# A method to generate multivariate data with moments arbitrary close to the desired moments

Johan Lyhagen*
Department of Economic Statistics, Stockholm School of Economics
*Working Paper Series in Economics and Finance, No. 481*

December 19, 2001

**Abstract**

We show how it is possible to generate multivariate data which have 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 examplified with an Monte Carlo simulation where the importance of the ability of generating data with control of higher moments is shown.

**Keywords:** Monte Carlo, skewness.
**JEL:** C15, C63.

## 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 depends on the generation of data. For a general treatment of generating data see Devroye (1986) and 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 characteristic. It is very convenient to assume (multivariate) normality but as noted by many, e.g. Fleishman (1978), the distribution of real world data typically are characterized with skewness and kurtosis which deviates from the normal distribution. Fleishman (1978) proposed to generate data according to

$$Y = a + bX + cX^2 + dX^3 \tag{1}$$

where $X$ is standard normal distributed. The distribution of $Y$ is generally unknown but it is some times possible to choose $a, b, c$ and $d$ such as $Y$ have the

*E-mail: johan.lyhagen@hhs.se

desired first four moments. Tadikamalla (1980) criticize this method as there are some combinations of skewness and kurtosis which is impossible to reach and review 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 lies within the feasible area and one of the methods reviewed by Tadikamalla (1980) to be used otherwise. One advantage of the Fleishman (1978) method is that it is easily generalized to the multivariate setting as done by Vale and Maurelli (1983). Define $\mathbf{x}$ as

$$\mathbf{x} = \left[1, X, X^2, X^3\right]^T \tag{2}$$

and the vector of weights

$$\mathbf{w} = [a, b, c, d]^T \tag{3}$$

then

$$Y = \mathbf{w}^T \mathbf{x} \tag{4}$$

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

$$\begin{aligned} \rho_{Y_1,Y_2} &= E\left(Y_1 Y_2\right) \\ &= E\left(\mathbf{w}_1^T \mathbf{x}_1 \mathbf{x}_2^T \mathbf{w}_2\right) \\ &= \mathbf{w}_1^T \mathbf{R} \mathbf{w}_2 \end{aligned} \tag{5}$$

where $\mathbf{R} = E\left(\mathbf{x}_1 \mathbf{x}_2^T\right)$ . The relationship between $\rho_{Y_1,Y_2}$ and the covariance of $X_1$ and $X_2$, expressed with the weights is

$$\begin{aligned} \rho_{Y_1,Y_2} &= \rho_{X_1 X_2}\left(b_1 b_2 + 3 b_1 d_2 + 3 d_1 b_2 + 9 d_1 d_2\right) \\ &\quad + \rho_{X_1 X_2}^2 \left(2 c_1 c_2\right) + \rho_{X_1 X_2}^3 \left(6 d_1 d_2\right) \end{aligned} \tag{6}$$

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 is 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. $EY_1^2 Y_2$ and $EY_1^2 Y_2^2$ 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 (1983) method.

The paper is as follows: Section 2 presents the proposed method and Section 3 generalize it to deal with the generation of multivariate data with arbitrary moments. How to choose between different competing distributions whereof 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 sufficient many moments existing, and $a_i$ the associated weight then

$$Y = a_0 + \sum_{i=1}^{p} a_i X_i \tag{7}$$

The expectation is

$$\begin{aligned} EY &= a_0 + \sum_{i=1}^{p} a_i E X_i \\ &= a_0 + \sum_{i=1}^{p} a_i \mu_{X_i} \end{aligned} \tag{8}$$

Similarly, the variance is

$$\begin{aligned} V(Y) &= E\left(a_0 + \sum_{i=1}^{p} a_i X_i - a_0 - \sum_{i=1}^{p} a_i \mu_{X_i}\right)^2 \\ &= \sum_{i=1}^{p}\sum_{j=1}^{p} a_i a_j Cov\left(X_i, X_j\right) \end{aligned} \tag{9}$$

In general, the $k$th central moment for $Y$ is

$$E\left(Y - \mu_Y\right)^k = \sum_{i_1=1}^{p}\sum_{i_2=1}^{p}\cdots\sum_{i_k=1}^{p} a_{i_1} a_{i_2}\cdots a_{i_k}\mu_{i_1 i_2\cdots i_k} \tag{10}$$

where $\mu_{i_1 i_2 \cdots i_k} = E\prod_{i\in i_1 i_2\cdots i_k}\left(X_i - \mu_{X_i}\right)$.

