Updates to the published version of *A Heuristic for Moment-matching Scenario Generation* by K. Høyland, M. Kaut, and S. W. Wallace

Michal Kaut
michal.kaut@iot.ntnu.no*

May 2003

Abstract

This note describes a new development of the scenario-generation algorithm from the paper "A Heuristic for Moment-matching Scenario Generation", published in *Computational Optimization and Applications*, vol. 24, pp. 169–185, 2003. The presented results lead to a better performance of the algorithm, so the note should be of interest to anybody considering implementing the algorithm.

Throughout the note, we assume that the reader is familiar with the paper, so we can use the notation and refer to parts of the algorithm.

Redundance of Step 4 of the algorithm

In the core (idealised) algorithm, presented in Section 2.4 of the paper, we start the generation with an independent random vector $\tilde{\mathcal{X}}$ with moments $TRSFMOM$. These moments are computed in such a way that $\tilde{\mathcal{Y}} = L\tilde{\mathcal{X}}$ has moments MOM (which then lead to the specified moments $TARMOM$). In the modified algorithm in Section 2.5, the outcomes $\mathbb{X}$ of the random vector $\tilde{\mathcal{X}}$ serve as a starting point for the iterative loop in Step 5 of the algorithm.

Unfortunately, the moments $TRSFMOM$ required for the random vector $\tilde{\mathcal{X}}$ are often quite extreme—they may not even exist. As a result, $\tilde{\mathcal{X}}$ may be both hard to obtain and, even worse, it may lead to "strange" (non-smooth and/or truncated) distributions.

On the other hand, our testing shows that the iterative procedure in Step 5 converges even if we start with a random vector $\tilde{\mathcal{X}}$ with moments different from $TRSFMOM$. In particular, the algorithm works well if we start with marginals $\tilde{\mathcal{X}}_i$ with standard normal distributions. Our recommendation is thus to skip Step 4 of the algorithm (generation of $\tilde{\mathcal{X}}$) altogether, and sample the marginals $\tilde{\mathcal{X}}_i$ from the standard normal distribution.

*Norwegian University of Science and Technology, N-7491 Trondheim, Norway

Instead of sampling, it is also possible to use a pre-defined discretization of the distribution. This variant then leads to an "almost deterministic" algorithm, i.e. it decreases the differences between trees coming from several runs of the algorithm. (Whether the algorithm becomes truly deterministic depends on the implementation of the solution method used for finding coefficients of the cubic transformation. In our case, the solution method contains some randomness, so the algorithm may give different trees even if we start with the same discretizations of $\tilde{\mathcal{X}}_i$.)

The normal distribution is chosen mostly for convenience, since it is easy to sample from. In addition, the normal distribution is smooth, which seems to be important for stability of optimization models—see *Kaut and Wallace, 2003* for a discussion of stability. Even if other distributions have not been tested, we believe that any smooth distribution would work as well.

Problems with low kurtosis

Unfortunately, not all distributions can be obtained by a single cubic transformation of the standard normal distribution: When the kurtosis is too small, the difference between the normal distribution and the target distribution becomes too big.

Throughout this section, we denote the higher moments by[1]

$$\text{skewness: } \gamma = \frac{\mu_3}{\sigma^3}$$
$$\text{kurtosis: } \delta = \frac{\mu_4}{\sigma^4} \ ,$$

where μ_k is the k-th central moment, $\mu_k = \mathbb{E}\big[\big(\tilde{X} - \mathbb{E}\big[\tilde{X}\big]\big)^k\big]$, and $\sigma^2 = \mu_2$. Note that both moments are independent of the value of the mean and variance. For the rest of the section, we thus set mean to zero and variance to one—in conformity with the paper.

First, we should explain what is meant by "too small kurtosis". *Pearson, 1916* proved that, for a given value of skewness γ, there is a lower bound on the possible value of the kurtosis,

$$\delta \geq 1 + \gamma^2 \ .$$

In addition, *Klaassen et al., 2000* showed that, in the case of unimodal distributions, the bound is

$$\delta \geq \frac{189}{125} + \gamma^2 = 1.52 + \gamma^2 \ ,$$

so all distributions between these two bounds are multi-modal. It is thus not surprising that they can not be obtained by a single cubic transformation of the (unimodal) normal distribution.

