>
<td>10</td>
<td>0</td>
<td>0.117</td>
<td>0.409</td>
<td>0.0250</td>
<td>0.9425</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.5</td>
<td>0.188</td>
<td>0.00486</td>
<td>0.0190</td>
<td>0.9165</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.7</td>
<td>0.306</td>
<td>-0.269</td>
<td>0.0090</td>
<td>0.8745</td>
</tr>
<tr>
<td>20</td>
<td>0</td>
<td>0.00511</td>
<td>0.329</td>
<td>0.0170</td>
<td>0.9655</td>
<td>0.0090</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.5</td>
<td>0.189</td>
<td>0.112</td>
<td>0.0170</td>
<td>0.9255</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.7</td>
<td>0.3106</td>
<td>-0.0891</td>
<td>0.0025</td>
<td>0.8375</td>
</tr>
<tr>
<td>5</td>
<td>10</td>
<td>0</td>
<td>0.00128</td>
<td>0.188</td>
<td>0.0320</td>
<td>0.9505</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.5</td>
<td>0.183</td>
<td>-0.179</td>
<td>0.0175</td>
<td>0.9360</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.7</td>
<td>0.302</td>
<td>-0.407</td>
<td>0.0090</td>
<td>0.9005</td>
</tr>
<tr>
<td>20</td>
<td>0</td>
<td>-0.00869</td>
<td>0.215</td>
<td>0.0215</td>
<td>0.9695</td>
<td>0.0150</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.5</td>
<td>0.183</td>
<td>-0.101</td>
<td>0.0145</td>
<td>0.9375</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.7</td>
<td>0.308</td>
<td>-0.314</td>
<td>0.0045</td>
<td>0.8445</td>
</tr>
<tr>
<td>10</td>
<td>10</td>
<td>0</td>
<td>0.001364</td>
<td>0.0523</td>
<td>0.0365</td>
<td>0.9590</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.5</td>
<td>0.1875</td>
<td>-0.271</td>
<td>0.0165</td>
<td>0.9470</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.7</td>
<td>0.309</td>
<td>-0.482</td>
<td>0.0060</td>
<td>0.9080</td>
</tr>
<tr>
<td>20</td>
<td>0</td>
<td>-0.00731</td>
<td>0.0503</td>
<td>0.0370</td>
<td>0.9665</td>
<td>0.0365</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.5</td>
<td>0.188</td>
<td>-0.174</td>
<td>0.0130</td>
<td>0.9370</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.7</td>
<td>0.315</td>
<td>-0.323</td>
<td>0.0035</td>
<td>0.8405</td>
</tr>
</tbody>
</table>

Table 1: Empirical 2.5 and 97.5 percentiles of the correlation of the sum of the squared Gaussian variables when bootstrapping second and third moments respectively. ρ is the correlation before squaring and the sum is made, μ is the estimated mean of correlation of the sum of the squared variables while ρ³ is the skewness of the same.
Response to comments by referee:

April 28, 2008

First of all I would like to thank the referee for the comments.

1. Here, I choose to relate the question of choosing distribution and \( p \) to the same kind of solution as in Section 4, i.e. choose the distribution and \( p \) which gives smallest variances of the moments of interest. Text is inserted at appropriate position, i.e. p. 5 line 49, to suggest this approach.

2. The reason to choose \( p = 3 \) is that then it is easily seen that this approach yields the same weights as the previous method of Vale and Maurelli, hence their method is a special case. This has been clarified in the text. The coefficients \( a - d \) have been defined.

3. At p. 10, line 46 the text is rewritten with an example and references to a situation where a bootstrap procedure incorporating skewness and kurtosis would most likely help.

4. Matrix \( A \) inserted.

5. Transpose inserted.

6. \( m \) changed to \( A \).

7. \( \mu_{1X,2X} \) and \( \mu_{2X,1X} \) defined with etc to indicate that the same applies to \( X \).