European Journal of Operational Research 255 (2016) 121–132

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Stochastics and Statistics

An empirical analysis of scenario generation methods for stochastic optimization

Nils Löhndorf*

Department of Information Systems and Operations, WU Vienna University of Economics and Business, Welthandelsplatz 1, Vienna 1020, Austria

ARTICLE INFO

Article history:
Received 10 February 2016
Accepted 10 May 2016
Available online 17 May 2016

Keywords:
Stochastic optimization
Sample average approximation
Scenario generation
Probability metrics
Monte Carlo methods

ABSTRACT

This work presents an empirical analysis of popular scenario generation methods for stochastic optimization, including quasi-Monte Carlo, moment matching, and methods based on probability metrics, as well as a new method referred to as *Voronoi cell sampling*. Solution quality is assessed by measuring the error that arises from using scenarios to solve a multi-dimensional newsvendor problem, for which analytical solutions are available. In addition to the expected value, the work also studies scenario quality when minimizing the expected shortfall using the conditional value-at-risk. To quickly solve problems with millions of random parameters, a reformulation of the risk-averse newsvendor problem is proposed which can be solved via Benders decomposition. The empirical analysis identifies *Voronoi cell sampling* as the method that provides the lowest errors, with particularly good results for heavy-tailed distributions. A controversial finding concerns evidence for the ineffectiveness of widely used methods based on minimizing probability metrics under high-dimensional randomness.

1. Introduction

A wide range of real-world problems that occur in finance, industrial engineering, operations, or marketing involve decision-making under uncertainty. If a statistical model can be used to describe this uncertainty, the decision problem can be modeled as a stochastic optimization problem. The most widely used technique to solve real-world stochastic optimization problems is sample average approximation, where the probability distribution is approximated by a set of discrete scenarios (Birge & Louveaux, 1997; Shapiro, 2003). Since the complexity of an optimization problem scales with the number of scenarios, a lot of work on stochastic optimization is dedicated to techniques that reduce the number of scenarios while retaining the quality of the stochastic solution. The goal of techniques for scenarios generation is hereby to select a set of scenarios which minimizes the approximation error.

Methods for generating a reduced set of scenarios from univariate distributions are well established and typically referred to as variance reduction techniques (Higle, 1998; Shapiro, 2003). Adapting these techniques for the more general multivariate case, however, is not straightforward. Let us briefly summarize what can be considered as state-of-the art to generate scenarios of multivariate random variables for sample average approximation: quasi-Monte Carlo methods, methods based on probability metrics, and moment matching.

A popular class of methods for Monte Carlo sampling in higher dimensions are quasi-Monte Carlo methods which have their roots in number theory (Niederreiter, 1992). Quasi-Monte Carlo methods rely on so called low-discrepancy sequences which produce a sequence of vectors that cover the unit hypercube as uniformly as possible, e.g., Sobol sequences. If combined with an adequate transformation, these sequences can be treated just like pseudo-random numbers and as such can be used to sample random number sequences from a large number of multivariate distributions. See Glasserman (2004) for an overview.

The idea behind scenario reduction using probability metrics is to compute the closest approximation of the probability distribution by a discrete distribution with smaller support. A probability metric serves as the objective criterion and can be related to the error from implementing the optimal solution of a stochastic optimization problem using sample average approximation (Heitsch & Römisch, 2003, 2007; Pennanen & Koivu, 2005; Pflug, 2001). A greedy algorithm that reduces a time series sample to a scenario tree is proposed in Heitsch and Römisch (2003).

If the scenario tree has only two stages, the problem of finding the closest discrete approximation, can be modeled as optimal quantization problem. Quantization has its roots in signal

* Tel.: +43 31336 5629.
E-mail address: nils.loehndorf@wu.ac.at

http://dx.doi.org/10.1016/j.ejor.2016.05.021
0377-2217/

processing, where a finite number of quantizers is needed to transmit a stationary signal (Gersho & Gray, 1992). Quantizers are real-valued vectors comparable to what a decision-maker would refer to as a scenario if the signal was a sample of prices or demands. An optimal quantization of the signal sets the location of the quantizers such that the average distortion is minimized, which is equivalent to finding the closest discrete approximation. Optimal quantization has reemerged as a method for numerical integration, mostly driven by the work of Pagès (1998) and Graf and Luschgy (2000). See Pagès and Printems (2008) for a review of optimal quantization applied in finance.

Another successful method to generate scenarios for stochastic optimization is moment matching. The original idea is due to Fleishman (1978) for the univariate which has been adapted by Høyland, Kaut, and Wallace (2003) who propose a heuristic algorithm for the multivariate case. Mehrotra and Papp (2013) use optimization to find a set of moment matching scenarios. In contrast to optimal quantization, where a random sample from the probability distribution is drawn to generate scenarios, moment matching tries to define another discrete distribution that exhibits the same first four moments as the original distribution.