To generate $Y$ values of $a_i$ must be decided upon such as the desired moments are gained. 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 get the wanted mean of $Y$. The relationship between the two vectors is

$$\mathbf{m}_Y = \mathbf{A}\mathbf{m}_X \tag{11}$$

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

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

$$\mathbf{A} = \left[ \begin{array}{ccc} a_0 & 0 & 0 \\ a_1 & 2a_1^2 & 8a_1^3 \end{array} \right]^T \tag{13}$$

$$\mathbf{m}_X = \left[ \begin{array}{cc} 1 & \lambda \end{array} \right]^T \tag{14}$$

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 unknown 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

$$f = \left(\mathbf{m}_Y - \mathbf{A}\mathbf{m}_X\right)^T \left(\mathbf{m}_Y - \mathbf{A}\mathbf{m}_X\right) \tag{15}$$

and then minimize $f$ with respect to the unknown parameters. This is simple to do with some numerical optimization technique commonly implemented in programing languages such as Gauss. 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 need not be unique. Depending on the situation increasing $p$ or changing the used distribution of the $X$ is to be preferred. E.g. if a non-zero third central moment of $Y$ is wanted and $X$ is normal 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, for some reason, one want to stick to a particular $p$ and distribution of $X$ then a better choice of function to minimize is

$$f = \left(\mathbf{m}_Y - \mathbf{A}\mathbf{m}_X\right)^T \mathbf{\Sigma}^{-1} \left(\mathbf{m}_Y - \mathbf{A}\mathbf{m}_X\right) \tag{16}$$

where a typical element of $\mathbf{\Sigma}$ 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 are 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 such as third or fourth cross moments are possible to. The above is easily generalized to generate multivariate data. Now, for $j = 1, \ldots, m$, we have

$$Y_j = a_{0j} + \sum_{i=1}^{p} a_{ij} X_i \tag{17}$$

and let $\mathbf{m}_Y$ be a vector of all moments of interest for the multivariate distribution of $Y$, i.e. also the cross-moments, and let $\mathbf{m}_X$ 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 a linear combination of the same components of $X$. This produces a very general structure for the moments.

The method of Fleishman (1978), and for that case also the multivariate version of Vale and Maurelli (1983), is a special case of the above proposed method for generating multivariate data. Consider the bivariate case and remember that $X$ is multivariate normal. For ease of notation, for $i = 1, 2$, let $\mathbf{m}_{Y_i} = E\left[Y_i, Y_i^2, Y_i^3, Y_i^4\right]^T$, $\mathbf{m}_{X_i} = E\left[1, X_i, X_i^2, \ldots, X_i^{12}\right]^T$ and

$$a_{y_i} = \begin{bmatrix} a & a^2 & a^3 & a^4 \\ b & 2a_i b_i & 3a^2 b & 4a^3 b \\ c & 2ac + b^2 & 3b^2 a + 3ca^2 & 6a^2 b^2 + 4a^3 c \\ d & 2ad + 2bc & b^3 + 6cab + 3da^2 & 4ab^3 + 12a^2 bc + 4a^3 d \\ 0 & 2bd + c^2 & 6dab + 3ac^2 + 3b^2 c & 12a^2 bd + 6a^2 c^2 + 12ab^2 c + b^4 \\ 0 & 2cd & 6acd + 3b^2 d + 3bc^2 & 12a^2 cd + 12ab^2 d + 12abc^2 + 4b^3 c \\ 0 & d^2 & 6bcd + 3ad^2 + c^3 & 6a^2 d^2 + 24bcda + 4b^3 d + 6b^2 c^2 + 4ac^3 \\ 0 & 0 & 3bd^2 + 3c^2 d & 12abd^2 + 12cdb^2 + 12c^2 da + 4bc^3 \\ 0 & 0 & 3cd^2 & 6d^2 b^2 + 12d^2 ac + 12c^2 db + c^4 \\ 0 & 0 & d^3 & 12cd^2 b + 4c^3 d + 4d^3 a \\ 0 & 0 & 0 & 4d^3 b + 6c^2 d^2 \\ 0 & 0 & 0 & 4cd^3 \\ 0 & 0 & 0 & d^4 \end{bmatrix}^T$$

