Stochastic scenario generation: An empirical approach

A. D. Oliveira[1], **T. P. Filomena** [1]

[1]Business School, Federal University of Rio Grande do Sul (UFRGS)

alan.delgado@ufrgs.br, tpfilomena@ea.ufrgs.br

Abstract. *We briefly discuss the differences among several methods to generate a scenario tree for stochastic optimization. First, the Monte Carlo Random sampling is presented, followed by the Fitting of the First Two Moments sampling, and lastly the Michaud sampling. Literature results are reviewed, taking into account distinctive features of each kind of methodology. According to the literature results, it is fundamental to consider the problem's unique characteristics to make the more appropriate choice on sampling method.*

1. Introduction

Stochastic optimization (SP) is used to model environments in a more realistic manner when uncertainty is inherent. Many situations in the real world are characterized by random events, which often are viewed as random variables. Random variables are combined into subsets, for that we denote by ξ. So that, we associate a collection of random events with a *probability space*, which contains a collection of all possible events and its probability. From these concepts, we formalize a generic stochastic optimization problem, based on [Dempster et al. 2011], as denoted by equation (1)

$$\max_{x \in X} \int_F f(x, \xi) dF(\xi), \tag{1}$$

where f is the value function defined in terms of both uncertainty and decision spaces, x is the decision variable defined over the feasible set $X \subset \mathbb{R}$, and ξ is a I-dimensional random realization of random variable ξ that is defined by the cumulative distribution function $F : \mathbb{R}^I \to [0, 1]$. Here, dF represents the probability measure on the probability space F of the underlying multivariate stochastic process.

The literature has discussed the way of finding the best approximation of the continuous distribution; efforts concentrating, for example, on having state-space distribution moments matched [Høyland et al. 2003, Høyland and Wallace 2001], minimizing Wassertein probability metrics [Heitsch and Romisch 2005, Romisch 2003], Latin hypercube sampling [McKay et al. 1979], or Michaud sampling [de Oliveira et al. 2016]. The main goal of this research is to design a discrete-space approximation model which approximates a continuous-space stochastic process. Moreover, empirical tests have been made in order to find the most suitable sampling option, [Dempster et al. 2011, Homem-de-Mello and Bayraksan 2014]. According to [Pflug 2001], continuous problems become easier to be solved if we reduce them to discrete-state multiperiod optimization problems. This logic structure, which contains a pertaining history process, may be seen as a tree. As the scenario generation methodology becomes a key part of process of stochastic optimization, we turn our attention to distinct manners of generating scenarios. We consider the Monte Carlo Random sampling detailed in Section 2, the Fitting of

First Two Moments sampling presented at Section 3, and the Michaud sampling discussed in Section 4. Following [de Oliveira et al. 2016], we concentrate our efforts on generating scenarios for Asset Liability Management (ALM).

2. ALM with the Monte Carlo Random sampling

The traditional way to generate a scenario tree for ALM is through Monte Carlo sampling, where uniformly distributed pseudo-random numbers are appropriately transformed into the target distribution, [Dempster et al. 2011]. Hence, we generate $W_1, \ldots, W_I$ random vectors from the standard normal distribution. As noted by [Homem-de-Mello and Bayraksan 2014], in this case, the vectors $W_1, \ldots, W_I$ are mutually independent. It is the detail that characterize this sampling method. This method is more detailed in [Glasserman 2003].

According to [Kouwenberg 2001], though this approach may be quite intuitive, when the states of scenario tree are sampled randomly, the mean and covariance matrix will not be correctly specified in most nodes of the tree. As this information is an input for the optimization model, the optimizer chooses an investment strategy from erratic or miss specified parameters. An obvious way to deal with this problem is to increase the number of nodes in the randomly sample event tree. However, the stochastic program might become computationally intractable due to the exponential growth rate of the tree.

3. ALM with the Fitting the First Two Moments sampling

In this approach, proposed by [Høyland and Wallace 2001], we construct an event tree that fits the mean and the covariance matrix of the underlying distribution. Besides that, [Høyland et al. 2003] provides a source code to this method. The first step is to generate the random vectors from the standard normal distribution, in the same way as described in Section 2. After that, with the I random vectors, we transform them to exhibit a given correlation by pre-multiplying the vectors with lower triangular matrix L of covariance matrix Σ ,

$$W_j^\top = LW_j, \Sigma = LL^\top, j = 1, \ldots, I, \quad (2)$$

where L can be obtained by applying Cholesky decomposition. In another words, as mentioned by [Kouwenberg 2001], we specify that the average of the disturbances should be zero, and they should have a covariance matrix equal to Σ. Therefore, we may denote this matching as (3) and (4),

$$\frac{1}{S}\sum_{s=1}^{S} W_{js} = 0 \quad \forall j \in 1, ..., I, \quad (3)$$

$$\frac{1}{S-1}\sum_{s=1}^{S} W_{js}W_{is} = \Sigma_{ij} \quad \forall j, i \in 1, ..., I, \quad (4)$$

As mentioned by [Kouwenberg 2001, Dempster et al. 2011, Löhndorf 2016], it is possible to argue that this sampling outperform other methods as Monte Carlo Random sampling, Wasserstein Distance sampling, or even Latin Hypercube sampling.