The easiest remedy of the problem is to repeat the cubic transformation several times. This, together with trying several starting samples, usually solves the problem, at least

[1] The standard notation would be γ_1 for skewness and γ_2 for normalised kurtosis, so our notation is $\gamma = \gamma_1$ and $\delta = \gamma_2 + 3$. The reason for the choice is that we need to divide by the kurtosis later, which is not possible with the standard definition, since γ_2 can be zero.

with our implementation of the cubic transformation. To illustrate the approach, we have tested 50, 000 combinations of skewness and kurtosis, sampled uniformly from

$$\{(\gamma, \delta),\ \gamma \in [0, 10],\ \delta \in [\delta_\gamma, 2\delta_\gamma]\}\ ,$$

where δ_γ denotes the minimal kurtosis, $\delta_\gamma = 1 + \gamma^2$. For every combination, we start with a sample of 10, 000 outcomes[2] from the standard normal distribution, and try to transform the sample to a distribution with the given skewness and kurtosis, using the cubic transformation. When we do not obtain the desired moments after ten transformations, we mark the combination as inaccessible, otherwise we store the number of transformations.

From the 50, 000 combinations, only 55 were not obtained in 10 transformations—and all were very close to the lower bound: Table 1 presents the maximal and average distances from the bound, both in absolute $(\delta - \delta_\gamma)$ and relative $({}^{\delta}/_{\delta_\gamma} - 1)$ values. For comparison: from the combinations we were able to obtain, the one closest to the lower bound had the absolute distance of 0.014, and relative distance of 0.01%.

Table 1: Distance of the combinations (γ, δ) we were not able to generate, from the theoretical lower bound (γ, δ_γ).

	distance from bound	
statistics	absolute	relative
average	0.015	0.17%
max	0.051	0.93%

We have also created a "map" showing the number of transformations needed to achieve different combinations of skewness and kurtosis. To obtain the map, we had to run the test on the whole region, not only along the bound as in the previous test. The result of the test is presented in Figure 1. An interesting observation is that there is also an upper bound for kurtosis that can be achieved by a single cubic transformation of the standard normal distribution—as far as we know, this has not been reported before. With our implementation, the upper bound is approximately 88.5. (There is a corresponding bound for two transformations, but it starts at kurtosis of about 3000 for zero skewness and increases to more than 4000 for skewness of 50. Hence, the maximal achievable kurtosis can be seen as unlimited for most practical purposes.)

It is important to realize that the repetition of the cubic transformation would be impossible with the original formulation from *Fleishman, 1978*, since it assumes that the starting distribution is *exactly* normal. Our implementation, however, allows starting with arbitrary distribution, as long as we can compute the first twelve moments.[3] This makes the computation of the coefficients more difficult, but on the other hand practically eliminates

[2]The relatively high number of scenarios was chosen in order to ensure that the starting discretization is sufficiently close to the standard normal distribution.

[3]This is why we can use it in the iterative loop in Step 5 of the algorithm.

