ARTICLE IN PRESS
JID: EOR [m5G;March 3, 2018;12:45]
European Journal of Operational Research 000 (2018) 1–5
Contents lists available at ScienceDirect
European Journal of Operational Research
journal homepage: www.elsevier.com/locate/ejor

Short Communication

Comment on "An algorithm for moment-matching scenario generation with application to financial portfolio optimization"

Juan Pablo Contreras, Paul Bosch*, Mauricio Herrera

Facultad de Ingeniería, Universidad del Desarrollo, Ave. Plaza 680, Las Condes, Santiago de Chile, Chile

ARTICLE INFO

Article history:
Received 29 August 2016
Accepted 13 February 2018
Available online xxx

Keywords:
Scenarios
Finance
Moment matching

ABSTRACT

A paper by Ponomareva, Roman, and Date proposed a new algorithm to generate scenarios and their probability weights matching exactly the given mean, the covariance matrix, the average of the marginal skewness, and the average of the marginal kurtosis of each individual component of a random vector. In this short communication, this algorithm is questioned by demonstrating that it could lead to spurious scenarios with negative probabilities. A necessary and sufficient condition for the appropriate choice of algorithm parameters is derived to correct this issue.

1. Introduction

Stochastic programming (*SP*) is a framework to model decision making under uncertainty. In *SP* some parameters are uncertain and described by statistical distributions. In cases when these distributions are unknown, it is usual to build a discrete approximation with a finite number of outcomes or *scenarios*, by taking advantage of the fact that probability distributions governing the data are known or can be estimated. An overview of scenario generation procedures for *SP* was given by Kaut and Wallace in Kaut and Wallace (2007).

An important class of scenario generation methods is based on matching a set of statistical properties, e.g. moments. Generally, the mean vector and covariance matrix computed from data are the usual statistics to match. However, higher order moments can also be used in order to get better approximation. The skewness and the kurtosis are some of the higher order moments that can be matched by using, for example, complex non-convex optimization problems (Høyland & Wallace, 2001).

In Date, Jalen, and Mamon (2008), a new sigma point filtering algorithm for state estimation in non-linear time series and non-Gaussian systems was developed. This algorithm combines numerical simplicity of the ensemble filter, along with exact moment matching properties. A modified version of this algorithm was proposed in Ponomareva and Date (2013) in which sigma points and the corresponding probability weights were modified at each step to match exactly the predicted values for the average marginal skewness and the average marginal kurtosis, besides matching the mean and covariance matrix. In Ponomareva, Roman, and Date (2015), this algorithm for generating scenarios and the corresponding probability weights was used in the context of financial optimization in order to decrease the computational complexity and th e time for solving *SP* problems. However, in the works of Ponomareva et al. (2015), some assumptions were inconsistent because several important considerations were simply overlooked. As a direct consequence of incorrect estimates for certain bounds, using the proposed algorithm, we obtained some scenarios with negative probabilities. In this short communication, we introduce a necessary and sufficient condition to be imposed on data and parameters in order to solve this issue. In addition, we suggest some changes in specific steps of the algorithm to enable it to deploy correctly.

This paper is organized as follows: Section 2 briefly reviews the algorithm given in Ponomareva et al. (2015). In Section 3, we present the main comment regarding the work by Ponomareva et al. (2015). Section 4 summarizes our proposition for a modified algorithm. In Section 5, we deploy and compare both algorithms on a data set from the Chilean stock exchange to illustrate all comments set out in Section 3. Conclusions are drawn in Section 6.

2. Algorithm for generating scenarios by matching: mean, covariance matrix, average marginal skewness and average marginal kurtosis

The method proposed in Ponomareva et al. (2015) generates $2ns + 3$ scenarios and their corresponding probabilities, where n is the dimension of the random vector and s is a positive integer chosen by the user. The inputs for the algorithm are: the mean vector

* Corresponding author.
E-mail addresses: juan.contrerasff@usach.cl (J.P. Contreras), pbosch@udd.cl (P. Bosch), mherrera@ingenieros.udd.cl (M. Herrera).

https://doi.org/10.1016/j.ejor.2018.02.028
0377-2217/© 2018 Elsevier B.V. All rights reserved.

μ, the covariance matrix Σ, and the average of third and fourth moments denoted by $\bar{\xi}$ and $\bar{\kappa}$, respectively.