4. ALM with the Michaud Sampling

Another approach is to simulate the generation of scenarios for multiple trees, and to take the solution of an optimization problem of smaller size for each tree. This technique has some similarities with the resampled efficient frontier method proposed by [Michaud and Michaud 2008] for the construction of portfolio of risky securities and it is applied by [de Oliveira et al. 2016]. This sampling has four steps. First, we define the number of trees to be solved for each parametrization, see, e.g., [Michaud and Michaud 2008]. The value of this realizations must be sufficiently large so that the portfolio allocations are stable and small enough, ensuring that the approach does not become computationally prohibitive. Once the number of instances is defined, the second step is to generate the scenarios for each tree in accordance with its model. In Step 3, we solve to optimality the optimization problem corresponding to each tree. In Step 4, we evaluate the results based on the optimal solutions of the trees.

The solutions provided by the resampling avoid corner allocations, i.e. extreme weight allocations, which may be seen in the classical mean-variance portfolio selection model. Furthermore, it decrease the estimation bias, [Fletcher and Hillier 2001]. However, [Scherer 2002] draws attention to when long and short positions are allowed. In this case, the resampled efficient frontier does not present notorious differences from the classical efficient frontier.

5. Final Considerations

In the previous sections, we addressed some literature results to identify and assess strengths and weaknesses of different scenario generation tree methods for stochastic programming. The main objective of these sampling methods is to consider many scenarios while preserving computational tractability. We noticed that main drawback of Monte Carlo Random sampling is the need to assign a large number of scenarios to achieve better results. If we consider, for example, a model with I independent random variables, each with only two possible alternatives; the total number of scenarios is thus 2^I for only one stage, and so even for moderate values of I it becomes impractical to take all possible outcomes into account. In case of the Fitting of the First Two Moments, although it has achieved good performance [Kouwenberg 2001, Löhndorf 2016], [Dempster et al. 2011] outline some weakness when it is applied to the risk control problems given underestimation of the problem's volatility. If we consider the Michaud sampling, the literature have not yet found a consensus on the pros and cons of the results delivered by resampling [Markowitz and Usmen 2003, Ulf and Raimond 2006, Becker et al. 2015]. Furthermore, this method still lacks a theoretical foundation. Therefore, as we can see, due to vulnerabilities of each sampling method, the choice of the appropriate sampling method depends on the problem to be modeled and its specific characteristics.

Referências

Becker, F., Gurtler, M., and Hibbeln, M. (2015). Markowitz versus Michaud: portfolio optimization strategies reconsidered. *The European Journal of Finance*, 21(4):269–291.

de Oliveira, A., Filomena, T., Perlin, M., Lejeune, M., and Macedo, G. (2016). A multistage stochastic programming asset-liability management model: an application to the brazilian pension fund industry. *Optimization and Engineering*, forthcoming.

Dempster, M., Medova, E., and Yong, Y. (2011). Comparison of sampling methods for dynamic stochastic programming. In *Stochastic Optimization Methods in Finance and Energy*, chapter 16, pages 389–425. Springer Science+Business Media, New York.

Fletcher, J. and Hillier, J. (2001). An examination of resampled portfolio efficiency. *Financial Analysts Journal*, 57(5):66–74.

Glasserman, P. (2003). *Monte Carlo Methods in Financial Engineering*, volume 53. Springer-Verlag New York, New York, 1 edition.

Heitsch, H. and Romisch, W. (2005). Generation of multivariate scenario trees to model stochasticity in power management. In *IEEE St. Petersburg Power Tech*.

Homem-de-Mello, T. and Bayraksan, G. (2014). Monte carlo sampling-based methods for stochastic optimization. *Surveys in Operations Research and Management Science*, 19(1):56–85.

Høyland, K., Kaut, M., and Wallace, S. (2003). A heuristic for moment-matching scenario generation. *Computational Optimization and Applications*, 24(2):169–185.

Høyland, K. and Wallace, S. (2001). Generating scenario trees for multistage decision problems. *Management Science*, 47(2):295–307.

Kouwenberg, R. (2001). Scenario generation and stochastic programming models for asset liability management. *European Journal of Operational Research*, 134(2):279 – 292.

Löhndorf, N. (2016). An empirical analysis of scenario generation methods for stochastic optimization. WU Vienna University of Economics and Business.

Markowitz, H. and Usmen, N. (2003). Resampled frontiers versus diffuse bayes: An experiment. *Journal of Investment Management*, 1(4):9–25.

McKay, M. D., Beckman, R. J., and Conover, W. J. (1979). A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21(2):239–245.

Michaud, R. and Michaud, R. (2008). *Efficient Asset Management: A Practical Guide to Stock Portfolio Optimization and Asset Allocation*. Oxford University Press, 2 edition.

Pflug, G. (2001). Scenario tree generation for multiperiod financial optimization by optimal discretization. *Mathematical Programming*, 89(2):251–271.

Romisch, W. (2003). Stability of stochastic programming problems. In Ruszczynski, A. and Shapiro, A., editors, *Stochastic Programming, Volume 10 of Handbooks in Operations Research and Management Science*, chapter 8, pages 483–554. Elsevier, Amsterdam.

Scherer, B. (2002). Portfolio resampling: Review and critique. *Financial Analysts Journal*, 58(6):98–109.

Ulf, H. and Raimond, M. (2006). Portfolio choice and estimation risk. a comparison of bayesian to heuristic approaches. *Astin Bulletin*, 36(1):135–160.