

A new moment matching algorithm for sampling from partially specified symmetric distributions

P. Date ^{a *}, R. Mamon ^b, L. Jalen^a

^aCenter for the Analysis of Risk and Optimization Modelling Applications, Department of Mathematical Sciences, Brunel University, United Kingdom

^bDepartment of Statistical and Actuarial Sciences, University of Western Ontario, Canada.

A new algorithm is proposed for generating scenarios from a partially specified symmetric multivariate distribution. The algorithm generates samples which match the first two moments exactly and match the marginal fourth moments approximately, using a semidefinite programming procedure. The performance of the algorithm is illustrated by a numerical example.

Keywords: scenario generation, semidefinite programming

1. Introduction

The problem of sampling from a partially specified multivariate distribution arises in many different areas. The work presented here was motivated by stochastic programming based optimization models in operations research, in which the key computational challenge is to generate scenarios from a distribution of the underlying random variables. For a large number of random variables, the scenario generation can be computationally very challenging. The distribution to be sampled may not be available in closed form and it may instead be characterized by moments obtained from empirical data. Even if the distribution is available in closed form, it may be very difficult to sample and an approximation may be necessary. Various heuristic methods exist for dealing with generation of scenarios under partially specified distributions. The literature on sampling from such probability distributions is too large to review here and we restrict our attention to the approaches only used in operations research and finance. These approaches can be roughly divided into two main classes:

- Under the first approach, the statistical

*corresponding author. mastpmd@brunel.ac.uk, Department of Mathematical Sciences, Brunel University, Middlesex, UB8 3PH, UK.

properties of the joint distribution are specified in terms of moments, usually including the covariance matrix. In [7], cubic transformation of univariate, standard normal random variables and Cholesky factorization of covariance matrix are used to produce a multivariate distribution which approximately matches a given set of marginal central moments and the covariance matrix. Similar moment matching approach is employed to generate probability weights and support points using non-convex optimization in [4]. In [14], entropy maximization method is used to generate a discrete approximation to a given continuous distribution.

- In the second approach, specified (parametric) marginal distributions are sampled independently and the samples are then used along with Cholesky factorization of the covariance matrix to generate the necessary multivariate distribution. An iterative procedure of this type is described in [11].

Other approaches to scenario generation with specific emphasis on operations research applications include principal component analysis-based simulation [15] and stochastic approximation based on transportation metrics ([13], [5]).

A detailed survey of different scenario generation methods appears in [10].

Despite the successes of the approaches described above in the practical applications of stochastic optimization, the procedures involved in the approaches of drawing samples from a partially specified multivariate distribution when the distribution is specified in terms of moments have certain limitations as detailed below:

1. All the moment-matching procedures in the above mentioned papers use non-convex optimization to generate scenarios which match a specified set of statistical properties, in addition to a needed factorization of the covariance matrix. Given a univariate random variable with known first 12 central moments, the approach used in [7] and [8] finds a cubic polynomial function of this random variable which has the required four central moments. This requires a non-convex optimization in terms of the coefficients of the polynomial. The procedure has to be repeated iteratively for each marginal distribution. Similarly, the algorithm in [11] requires a non-convex optimization over the space of lower triangular matrices.
2. The achieved moments of the generated samples match the target moments only approximately. There are two sources of error in these moment matching methods: one is due to the fact that only local optima are found for the non-convex optimization problem and the other is the inexact starting moments of samples of univariate random variables. Since these procedures employ samples from a known, “simple” univariate distribution, the achieved moments usually depend on the sample moments of univariate random variables used.

The primary objective of this technical note is to develop an algorithm based on *convex optimization* which *matches exactly* the mean, covariance matrix and marginal (zero) skewness of a symmetric distribution and also matches the marginal fourth moments approximately (by minimizing the worst case error between the achieved and

the target marginal fourth moments). An analytic solution to this optimization is known in the scalar case, as illustrated in section 2.3.