The method begins by choosing a vector $Z \in \mathbb{R}^n$ such that $\Sigma - ZZ^\top \succ 0$, and then a square matrix L is chosen such that $\Sigma = LL^\top + ZZ^\top$. Then, the s numbers $p_k \in (0,1)$, $k = 1, \ldots, s$ are chosen such that

$$\sum_{k=1}^{s} p_k < \frac{1}{2n} \qquad \wedge \qquad \sum_{k=1}^{s} \frac{1}{p_k} < \gamma, \tag{1}$$

where the upper bound γ is defined as

$$\gamma = 2s^2 \frac{n\bar{\kappa} - \frac{3}{4}\left(\sum_{i=1}^n Z_i^4\right)\left(\frac{n\bar{\xi}}{\sum_{i=1}^n Z_i^3}\right)^2}{\sum_{i,j} L_{ij}^4}. \tag{2}$$

In addition, the following parameters are defined:

- An extra probability $p_{s+1} = 1 - 2n\sum_{k=1}^s p_k$
- Auxiliary parameters $\phi_1 = \frac{n\bar{\xi}\sqrt{p_{s+1}}}{\sum_{i=1}^n Z_i^3}$ and $\phi_2 = p_{s+1}\frac{n\bar{\kappa} - \frac{1}{2s^2}(\sum_{i,j} L_{ij}^4)(\sum_{k=1}^s \frac{1}{p_k})}{\sum_{i=1}^n Z_i^4}$
- The symmetric parameters $\alpha = \frac{1}{2}(\phi_1 + \sqrt{4\phi_2 - 3\phi_1^2})$ and $\beta = \frac{1}{2}(-\phi_1 + \sqrt{4\phi_2 - 3\phi_1^2})$
- The coefficients $\omega_0 = 1 - \frac{1}{\alpha\beta}$, $\omega_1 = \frac{1}{\alpha(\alpha+\beta)}$ and $\omega_2 = \frac{1}{\beta(\alpha+\beta)}$

The first $2ns$ scenarios of the distribution and their corresponding probabilities are fixed as:

$$\mathbb{P}\left\{X_{ik}^+ = \mu + \frac{1}{\sqrt{2sp_k}}L_i\right\} = \mathbb{P}\left\{X_{ik}^- = \mu - \frac{1}{\sqrt{2sp_k}}L_i\right\}$$
$$= p_k, \quad i = 1, \ldots, n; k = 1, \ldots, s,$$

were L_i represents the i-column of matrix L. Then, three extra scenarios are added, $\mathbb{P}\{X_0 = \mu\} = p_{s+1}w_0$, and

$$\mathbb{P}\left\{X_\alpha = \mu + \frac{\alpha}{\sqrt{p_{s+1}}}Z\right\} = p_{s+1}w_1, \qquad \mathbb{P}\left\{X_\beta = \mu - \frac{\beta}{\sqrt{p_{s+1}}}Z\right\} = p_{s+1}w_2. \tag{3}$$

Finally, Proposition 1 in Ponomareva et al. (2015) summarizes the properties of the generated discrete asymmetric distribution denoted by $\mathcal{X}$. The authors showed that $\mathbb{E}[\mathcal{X}] = \mu$, $\mathbb{E}[(\mathcal{X} - \mu)(\mathcal{X} - \mu)^\top] = \Sigma$, and:

i. $2n\sum_{k=1}^s p_k + p_{s+1}\sum_{i=0}^3 w_i = 1$,
ii. $\sum_{i=1}^n \sum_{k=1}^{2ns+3} \mathbb{E}(\mathcal{X}_{ij} - \mu_i)^3 = n\bar{\xi}$,
iii. $\sum_{i=1}^n \sum_{k=1}^{2ns+3} \mathbb{E}(\mathcal{X}_{ij} - \mu_i)^4 = n\bar{\kappa}$.

Thus, the average of the marginal skewness and the average of the marginal kurtosis are matched.

3. Comment on the previous algorithm

Our main concerns regarding the work by Ponomareva et al. (2015) are summarized in three points, which will be discussed separately.

3.1. Constraints over data

The condition $\sum_{k=1}^s \frac{1}{p_k} < \gamma$, which restricts the definition of the probabilities, is a necessary condition to ensure $4\phi_2 - 3\phi_1^2 > 0$, and to guarantee that previously defined parameters α and β be real numbers.

There is a probability vector with properties (1) *if and only if* the interval $(\frac{s}{\gamma}, \frac{1}{2ns})$ is not empty. In this case, the probabilities can be obtained using a random generation of $p_k \sim U(\frac{s}{\gamma}, \frac{1}{2ns})$.

