Int. J. Production Economics 182 (2016) 230–252

Contents lists available at ScienceDirect

Int. J. Production Economics

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

An enhanced sample average approximation method for stochastic optimization

Adindu Emelogu [a], Sudipta Chowdhury [a], Mohammad Marufuzzaman [a,*], Linkan Bian [a], Burak Eksioglu [b]

[a] Department of Industrial and Systems Engineering Mississippi State University, Starkville, MS 39759-9542, USA
[b] Department of Industrial Engineering Clemson University, Clemson, SC 29634-0920, USA

ARTICLE INFO

Article history:
Received 1 April 2016
Received in revised form
17 August 2016
Accepted 31 August 2016
Available online 1 September 2016

Keywords:
Sample average approximation
K-means
K-means++
K-means||
Fuzzy C-means
Facility location problem
Scenario clustering

ABSTRACT

Choosing the appropriate sample size in Sample Average Approximation (SAA) method is very challenging. Inappropriate sample size can lead to the generation of low quality solutions with high computational burden. To overcome this challenge, our study proposes an enhanced SAA algorithm that utilizes clustering techniques to dynamically update the sample sizes and offers high quality solutions in a reasonable amount of time. We evaluate this proposed algorithm in the context of a facility location problem **[FLP]**. A number of numerical experiments (e.g., impact of different clustering techniques, fixed vs. dynamic clusters) are performed for various problem instances to illustrate the effectiveness of the proposed method. Results indicate that on average, enhanced SAA with fixed clustering size and dynamic clustering size solves [**FLP**] almost 631% and 699% faster than the basic SAA algorithm, respectively. Furthermore, it is observed that there is no single winner among the clustering techniques to solve all the problem instances of enhanced SAA algorithm and the performance is highly impacted by the size of the problems.

1. Introduction

Solving large scale stochastic optimization problem is extremely challenging because of their inherent analytical complexities and high computational requirements (Kleywegt et al., 2002 and Shastri and Diwekar, 2006). Sample Average Approximation (SAA) is a popular approach which is frequently employed to solve large scale stochastic optimization problems. In this method, the objective function value of the stochastic problem is unknown and approximated using a sample average estimate derived from a random sample (Ahmed and Shapiro, 2002; Blomvall and Shapiro, 2006; Chepuri and Homem-de-mello, 2005; Homem-de-Mello and Bayraksan, 2014; Royset, 2013 and Alex Shapiro and Homem-de-Mello, 2000). SAA provides a straightforward framework which is amenable to parallel implementation and variance reduction techniques. Moreover, it possesses good convergence properties and well-developed statistical methods for validating solutions and conducting error analysis.

SAA has been successfully utilized to serve a wide range of applications, some of which include: reliability-based optimal design of engineering systems where the failure probabilities of highway bridges are replaced by corresponding Monte Carlo sampling estimates (Royset and Polak, 2004); speech recognition optimization problem (Byrd et al., 2012); investment problem with conditional value at risk (CVaR) constraints (Branda, 2014); portfolio selection and blending problems with chance-constraints (Wang and Ahmed, 2008); stochastic knapsack problem to determine an optimum resource allocation strategy (Kleywegt et al., 2002); stochastic supply chain design problems with extremely large number of scenarios (Santoso et al., 2005), and many others. The most significant challenge of using SAA confronted by the researchers in prior works is to choose the sample size for the algorithm. This is a critical step since it highly impacts the computational performance of the SAA algorithm. To address this challenge, a number of studies are conducted to determine the best scheme for choosing the sample size of the SAA algorithm. One stream of research focuses on keeping the sample size constant throughout the optimization process (e.g., Santoso et al., 2005; Verweij et al., 2003 and Nemirovski et al., 2009). The major drawback of this approach is that it may lead to a bad sample path (Homem-De-Mello, 2003). Another stream of research focuses on variable sample approach in which a schedule of sample sizes is used to solve the SAA problem (e.g., Royset, 2013; Byrd et al., 2012; Homem-De-Mello, 2003; Balaprakash et al., 2009 and Deng and Ferris, 2009). The general idea employed by the authors in variable sample scheme is to start the early iteration of the optimization