This algorithm may be used as a scenario generator on its own or its scalar version may be adopted to produce an initial guess for the optimization routines proposed by other authors. Being able to match a small set of statistical properties exactly, possibly with a very small set of scenarios, may be preferable to generating a very large number of scenarios to model the entire distribution. This is especially true when the scenarios are to be used in stochastic optimization procedures.

The rest of the paper is organized as follows. The next section introduces the notation. We develop in sections 2.2-2.3 the main sampling algorithm of this paper and provide a discussion of its properties. Section 3 presents a numerical study demonstrating the utility and efficiency of the algorithm. Finally section 4 concludes and outlines the directions of further research.

2. The sampling algorithm

2.1. Notation

We first outline here the notation used in our development of the algorithm.

| | |
|---------------------------|---|
| n | number of random variables |
| s | number of scenarios |
| \mathcal{X} | discrete n -dimensional random variable |
| \mathcal{X}_i | i^{th} random variable |
| Φ | target mean vector for \mathcal{X} |
| \mathbf{R} | target covariance matrix for \mathcal{X} |
| κ_i | target marginal 4 th central moment for the i^{th} random variable \mathcal{X}_i |
| \mathbf{L}_{ij} | entry in the i^{th} row and j^{th} column of a matrix \mathbf{L} |
| $\mathbb{P}(A)$ | probability of an event A |
| $\mathbb{E}[\mathcal{Y}]$ | expected value of a random variable \mathcal{Y} |

$\mathbf{1}_s$ denotes s -dimensional vector with all entries 1. $\mathbf{diag}(x_i)^{(s)}$ denotes a diagonal matrix with x_1, x_2, \dots, x_s on the diagonal. For a symmetric matrix \mathbf{P} , $\mathbf{P} \geq 0$ indicates that the matrix is

positive semi-definite, i.e. has all non-negative eigenvalues.

To re-iterate our objective, we aim to generate samples from a symmetric distribution with a specified mean vector Φ and a specified (positive definite) covariance matrix \mathbf{R} . These target moments will usually be obtained from the data. If the covariance matrix obtained from the data is not positive definite, an adjustment may be necessary, such as the one suggested in [11]. In addition, we wish to minimize the worst case mismatch between the achieved marginal fourth moments and the target marginal fourth moments. We will describe the algorithm from an optimization point of view first and then provide the closed-form solution in the scalar case. The rationale behind the key steps in the algorithm will become clear from the proofs of subsequent results and accompanying discussion.

2.2. Algorithm for moment matching scenario generation

- (i) Find a symmetric positive definite matrix \mathbf{L} such that $\mathbf{R} = \mathbf{L}\mathbf{L}^\top$. For a symmetric positive definite \mathbf{R} , \mathbf{L} is unique. If \mathbf{R} has distinct eigenvalues, this may be found using singular value decomposition; see, e.g. [6] and the references therein for methods of finding \mathbf{L} . This matrix \mathbf{L} is usually referred to as the *square root* of the matrix \mathbf{R} .
- (ii) Solve the following optimization problem:

$$\min_{\epsilon, q_1, q_2, \dots, q_s} \epsilon \quad (1)$$

subject to

$$\begin{bmatrix} \epsilon & \psi_i \\ \psi_i & \epsilon \end{bmatrix} \geq 0, \quad i = 1, 2, \dots, n, \quad (2)$$

$$\mathbf{diag}(q_k)^{(s)} \geq 0, \quad (3)$$

$$\begin{bmatrix} 1 & \mathbf{1}_s^\top \\ \mathbf{1}_s & \frac{1}{2n} \mathbf{diag}(q_k)^{(s)} \end{bmatrix} \geq 0, \quad (4)$$

where $\psi_i = \frac{1}{2s^2} \sum_{j=1}^n \mathbf{L}_{ij}^4 \sum_{k=1}^s q_k - \kappa_i$, $i = 1, 2, \dots, n$. Note that this is a convex problem with a linear objective and affine matrix inequality (AMI) constraints. These

problems can be solved in polynomial time using interior point methods and extensive software packages are available to implement interior point methods for solving convex problems of this type (which are also called semidefinite programming problems); see [2], [12] and [16]. Let $\hat{q}_k, k = 1, 2, \dots, s$ and $\hat{\epsilon}$ be the arguments which solve the above problem within a specified degree of accuracy.