where the index $i$ for the weights is suppressed to save space. Then $\mathbf{m}_Y, \mathbf{A}$ and $\mathbf{m}_X$ are as follows:

$$\mathbf{m}_Y = \begin{bmatrix} \mathbf{m}_{Y_1}^T & \mathbf{m}_{Y_2}^T & \rho_{Y_1,Y_2} \end{bmatrix}^T$$

$$\mathbf{A} = \begin{bmatrix} a_{y_1} & 0 & 0 & 0 & 0 \\ 0 & a_{y_2} & 0 & 0 & 0 \\ 0 & 0 & b_1 b_2 + 3b_1 d_2 + d_1 b_2 + 9d_1 d_2 & 2c_1 c_2 & 6d_1 d_2 \end{bmatrix}$$

$$\mathbf{m}_X = \begin{bmatrix} \mathbf{m}_{X_1} & \mathbf{m}_{X_2} & \rho_{X_1,X_2} & \rho_{X_1,X_2}^2 & \rho_{X_1,X_2}^3 \end{bmatrix}^T$$

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 (1983) but in one step and are in that way simpler and more intuitive.

# 4 Choosing amongst competing distributions

It is well known that there is a 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 depends on the choice. Assume there is for convenience $L$ competing sets of $X$ vectors which gives $Y$ 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 $\mathbf{\Sigma}$ be the covariance matrix of $\mathbf{m}_Y$ then the best $X$ would be the one which has the smallest $\mathbf{\Sigma}$ using some measure. One natural choice of would be the determinant, $|\mathbf{\Sigma}|$ , referred to as the generalized variance by Mardia et al (1979). It is also possible to use the trace of $\mathbf{\Sigma}$, i.e. the total variation.

For example, let $X$ be lognormal distributed and as in the example above the three first moments are of interest then the three matrices are:

$$\mathbf{m}_Y = \left[ \begin{array}{ccc} \mu_Y & E\left(Y-\mu_Y\right)^2 & E\left(Y-\mu_Y\right)^3 \end{array} \right]^T \tag{18}$$

$$\mathbf{A} = \left[ \begin{array}{cccc} a_0 & 0 & 0 & 0 \\ 0 & a_1 & 0 & 0 \\ 0 & 0 & a_1^2 & 0 \\ 0 & 0 & 0 & a_1^3 \end{array} \right] \tag{19}$$

$$\mathbf{m}_X = \left[ \begin{array}{cccc} 1 & e^{\mu+\sigma^2/2} & e^{2\mu+2\sigma^2} - e^{2\mu+\sigma^2} & e^{3\mu+\frac{9}{2}\sigma^2} - 3e^{3\mu+\frac{5}{2}\sigma^2} + 2e^{3\mu+\frac{3}{2}\sigma^2} \end{array} \right]^T \tag{20}$$

Assume the aim is to generate data with

$$\begin{aligned} \mathbf{m}_Y &= \left[ \begin{array}{ccc} \mu_Y & E\left(Y-\mu_Y\right)^2 & E\left(Y-\mu_Y\right)^3 \end{array} \right]^T \\ &= \left[ \begin{array}{ccc} 0 & 1 & 1 \end{array} \right]^T \end{aligned} \tag{21}$$