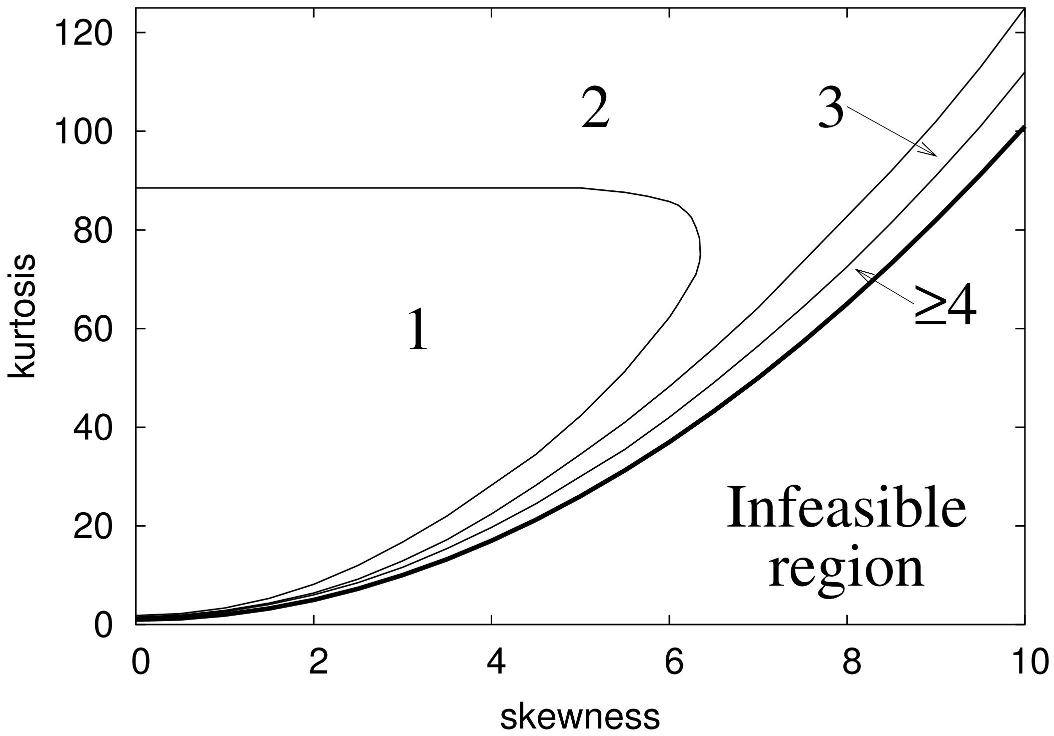

Figure 1: Combinations of skewness and kurtosis accessible by a repeated cubic transformation of the standard normal distribution. The numbers show the number of cubic transformations needed to obtain distributions from the corresponding areas. The lowest region contains infeasible combinations of skewness and kurtosis.

the biggest problem of Fleishman's method, the inability of generating distributions with low kurtosis.

An alternative to repeating the cubic transformation is to start with a different distribution than the standard normal. For example, if the kurtosis is below the bound for unimodal distributions, we may try a mixture of two normal distributions: Kurtosis decreases with the distance of the two components of the mixture. In some occasions, uniform distribution may help—this is, however, not recommended because of the non-smoothness of the distribution.

As a more sophisticated alternative, we may consider using a different method than the cubic transformation to obtain the starting value $\mathbb{Y}_i$ for the random variable $\tilde{\mathcal{Y}}_i$ in Step 5 of the algorithm. A good source of information is *Tadikamalla, 1980*, who presents and compares six different methods, where three are capable of generating distributions with any feasible combination of skewness and kurtosis: the Johnson system of distributions from *Johnson, 1949*, the Tadikamalla-Johnson system from *Tadikamalla and Johnson, 1979*, and the Schmeiser-Deutch system from *Schmeiser and Deutch, 1997*.

Note that the special approach—either repeating the cubic transformation, or using some of the mentioned alternatives—is needed only in order to get a starting point for the iterative procedure in Section 5. Inside the loop, we always use the cubic transformation: Even if the cubic transformation is not capable of entering the low-kurtosis region, it works inside it.

New implementation

As mentioned in the "Future work" part of the paper, we have implemented the algorithm in the `C` programming language. The cubic transformation, which is the crucial part of the algorithm, was implemented by Diego Mathieu from INSA Toulouse, France, during his visit at Molde University College in the summer of 2002. The code has been compiled for `Win32` using both `Microsoft Visual C++` and `MinGW` (`GCC` for `Win32`), so it should compile on other platforms as well.

The new implementation provides two major improvements compare to the original `AMPL` implementation: It is more than ten times faster, and it is a stand-alone code, so it is not dependent on any commercial solver.

The issue of distributions with low kurtosis has been addressed by allowing several repetitions of the cubic transformation. Diego Mathieu has also implemented the Schmeiser-Deutch system from *Schmeiser and Deutch, 1997*, but it has not yet been included in the code—mostly because we have not yet encountered a real case where we could not generate the scenarios using the current implementation.

References

Fleishman, A. I. (1978). A method for simulating nonnormal distributions. *Psychometrika*, 43:521–532.

Johnson, N. L. (1949). Systems of frequency curves generated by methods of translation. *Biometrika*, 36(1):149–176.

Kaut, M. and Wallace, S. W. (2003). Evaluation of scenario-generation methods for stochastic programming. Stochastic Programming E-Print Series, `http://www.speps.info`.

Klaassen, C., Mokveld, P., and Es, B. V. (2000). Squared skewness minus kurtosis bounded by 186/125 for unimodal distributions. *Statistics & Probability Letters*, 50:131–135.

Pearson, K. (1916). *Mathematical Contributions to the Theory of Evolution, XIX: Second Supplement to a Memoir on Skew Variation.* Ser. A. Philos. Trans. Roy. Soc., London.

Schmeiser, B. W. and Deutch, S. J. (1997). A versatile four parameter family of probability distributions suitable for simulation. *AIIE Transactions*, 9(2):176–182.

Tadikamalla, P. and Johnson, N. L. (1979). Systems of frequency curves generated by transformations of logistic variables. Mimeo Series 1226, Department of Statistics, University of North Carolina at Chapel Hill, Chapel Hill.

Tadikamalla, P. R. (1980). On simulating non-normal distributions. *Psychometrika*, 45(2):273–279.