- (iii) Set $p_i = \frac{1}{\hat{q}_i}$, $i = 1, 2, \dots, s$ and $p_{s+1} = 1 - 2n \sum_{i=1}^s p_i$.
- (iv) Define a discrete n -dimensional random variable \mathcal{X} over a support of $2ns + 1$ points as follows:

$$\begin{aligned} \mathbb{P}\left(\mathcal{X} = \Phi \pm \frac{1}{\sqrt{2sp_i}} \mathbf{L}_j\right) &= p_i, \\ j &= 1, 2, \dots, n, \quad i = 1, 2, \dots, s, \\ \mathbb{P}(\mathcal{X} = \Phi) &= p_{s+1}. \end{aligned} \quad (5)$$

where \mathbf{L}_j denotes the j^{th} column of a matrix \mathbf{L} .

Steps (i)-(iv) constitute the entire set of procedures needed to construct the required samples. Before we prove that it has the required moment properties, we need to show that $p_i, i = 1, 2, \dots, s + 1$ defines a probability measure over the chosen $2ns + 1$ support points. Since \hat{q}_i satisfy (3), $\hat{q}_i = \frac{1}{p_i} > 0 \forall i$. It remains to be shown that $p_{s+1} = 1 - 2n \sum_{i=1}^s p_i$ is non-negative. This is demonstrated in the following lemma:

Lemma 1 For \hat{q}_i defined as above, $2n \sum_{i=1}^s \hat{q}_i^{-1} \leq 1$.

Proof: The proof depends on the well-known property of positive definite block matrices:

$$\begin{aligned} \mathbf{M} := \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix} \geq 0 \text{ and } \mathbf{D} \geq 0 \\ \Leftrightarrow \mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C} \geq 0. \end{aligned}$$

The block matrix $\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C}$ is called the *Schur complement* of \mathbf{D} in \mathbf{M} ; see, e.g. [2] for details.

In the present case, let $\mathbf{Q} = \frac{1}{2n} \mathbf{diag}(\hat{q}_k)^{(s)}$. Since $\hat{q}_k, k = 1, 2, \dots, s$ satisfies (3)-(4), we have

$$\begin{aligned} \begin{bmatrix} 1 & \mathbf{1}_s^\top \\ \mathbf{1}_s & \mathbf{Q} \end{bmatrix} \geq 0 \text{ and } \mathbf{diag}(\hat{q}_k)^{(s)} \geq 0 \\ \Leftrightarrow 1 - \mathbf{1}_s^\top \mathbf{Q}^{-1} \mathbf{1}_s \geq 0, \end{aligned}$$

from which the result follows. \blacksquare

Next, we establish a relationship between the optimal argument $\hat{\epsilon}$ (which also equals the optimal cost in (2)) and the target fourth marginal moments κ_i .

Lemma 2 For \mathcal{X} defined as above,

$$\max_i |\kappa_i - \mathbb{E}(\mathcal{X}_i - \Phi_i)^4| \leq \hat{\epsilon}. \quad (6)$$

Proof: Note that

$$\begin{aligned} \mathbb{E}[(\mathcal{X}_i - \Phi_i)^4] &= \sum_{k=1}^s p_k \frac{1}{2p_k^2 s^2} \sum_{j=1}^n \mathbf{L}_{ij}^4 \\ &= \left(\sum_{k=1}^s \hat{q}_k \right) \frac{1}{2s^2} \left(\sum_{j=1}^n \mathbf{L}_{ij}^4 \right). \end{aligned}$$

Let $\hat{\psi}_i = \frac{1}{2s^2} \sum_{j=1}^n \mathbf{L}_{ij}^4 \sum_{k=1}^s \hat{q}_k - \kappa_i$. The result then follows by observing that $\hat{q}_k, \hat{\epsilon}$ satisfy constraint (2) and

$$\begin{bmatrix} \epsilon & \hat{\psi}_i \\ \hat{\psi}_i & \epsilon \end{bmatrix} \geq 0 \Leftrightarrow \epsilon^2 - \hat{\psi}_i^2 \geq 0 \quad \blacksquare$$