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

$$Y_1 = -2.00 + 0.25X_1 \tag{22}$$

$$Y_2 = -3.10 + 0.21X_2 \tag{23}$$

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

$$V\left(\mathbf{m}_{Y_1}\right) = \left[ \begin{array}{ccc} 1.00 & 1.00 & 1.50 \\ 1.00 & 3.50 & 9.00 \\ 1.50 & 9.00 & 36.00 \end{array} \right] \tag{24}$$

$$V\left(\mathbf{m}_{Y_2}\right) = \left[ \begin{array}{ccc} 1.00 & 1.00 & 1.83 \\ 1.00 & 3.83 & 11.03 \\ 1.83 & 11.03 & 50.26 \end{array} \right] \tag{25}$$

with determinants $|V(\mathbf{m}_{Y_1})| = 28.13$ and $|V(\mathbf{m}_{Y_2})| = 48.10$. 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 (1978) 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 the eighteenth! Considering kurtosis the case become even more extreme with expectations of powers up to the twenty-fourth.

# 5 Generating data for the Bootstrap

In many applications, such as empirical finance, there are skewness and kurtosis in the estimated residual which signals deviation from the normal distribution. This can be dealt with using the bootstrap. The basic idea with the bootstrap is to mimic the small sample properties by a Monte Carlo based method which is better in small samples than asymptotic methods. It can be shown that under rather weak conditions the bootstrap converges quicker to the empirical distribution than standard asymptotic methods. There are parametric and non-parametric bootstrap. 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-normal 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 $N = \begin{pmatrix} 2T-1 \\ T \end{pmatrix}$ distinct bootstrap samples, hence, a $B$ larger than that would not help.

In a system of $m$ equations it is possible to bootstrap rows to save the dependence between the $m$ residuals. Also in this case there would still be $N$ number of 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 $\begin{pmatrix} 2mT-1 \\ mT \end{pmatrix}$ 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 $\Omega$ be the covariance matrix of order $m \times m$ and use the decomposition $\Omega = PP^T$ on the $T \times m$ data matrix $\mathbf{X}$ then each element of $\mathbf{Y} = \mathbf{XP}^{-T}$ is iid. The bootstrap samples can now be made as $\mathbf{X}^* = \mathbf{Y}^* P^T$ 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 as 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 as each row of $\mathbf{X}^* = \mathbf{Y}^* P^T$ 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 have moments which obeys $E\left(Y_i - \mu_{Y_i}\right)^{k_i}\left(Y_j - \mu_{Y_j}\right)^{k_j} = E\left(Y_i - \mu_{Y_i}\right)^{k_i} E\left(Y_j - \mu_{Y_j}\right)^{k_j}$, $E\left(Y_i - \mu_{Y_i}\right)^{k_i} = E\left(Y_j - \mu_{Y_j}\right)^{k_j}$ and $EY_i = EY_j$ up to a certain order of interest $(k_i + k_j \leq) k$. This can be done by not deciding the elements of $\mathbf{m}_Y$ apriori but minimizing $f$ with the restriction that the elements obey the above mentioned moment restrictions. When $\mathbf{Y}$ is formed one can continue as in the first case.

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

$$f = \left(\mathbf{m}_Y - \mathbf{A}\mathbf{m}_X\right)^T \left(\mathbf{m}_Y - \mathbf{A}\mathbf{m}_X\right) + \left(\mathbf{m}_X - \mathbf{B}\mathbf{m}_{\mathbf{Y}}\right)^T \left(\mathbf{m}_X - \mathbf{B}\mathbf{m}_Y\right) \quad (26)$$

At minimum $\mathbf{m}_Y = \mathbf{A}\mathbf{m}_X$ hence substitution gives

$$f = \left(\mathbf{m}_Y - \mathbf{A}\mathbf{m}_X\right)^T \left(\mathbf{m}_Y - \mathbf{A}\mathbf{m}_X\right) + \left(\mathbf{m}_X - \mathbf{B}\mathbf{A}\mathbf{m}_{\mathbf{X}}\right)^T \left(\mathbf{m}_X - \mathbf{B}\mathbf{A}\mathbf{m}_{\mathbf{X}}\right) \quad (27)$$

where the second part is zero iff $\mathbf{BA} = \mathbf{I}$.