In the extant literature, only few studies compare the performance of different scenario generation techniques across multiple different optimization problems, with respect to solution quality of sample average approximation. Studies with a focus on Monte Carlo methods are Koivu (2005), Linderoth, Shapiro, and Wright (2006), Freimer, Linderoth, and Thomas (2012), and Homem-de Mello, de Matos, and Finardi (2011). A notable exception are Dempster, Medova, and Yong (2011) who compare the performance of Monte Carlo methods, moment matching, and methods based on probability metrics on a specific asset liability management problem.

The goal of this research is study the error that arises from using a small set of scenarios to approximate a continuous multivariate distribution with the objective to numerically solve a multidimensional newsvendor problem, for which analytical solutions are still available. The computational study compares the current state-of-the-art, including quasi-Monte Carlo, moment matching, as well as optimal quantization, and considers a range of distributions including normal, uniform, log-normal, as well as a heavy-tailed t-distribution.

Discovered as a by-product of initial experimental work, the work furthermore introduces *Voronoi cell sampling* as an alternative method for scenario generation. Voronoi cell sampling integrates stratified sampling with probability metrics in a simple yet effective way and performs favorably compared to the other methods.

Since the objective function of the profit maximizing newsvendor is separable, minimization of the expected shortfall is also considered, which is modeled using the conditional value-at-risk formulation (CVaR) of Rockafellar and Uryasev (2002). To measure the error of approximating the continuous distribution by a small set of scenarios, a tailored solution approach based on Benders decomposition is developed. The method is capable of approximating a solution to the CVaR newsvendor problem with several million random parameters which then serves as proxy of the true solution.

The paper is organized as follows. Section 2 gives a brief introduction to stochastic optimization in general, describes existing approaches to generate scenarios for stochastic optimization, and introduces Voronoi cell sampling as a new scenario reduction method. Section 3 presents the newsvendor models that are used to evaluate the sample average approximation error. Section 4 outlines the experimental design and discusses the results. Section 5 concludes with a summary and gives implications for future work.

2. Methodological background

2.1. Sample average approximation

Denote x as a decision variable defined over the feasible set $X \subseteq \mathbb{R}$, z as a d-dimensional random realization of random variable Z that is defined by the distribution function $F : \mathbb{R}^d \to [0, 1]$, and $c(x, z)$ as the cost function. The stochastic optimization problem is to minimize

$$C(x) = \int c(x, z) dF(z) \tag{1}$$

by choosing an optimal decision

$$x^* \in \text{argmin}_{x \in X} C(x). \tag{2}$$

With the exception of very simple well-behaved problems, solutions are typically calculated numerically by either drawing a sample from F or by approximating F by a discrete distribution $\hat{F}$. Using $\hat{F}$ in place of F yields the problem of minimizing

$$\hat{C}(x) = \sum_{\hat{z} \in \hat{F}} \hat{p}(z) c(x, \hat{z}), \tag{3}$$

where $\hat{p}(z)$ is the probability of the mass points of $\hat{F}$.

This approach is referred to by different names in different communities, e.g., numerical integration (Judd, 1998), Monte Carlo methods (Glasserman, 2004), or sample average approximation (Shapiro, 2013). Since the optimization literature mostly refers to this method by the latter name, the term sample average approximation (SAA) will be used from hereon.

2.2. Monte Carlo and quasi-Monte Carlo

A common approach to solve a stochastic optimization problem using SAA is based on Monte Carlo sampling. To use Monte Carlo sampling for SAA, we generate a set of M uniformly distributed, pseudo-random realizations $u_1, \ldots, u_M$ with $u_i \in [0, 1]^d$, and then construct a sample from F by an appropriate transformation $U \to Z$ (Glasserman, 2004).

If we view $z_1, \ldots, z_M$ as a sample of random realizations from the same distribution as F, the SAA is given by

$$\hat{C}(x) = \frac{1}{M} \sum_{i=1}^{M} c(x, z_i), \tag{4}$$

with the variance of the estimate given by

$$\hat{\sigma}^2_{MC} = \frac{1}{M} \mathbb{V}ar[c(x, Z)]. \tag{5}$$

Although the Monte Carlo estimate $\hat{C}_{MC}(x)$ converges to $C(x)$ with probability one as $M \to \infty$, it is desirable to reduce the variance of the estimate more quickly to speed up convergence of the error bounds (Shapiro, 2003). In this respect, a number of variance reduction techniques have been proposed for Monte Carlo sampling. See Glasserman (2004) or Shapiro (2003) for an overview.

A popular approach to reduce the variance for small sample sizes is to use so-called low-discrepancy sequences that cover the unit hypercube as uniformly as possible (Niederreiter, 1992). It can be shown that using such low-discrepancy sequences instead of pseudo-random numbers speeds up the rate of convergence of the Monte Carlo estimate. Since these sequences are also referred to as quasi-random numbers, the sampling method is typically referred to as quasi-Monte Carlo. See Koivu (2005) for an analysis of using quasi-Monte Carlo sampling for SAA.

2.3. Probability metrics

Another approach to reduce the number of scenarios is to address the approximation error directly. Let us therefore define the