Lastly, we collect together all the moment matching properties of \mathcal{X} in the following result:

Theorem 1 The distribution defined in (5) satisfies the following properties:

$$\mathbb{E}[\mathcal{X}] = \Phi, \quad (7)$$

$$\mathbb{E}[(\mathcal{X} - \Phi)(\mathcal{X} - \Phi)^\top] = \mathbf{R}, \quad (8)$$

$$\mathbb{E}[(\mathcal{X}_i - \Phi_i)^3] = \mathbf{0}, \quad (9)$$

$$\max_i |\kappa_i - \mathbb{E}(\mathcal{X}_i - \Phi_i)^4| \leq \hat{\epsilon}. \quad (10)$$

Proof: Equations (7) and (9) are immediate due to the symmetry of the support points around

the target mean vector Φ and (10) was proven in lemma 2. Equation (8) follows by noting that

$$\begin{aligned} \mathbb{E}[(\mathcal{X} - \Phi)(\mathcal{X} - \Phi)^\top] &= 2 \sum_{i=1}^s \frac{p_i}{2sp_i} \left(\sum_{j=1}^n \mathbf{L}_j \mathbf{L}_j^\top \right) \\ &= \sum_{j=1}^n \mathbf{L}_j \mathbf{L}_j^\top = \mathbf{R} \quad \blacksquare \end{aligned}$$

Several remarks concerning the above results are in order.

- (a) Note that the optimization problem (2)-(4) finds the smallest $\hat{\epsilon}$ and the corresponding \hat{q}_k such that (10) holds. In other words, the algorithm minimizes an upper bound on the worst case error in matching the fourth marginal moment. A small value of $\hat{\epsilon}$ indicates that the fourth moment is approximately matched (with the maximum approximation error being $\hat{\epsilon}$). This upper bound can be made zero in the scalar case, as will be seen in the next section.
- (b) Even if the chosen q_k are not optimal, (7)-(9) will still hold provided q_k 's satisfy the condition in lemma 1 to define a probability measure. If we are not concerned with matching the fourth marginal moment, we may choose *not* to solve the optimization problem and choose *any* q_k such that the condition in lemma 2 holds, *e.g.*, we can choose $q_k > 2ns, \forall k$ which automatically satisfies the required condition. The actual choice of q_k subject to the lower bound $2ns$ can be made using any deterministic or stochastic algorithm. This provides s additional degrees of freedom, which may, in principle, be used to match other statistical properties (e.g. certain quantiles of interest). We have restricted our attention to matching fourth marginal moment only since matching these moments is relevant from a practical point of view and the associated optimization, being convex, is numerically tractable.
- (c) The downside, of course is that the algorithm is limited to symmetric distributions.

However, even in cases when the underlying distribution is known to be asymmetric, the proposed algorithm may still have a useful role to play. In computation or optimization of risk of a financial portfolio, the leptokurtic behavior of the loss distribution is often far more important than the asymmetry and a symmetric approximation which captures the tail behavior of loss distribution correctly may be admissible.

- (d) In a rather unrelated field, similar sample point generation methods are also employed in the development of sigma point filters (also called unscented filters) widely adopted in engineering; see [9] and the references therein. These methods have become quite popular as a computationally cheaper alternative to particle filters for state estimation problems in nonlinear systems. However, the sampling methods in existing sigma point filtering techniques do not guarantee that the weights assigned to each sample point will always be non-negative. Our proposed algorithm for sampling distributions avoids this problem and its application in sigma point filtering has now been reported in [3].