Consider a bivariate example with moment matrices

$$\begin{aligned}\mathbf{m}_Y = \big[\ 1 \quad \mu_{Y_1} \quad V\left(Y_1\right) \quad E\left(Y_1 - \mu_{Y_1}\right)^3 \quad \mu_{Y_2} \quad V\left(Y_2\right) \\ E\left(Y_2 - \mu_{Y_2}\right)^3 \quad \rho_{Y_1,Y_2} \quad \mu_{1_{Y_1}2_{Y_2}} \quad \mu_{2_{Y_1}1_{Y_2}}\ \big]^T\end{aligned} \quad (28)$$

and

$$\begin{aligned}\mathbf{m}_X = \big[\ 1 \quad \mu_{X_1} \quad V\left(X_1\right) \quad E\left(X_1 - \mu_{X_1}\right)^3 \quad \mu_{X_2} \quad V\left(X_2\right) \\ E\left(X_2 - \mu_{X_2}\right)^3 \quad \rho_{X_1,X_2} \quad \mu_{1_{X_1}2_{X_2}} \quad \mu_{2_{X_1}1_{X_2}}\ \big]^T\end{aligned} \quad (29)$$

The linear combinations of interest are

$$Y_1 = a_{10} + a_{11}X_1 + a_{12}X_2 \quad (30)$$

$$Y_2 = a_{20} + a_{21}X_1 + a_{22}X_2 \quad (31)$$

In the bootstrap exercise, the elements in $\mathbf{m}_X$ must be estimated from data. To get $Y_1$ and $Y_2$ which are identical distributed up to order three and obey the moments restrictions outlined above the following restrictions are set

$$\mathbf{m}_Y = \begin{bmatrix} 1 & \mu_Y & E(Y-\mu_Y)^2 & E(Y-\mu_Y)^3 & \mu_Y & E(Y-\mu_Y)^2 \\ E(Y-\mu_Y)^3 & 0 & \mu_Y E(Y-\mu_Y)^2 & \mu_Y E(Y-\mu_Y)^2 \end{bmatrix}^T \quad (32)$$

while the estimated moments are used in $\mathbf{m}_X$. 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 number as elements in $\mathbf{m}_X$, hence, it might be possible to get an exact solution. In the appendix the relationship between the moments are shown which are to be used to form $\mathbf{A}$.

## 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. A standard bivariate normal process with correlation $\rho = 0, 0.5, 0.7$ is considered. For each observation $N = 2, 5, 10$ of the variates are generated,squared and summed, and in a total of $T = 10, 20$ observations are generated. As the variables is 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 Krishnamoorthy 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 decomposition of the covariance matrix to derive data. This implies that the first two moments are mimiced correctly. 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 calculated 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 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), shows 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 improve the percentiles but it still have 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 $T =$

| N | T | $\rho$ | $\bar{\rho}$ | $\rho^3$ | 2nd moments | | 3rd moments | |
|---|---|---|---|---|---|---|---|---|
| | | | | | $p_{0.025}$ | $p_{0.975}$ | $p_{0.025}$ | $p_{0.975}$ |
| 2 | 10 | 0 | 0.0117 | 0.409 | 0.0250 | 0.9425 | 0.0155 | 0.9470 |
| | | 0.5 | 0.188 | 0.00486 | 0.0190 | 0.9165 | 0.0295 | 0.9190 |
| | | 0.7 | 0.306 | -0.269 | 0.0090 | 0.8745 | 0.0355 | 0.9175 |
| | 20 | 0 | 0.00511 | 0.329 | 0.0170 | 0.9655 | 0.0090 | 0.9695 |
| | | 0.5 | 0.189 | 0.112 | 0.0170 | 0.9255 | 0.0175 | 0.9480 |
| | | 0.7 | 0.3106 | -0.0891 | 0.0025 | 0.8375 | 0.0250 | 0.9470 |
| 5 | 10 | 0 | 0.00128 | 0.188 | 0.0320 | 0.9505 | 0.0285 | 0.9490 |
| | | 0.5 | 0.183 | -0.179 | 0.0175 | 0.9360 | 0.0360 | 0.9380 |
| | | 0.7 | 0.302 | -0.407 | 0.0090 | 0.9005 | 0.0390 | 0.9345 |
| | 20 | 0 | -0.00869 | 0.215 | 0.0215 | 0.9695 | 0.0150 | 0.9670 |
| | | 0.5 | 0.183 | -0.101 | 0.0145 | 0.9375 | 0.0295 | 0.9535 |
| | | 0.7 | 0.308 | -0.314 | 0.0045 | 0.8445 | 0.0340 | 0.9510 |
| 10 | 10 | 0 | 0.001364 | 0.0523 | 0.0365 | 0.9590 | 0.0355 | 0.9585 |
| | | 0.5 | 0.1875 | -0.271 | 0.0165 | 0.9470 | 0.0430 | 0.9495 |
| | | 0.7 | 0.309 | -0.482 | 0.0060 | 0.9080 | 0.0370 | 0.9475 |
| | 20 | 0 | -0.00731 | 0.0503 | 0.0370 | 0.9665 | 0.0365 | 0.9655 |
| | | 0.5 | 0.188 | -0.174 | 0.0130 | 0.9370 | 0.0290 | 0.9550 |
| | | 0.7 | 0.315 | -0.323 | 0.0035 | 0.8405 | 0.0275 | 0.9470 |