The existence of this interval requires some restrictions over the data set in order that inequality $\frac{s}{\gamma} < \frac{1}{2ns}$ be fulfilled. After some algebraic manipulation, we get the following relationship that the data set should satisfy to be used for generating scenarios:

$$n < \frac{n\bar{\kappa}}{\sum_{i,j} L_{ij}^4} - \frac{3\sum_{i=1}^n Z_i^4}{4\sum_{i,j} L_{ij}^4}\left(\frac{n\bar{\xi}}{\sum_{i=1}^n Z_i^3}\right)^2. \tag{4}$$

As in Ponomareva et al. (2015), the choice of Z and L depends only on data. Similar constraints over data appear in Date et al. (2008), but are less restrictive than condition (4).

It is easy to see that constraints over data are not sufficient to guarantee that w_1, w_2, w_0 be positives. If one of these parameters becomes negative, then it could result in ill-defined or negative probabilities for some scenarios. To avoid this situation more restrictive constraints over data should be imposed.

First, to guarantee $w_1 > 0$ and $w_2 > 0$, it is sufficient to ensure $\alpha > 0$ and $\beta > 0$, that we get by putting $\sqrt{4\phi_2 - 3\phi_1^2} > |\phi_1|$, or equivalently $\phi_2 > \phi_1^2$. Second, in order to get $w_0 > 0$, the condition $\phi_2 > \phi_1^2 + 1$ is required. It is easy to see that this last condition is also sufficient to get all the previous conditions; thus, we arrive at a *necessary and sufficient condition* to correctly apply the method for generation of scenarios with well-defined probabilities:

$$\phi_2 - \phi_1^2 > 1 \Leftrightarrow p_{s+1}\left[A - B\sum_{k=1}^s \frac{1}{p_k}\right] > 1, \tag{5}$$

where parameters A and B are defined by:

$$A = \frac{n\bar{\kappa}}{\sum_{i=1}^n Z_i^4} - \left(\frac{n\bar{\xi}}{\sum_{i=1}^n Z_i^3}\right)^2, \qquad B = \frac{\sum_{i,j} L_{ij}^4}{2s^2\sum_{i=1}^n Z_i^4}. \tag{6}$$

3.2. The choice of parameters

Condition (5) is sensitive to the choice of parameters Z, L and probabilities p_k. For choosing parameters we start with Z, then L and finally the probabilities p_k. We present some alternatives to choose these parameters.

3.2.1. The choice of Z

Vector $Z \in \mathbb{R}^n$ should be taken such that $\Sigma - ZZ^\top \succ 0$. Authors in Ponomareva et al. (2015) proposed to take $Z = \rho\sqrt{\text{diag}(\Sigma)}$, where diag($\Sigma$) is the diagonal of the covariance matrix and $\rho \in (0, 1)$. Under the assumption that Σ is a positive definite matrix, this choice works well for some sufficiently small values of ρ.

Another alternative is to choose Z using eigenvalues and eigenvectors of the covariance matrix. Let $0 < \lambda_1 \le \cdots \le \lambda_n$ be the eigenvalues of Σ and $v^1, \ldots, v^n$ be the respective orthonormal eigenvectors, and set $Z = \rho\sqrt{\lambda_l}\,v^l$ with $\rho \in (0, 1)$ and v^l any eigenvector, then:

$$\left(\Sigma - ZZ^\top\right)v^j = \begin{cases} \lambda_j v^j & \text{if } j \neq l \\ \lambda_l(1-\rho^2)v^l & \text{if } j = l \end{cases}. \tag{7}$$

The eigenvectors and eigenvalues of $\Sigma - ZZ^\top$ are the same as the Σ, except for lth eigenvalue, which is $\lambda_l(1-\rho^2)$. This eigenvalue is also positive and therefore $\Sigma - ZZ^\top$ is a positive definite matrix.

3.2.2. The choice of L

The matrix L must be chosen such that $LL^\top = \Sigma - ZZ^\top$. The authors in Ponomareva et al. (2015) propose to use the square root of matrix $\Sigma - ZZ^\top$. Because $\Sigma - ZZ^\top$ is a positive definite matrix, there is a unique symmetric positive definite matrix L that satisfies $LL^\top = \Sigma - ZZ^\top$.

The above election is unique, however there are other nonsymmetric matrices that can be used. For instance, L can be calculated as Cholesky decomposition of $\Sigma - ZZ^\top$, obtaining in this

Please cite this article as: J.P. Contreras et al., Comment on "An algorithm for moment-matching scenario generation with application to financial portfolio optimization", European Journal of Operational Research (2018), https://doi.org/10.1016/j.ejor.2018.02.028