2.3. Closed-form solution in the scalar case

Having shown that finding positive q_k satisfying $2n \sum_{k=1}^s q_k^{-1} < 1$ and minimizing the worst case error in matching the fourth marginal moment is a convex optimization problem, a natural question to ask is whether it is possible to find a closed-form solution to this problem in specific instances. As mentioned earlier, choosing $q_k > 2ns \forall k$ will automatically satisfy the necessary constraint on the sum of q_k^{-1} . At this stage, it still remains to be seen whether we can choose $q_k > 2ns$ which will also satisfy the condition for matching the fourth moment, *i.e.*, whether we can choose q_k such that $\kappa_i = \mathbb{E}(\mathcal{X} - \Phi)^4$ holds. The answer is affirmative in the scalar case, *i.e.*, when $n = 1$, as shown in the next result.

Lemma 3 *Suppose that $n = 1$ and that $\frac{\kappa_1}{\mathbf{L}_{11}^4} > 1$. Let $\hat{q}_i \in [2sn, \psi]$, $i = 1, 2, \dots, s - 1$ be $s - 1$ real*

numbers, where the constant ψ is given by

$$\psi = \frac{2s^2 \kappa_1}{(s-1)\mathbf{L}_{11}^4} - \frac{2s}{s-1},$$

$$\text{and let } \hat{q}_s = \frac{2s^2 \kappa_1}{\mathbf{L}_{11}^4} - \sum_{i=1}^{s-1} \hat{q}_i. \quad (11)$$

Finally, let \mathcal{X} as in (5), with $n = 1$. Then the random variable \mathcal{X} satisfies the properties (7)-(9). In addition, $\kappa_1 = \mathbb{E}(\mathcal{X} - \Phi)^4$ holds.

Proof: Verifying that \mathcal{X} satisfies (7)-(9) is straightforward. To verify that $\kappa_1 = \mathbb{E}(\mathcal{X} - \Phi)^4$ holds, first note that $\hat{q}_k \leq \psi, i = 1, 2, \dots, s - 1$ ensures that $q_s > 2ns$ and

$$\mathbb{E}(\mathcal{X} - \Phi)^4 = \sum_{k=1}^s \frac{p_k \mathbf{L}_{11}^4}{2p_k^2 s^2} = \left(\sum_{k=1}^s \hat{q}_k \right) \frac{\mathbf{L}_{11}^4}{2s^2}.$$

The result then follows due to definition of \hat{q}_s in (11). ■

Remark 1 *Note that the condition $\frac{\kappa_1}{\mathbf{L}_{11}^4} > 1$ is not particularly restrictive, and is in fact satisfied by all elliptic distributions including Gaussian distribution and t -distribution; see [1].*

The above result gives an optimization-free methodology of matching the first four moments of a symmetric scalar random variable. This fact is quite important in itself and our algorithm as proposed above can be used as an efficient alternative to cubic transformation-based approaches for generating random samples which match a given set of four central moments. Furthermore, there is a lot of extra freedom in the choice of q_i which may be utilized to match further higher moments. Alternatively, q_i may be generated using any random number generator or using an appropriate deterministic algorithm.

3. Numerical experiments

To test the computational efficiency of the optimization procedure, we used LMI toolbox of MATLAB (version 6.5), running on a desktop with a 3 GHz Pentium processor. To derive a covariance matrix and marginal kurtosis, which is

guaranteed to correspond to a feasible distribution, we used MATLAB’s random number generator for t-distribution with 10 degrees of freedom. The sample covariance matrix of the resulting random samples was used as a target covariance matrix and the sample marginal kurtosis values were used as the target kurtosis in our optimization. We ran the numerical experiments for various combinations of number of variables (n), scenarios ($2ns + 1$) and kurtosis values. Some of the results are reported in Table 1 with $\hat{\epsilon}$ defined in theorem 1. We have reported only the mean target kurtosis, rather than the individual kurtosis values, for brevity. Note that the mean vector, the covariance matrix and the zero skewness are exactly matched in all cases. The specific choice of these first three moments has very little impact on the matching of the fourth marginal moment.