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. $\rho$ is the correlation before squaring and the sum is made, $\bar{\rho}$ is the estimated mean of correlation of the sum of the squared variables while $\rho^3$ is the skewness of the same.

10 to 20 do 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 show 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 (1978) and in multivariate form by Vale and Maurelli (1983) as a special case. It is also explained why that method have 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 show 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 Andesson for helpful comments and to The Bank of Sweden Tercentenary Foundation and The Jan Wallander and Tom Hedelius Foundation for financial support.

## References

Cook, M.B. (1951) Bi-variate k-statistics and cumulants and their joint sampling distribution, *Biometrika* 38, 179-195.

Cramér, H. (1946) *Mathematical Methods of Statistics*, Princeton: Princeton University Press.

Devroye, L. (1986) *Non-uniform random variate generation*, New York: Springer Verlag.

Efron, B. and Tibshirani, R.J. (1993) *An introduction to the bootstrap*, New York : Chapman & Hall.

Fleishman, A.I. (1978) A method for simulating non-normal distributions, *Psychometrika* 43, 51-532.

Kendall, M.C. and Stewart, A. (1969) *The advanced theory of statistics*, vol. 1, London: Charles Griffin & Co.

Krishnamoorthy, A.S. and Parthasarathy, M. (1951) A multivariate gamma-type distribution, *Annals of Mathematical Statistics* 22, 549-557.

Kullback, S. and Leibler, R. (1951) On information and sufficiency, *Annals of Mathematical Statistics* 22, 79-86.

Mardia, K.V., Kent, J.T. and Bibby, J.M. (1979) *Multivariate analysis*, London: Academic Press.

Tadikamalla, P.R. (1980) On simulating nonnormal distributions, *Psychometrika* 45, 273-279.

Vale, C.D. and Maurelli, V.A. (1983) Simulating multivariate nonormal distributions, *Psychometrika* 48, 465-471.

Wishart, J. (1928) The generalised product moment distribution in samples from a normal multivariate population, *Biometrika* 20A, 32-52.

## Appendix

Notation: $EX^r = \mu_r'$, $E\left(X - \mu_1'\right)^r = \mu_r$, $EX_1^r X_2^s = \mu_{rs}'$, $E\left(X_1 - \mu_{10}'\right)^r \left(X_2 - \mu_{01}'\right)^s = \mu_{rs}$, $E\left(X_1 - \mu_{10}'\right)^r X_2 = \mu_{r1'}$ and when the moments is for $Y$ we use $\lambda$ instead of $\mu$.

*Ok* we have two variables which are linear combination of two variables

$$Y_1 = a_{10} + a_{11}X_1 + a_{12}X_2 \tag{33}$$
$$Y_2 = a_{20} + a_{21}X_1 + a_{22}X_2 \tag{34}$$

The moments we want is

$$\mathbf{m}_Y = \begin{bmatrix} 1 & \lambda'_{10} & \lambda_{20} & \lambda_{30} & \lambda'_{01} & \lambda_{02} \\ \lambda_{03} & \lambda_{11} & \lambda_{21'} & \lambda_{1'2} \end{bmatrix}^T \tag{35}$$

and the moments $X_1$ and $X_2$ are

$$\mathbf{m}_X = \begin{bmatrix} 1 & \mu'_{10} & \mu_{20} & \mu_{30} & \mu'_{01} & \mu_{02} \\ \mu_{03} & \mu_{11} & \mu_{21'} & \mu_{1'2} \end{bmatrix}^T \tag{36}$$

Below we show the relationship between the moments which is used to form **A**.

1. $1 = 1$

2. $\lambda'_{10} = a_{10} + a_{11}\mu'_{10} + a_{12}\mu'_{01}$

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

4. $\lambda_{30} = E\left(Y_1 - \lambda'_{10}\right)^3 = E\left(a_{11}\left(X_1 - \mu'_{10}\right) + a_{12}\left(X_2 - \mu'_{01}\right)\right)^3$
$= E\Big[a_{11}^3\left(X_1 - \mu'_{10}\right)^3 + a_{11}^2a_{12}\left(X_1 - \mu'_{10}\right)^2\left(X_2 - \mu'_{01}\right)$
$+a_{11}a_{12}^2\left(X_1 - \mu'_{10}\right)\left(X_2 - \mu'_{01}\right)2 + a_{12}^3\left(X_2 - \mu'_{01}\right)^3\Big]$
$= a_{11}^3\mu_{30} + 3a_{11}^2a_{12}\mu_{21} + 3a_{11}a_{12}^2\mu_{12} + a_{12}^3\mu_{03}$

5. $\lambda_{11} = E\left(Y_1 - \lambda'_{10}\right)\left(Y_2 - \lambda'_{01}\right) =$
$E\left[\left(a_{11}\left(X_1 - \mu'_{10}\right) + a_{12}\left(X_2 - \mu'_{01}\right)\right)\left(a_{21}\left(X_1 - \mu'_{10}\right) + a_{22}\left(X_2 - \mu'_{01}\right)\right)\right]$
$= E\left[a_{11}a_{21}\left(X_1 - \mu'_{10}\right)\left(X_1 - \mu'_{10}\right) + a_{11}a_{22}\left(X_1 - \mu'_{10}\right)\left(X_2 - \mu'_{01}\right)\right.$
$\left.+ a_{12}a_{21}\left(X_2 - \mu'_{01}\right)\left(X_1 - \mu'_{10}\right) + a_{12}a_{22}\left(X_2 - \mu'_{01}\right)\left(X_2 - \mu'_{01}\right)\right]$
$= a_{11}a_{21}\mu_{20} + a_{11}a_{22}\mu_{11} + a_{12}a_{21}\mu_{11} + a_{12}a_{22}\mu_{02}$
$= a_{11}a_{21}\mu_{20} + \left(a_{11}a_{22} + a_{12}a_{21}\right)\mu_{11} + a_{12}a_{22}\mu_{02}$

6. $\lambda_{21'} = E\left(Y_1 - \lambda'_{10}\right)^2 X_2 = E\left(Y_1^2Y_2 - 2\lambda'_{10}Y_1Y_2 + \lambda'^2_{10}Y_2\right)$
$= \lambda'_{21} - 2\lambda'_{10}\lambda_{11} + \lambda'^2_{10}\lambda'_{01}$
As $\mathbf{m}_X$ contains various central moments and means $\lambda_{21'}$ have to be rewritten as a function of those quantities. From Cramér (1946), p. 263, we have the result:
$\lambda_{11} = \lambda'_{11} - \lambda'_{10}\lambda'_{01}$