It can be seen that it took less than 15 seconds to generate 7201 samples for 60 random variables. The worst kurtosis matching error over the scenarios and dimensions under consideration was around 15%. The *average* error between the target and the achieved kurtosis values over n dimensions was significantly smaller and was under 5% in all cases. The computation times can easily be improved by employing a higher specification machine and a purpose-written optimization code, e.g. which exploits the sparsity in (4).

Table 1 : Results of numerical experiments.

| n | s | $\frac{1}{n} \sum_i \kappa_i$ | $\hat{\epsilon}$ | time in seconds |
|-----|-----|-------------------------------|------------------|-----------------|
| 2 | 20 | 5.6118 | 0.3037 | 0.16 |
| 4 | 5 | 6.1847 | 0.4875 | 0.03 |
| 10 | 2 | 6.3868 | 0.6344 | 0.04 |
| 50 | 50 | 6.0837 | 0.8347 | 8.01 |
| 60 | 60 | 6.2017 | 0.7547 | 14.21 |

4. Future research

Our proposed method deals only with single stage scenarios. An extension of this algorithm to generation of scenario trees for multi-stage decision problems and an implementation of a large scale stochastic programming model demonstrating the use of this method in financial opti-

mization are topics of ongoing research. From a theoretical point of view, the relationship between the proposed optimization procedure and the semi-definite optimization procedures to determine whether a given vector of moments is feasible (e.g., as discussed in chapter 16 of [16]) is worth investigating.

Acknowledgement

A part of this research was carried out when the first author was visiting University of Western Ontario, Canada on an Overseas Travel grant from Royal Society, UK.

References

- [1] P.M. Bentler and M. Berkane. Greatest lower bound to the elliptical theory kurtosis parameter, *Biometrika* 73 (1986) 240-241.
- [2] S. Boyd, L.El Ghaoui, E. Feron, and V. Balakrishnan, *Linear Matrix Inequalities in System and Control Theory*, SIAM, 1994.
- [3] P. Date, L. Jalen, and R. Mamon, A new algorithm for latent state estimation in nonlinear time series models, *Applied Mathematics and Computation* (2008), in press.
- [4] N. Gulpinar, B. Rustem, and R. Settergren, Optimisation and simulation approaches to scenario tree generation, *Journal of economic dynamics and control* 28 (2004), 1291-1315.
- [5] R. Hochreiter and G.Ch. Pflug, Financial scenario generation for stochastic multi-stage decision processes as facility location problems, *Annals of Operations Research* 152 (2007), 257-272.
- [6] R.A. Horn and C.R. Johnson, *Matrix Analysis*, Cambridge University Press, 1999.
- [7] K. Høyland, M. Kaut, and S.W. Wallace, A heuristic for moment matching scenario generation, *Computational Optimization and Applications* 24 (2003), 169-185.

- [8] K. Høyland and S.W. Wallace, Generating scenario trees for multistage decision problems, *Management Science* 47 (2001), 295-307.
- [9] S.J. Julier and J.K. Uhlmann, Unscented filtering and nonlinear estimation, *Proceedings of the IEEE* 92 (2004), 401-422.
- [10] M. Kaut and S.W. Wallace, Evaluation of scenario generation methods for stochastic programming, *Pacific Journal of Optimization* 3 (2007), 257-271.
- [11] P.M. Lurie and M.S. Goldberg, An approximate method for sampling correlated random variables from partially specified distributions, *Management Science* 44 (1998), 203-218.
- [12] Mathworks Inc. *LMI control toolbox manual*, 1995.
- [13] G.Ch. Pflug, Scenario tree generation for multiperiod financial optimization by optimal discretization, *Mathematical Programming* 89 (2001), 251-271.
- [14] J.E. Smith, Moment methods for decision analysis, *Management Science* 39 (1993), 340-358.
- [15] N. Topaloglou, H. Vladimirou, and S.A. Zenios, CVaR models with selective hedging for international asset allocation, *Journal of banking and finance* 26 (2002), 1535-1561.
- [16] H. Wolkowicz, R. Saigal, and L. Vandenberghe, *Handbook of semidefinite programming: theory, algorithms and applications*, Kluwer, 2000.