and in Cook (1951), where $k$ is the kumulant, $\lambda'_{21} = k_{12}+k_{20}k_{01}+2k_{11}k_{10}+k_{10}^2 k_{01} = \lambda_{21} + \lambda_{20}\lambda'_{01} + 2\lambda_{11}\lambda'_{10} + \lambda_{10}^2\lambda'_{01}$
Subsitute these two results we have
$\lambda_{21'} = \lambda_{21} + \lambda_{20}\lambda'_{01} + 2\lambda_{11}\lambda'_{10} + \lambda_{10}^2\lambda'_{01} - 2\lambda'_{10}\left(\lambda_{11} + \lambda^{'}_{10}\lambda'_{01}\right) + \lambda'^2_{10}\lambda'_{01}$
$= \lambda_{21} + \lambda_{20}\lambda'_{01}$
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) $\alpha_{123} = \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n a_{1i}a_{2j}a_{3k}\beta_{ijk},\ \beta_{ijk} = EZ_iZ_jZ_k,\ Z_1 = (X_1 - \mu'_{10})$ 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.
$\lambda_{21} = \sum_{i=1}^2 \sum_{j=1}^2 \sum_{k=1}^2 a_{1i}a_{1j}a_{2k}\beta_{ijk} = a_{11}a_{11}a_{21}\beta_{111} + a_{11}a_{11}a_{22}\beta_{112} + a_{11}a_{12}a_{21}\beta_{121} + a_{11}a_{12}a_{22}\beta_{122}+$
$a_{12}a_{11}a_{21}\beta_{111} + a_{12}a_{11}a_{22}\beta_{112} + a_{12}a_{12}a_{21}\beta_{121} + a_{12}a_{12}a_{22}\beta_{122} =$
$= (a_{11}a_{11}a_{21} + a_{12}a_{11}a_{21})\,\beta_{111}+(a_{11}a_{11}a_{22} + a_{11}a_{12}a_{21} + a_{12}a_{11}a_{22} + a_{12}a_{12}a_{21})\,\beta_{112}$
$+ (a_{11}a_{12}a_{22} + a_{12}a_{12}a_{22})\,\beta_{122}$
$= (a_{11}a_{11}a_{21} + a_{12}a_{11}a_{21})\,\mu_{30}+(a_{11}a_{11}a_{22} + a_{11}a_{12}a_{21} + a_{12}a_{11}a_{22} + a_{12}a_{12}a_{21})\,\mu_{21}$
$+ (a_{11}a_{12}a_{22} + a_{12}a_{12}a_{22})\,\mu_{12}$

Hence,

$\lambda_{21'} = (a_{11}a_{11}a_{21} + a_{12}a_{11}a_{21})\,\mu_{30}+(a_{11}a_{11}a_{22} + a_{11}a_{12}a_{21} + a_{12}a_{11}a_{22} + a_{12}a_{12}a_{21})\,\mu_{21}$
$+ (a_{11}a_{12}a_{22} + a_{12}a_{12}a_{22})\,\mu_{12}+\left(a_{11}^2\mu_{20} + 2a_{11}a_{12}\mu_{11} + a_{12}^2\mu_{02}\right)(a_{20} + a_{21}\mu'_{10} + a_{22}\mu'_{01})$
$= (a_{11}a_{11}a_{21} + a_{12}a_{11}a_{21})\,\mu_{30}+(a_{11}a_{11}a_{22} + a_{11}a_{12}a_{21} + a_{12}a_{11}a_{22} + a_{12}a_{12}a_{21})\,\mu_{21}$
$+ (a_{11}a_{12}a_{22} + a_{12}a_{12}a_{22})\,\mu_{12} + a_{11}^2\mu_{20}a_{20} + a_{11}^2\mu_{20}a_{21}\mu'_{10} + a_{11}^2\mu_{20}a_{22}\mu'_{01} +$
$2a_{11}a_{12}\mu_{11}a_{20}$
$+2a_{11}a_{12}\mu_{11}a_{21}\mu'_{10}+2a_{11}a_{12}\mu_{11}a_{22}\mu'_{01}+a_{12}^2\mu_{02}a_{20}+a_{12}^2\mu_{02}a_{21}\mu'_{10}+a_{12}^2\mu_{02}a_{22}\mu'_{01}$