* Corresponding author.
E-mail addresses: aae39@msstate.edu (A. Emelogu), sc2603@msstate.edu (S. Chowdhury), maruf@ise.msstate.edu (M. Marufuzzaman), bian@ise.msstate.edu (L. Bian), burak@clemson.edu (B. Eksioglu).

http://dx.doi.org/10.1016/j.ijpe.2016.08.032
0925-5273/

algorithm with a small sample size and then gradually increase the sample size as the algorithm progresses. Note that starting with a small sample size may save some computational time; however, a large sample size eventually needs to be investigated to obtain a solution that is close to the true solution (Morton, 1998). Therefore, all the methods discussed above may not perform well to solve stochastic discrete optimization problems. Although there is a theoretical sample size that can be used to compute sample sizes for discrete optimization problems (as shown in Kleywegt et al., 2002 and Homem-De-Mello, 2003), this estimate is too conservative for practical applications.

To address this challenge, this paper proposes a methodological approach to enhance the performance of the basic SAA by incorporating a dynamic clustering strategy within the algorithmic framework. In basic SAA, a small number of scenarios are generated in each iteration and the objective function is evaluated iteratively until the optimality gap falls below a certain threshold value. In our approach, a larger number of scenarios are considered as an initial sample size (much larger than the one used in basic SAA) and then clustering methods are employed to reduce this large sample size into small number of clusters. We assume that the average of each cluster is the most representative of all the samples in each cluster. We then represent those clusters as scenarios and use them to solve the SAA problem. Unlike prior studies where the sample size is either kept fixed or increased monotonically, our enhanced SAA approach provides the flexibility to either increase or decrease the sample size based on the computational performance obtained from previous iterations. This approach is then experimentally validated in the context of a facility location problem **[FLP]** with stochastic demand. We create different variants of the enhanced SAA algorithm (i.e., different clustering strategy, fixed clusters vs. dynamic clusters) and compare the computational performance of those variants with the basic SAA algorithm. Finally, we employ five different clustering techniques (e.g., K-means, K-means++, K-means||, Fuzzy C-means, and Mixed Integer Programming (MIP) based clustering techniques) and check how these clustering techniques affect the solution quality of the SAA algorithm.

The remainder of this paper is organized as follows. Section 2 provides the literature review on SAA. Section 3 introduces the enhanced SAA algorithm. Section 4 conducts numerical experiments to verify the performance of the enhanced SAA algorithm. Section 5 concludes this paper and discusses future research directions.

2. Literature review

We review the existing literatures related to SAA and categorize them into two major groups: (1) SAA with fixed sample size and (2) SAA with variable sample size.

2.1. SAA with fixed sample size

The first set of literature considers the basic SAA where the sample size remain fixed in all iterations. This approach is also referred to as *sample-path approximation* method (Gürkan et al., 1994 and Plambeck et al., 1996) or *stochastic counterpart* method (Rubinstein and Shapiro, 1990; Rubinstein and Shapiro, 1993 and Alexander Shapiro and Homem-de-Mello, 1998). Linderoth et al. (2006) exploit the parallel implementation capability of the SAA to solve various two-stage stochastic linear programming problems with recourse. Kenyon and Morton (2003) embed branch-and-cut inside the SAA to solve a stochastic vehicle routing problem under random travel and service times. Morton (1998) develops an SAA procedure to solve a stochastic knapsack problem (SKP). Schütz et al. (2009) embed dual decomposition inside the SAA to solve a meat packing supply chain network designing problem. The authors investigate the effect of sample size on solution quality and find that increasing the sample size improves the solution quality of the SAA algorithm. Wang and Ahmed (2008) use SAA to solve a conditional value-at-risk (CVaR) problem and find that the SAA solution is acceptable to the true CVaR problem with a probability of at least 97.7%.

Some other studies involving SAA implementations with fixed sample sizes are conducted by Kleywegt et al. (2002), Verweij et al. (2003), Santoso et al. (2005) and Nemirovski et al. (2009). The major concurring theme among the studies in fixed sample SAA literature is that estimating the sample size in practice is not trivial and selecting the sample size involves two conflicting trade-offs: (i) larger sample sizes yield SAA solution comparable to the true solution and (ii) the computation effort required to solve the SAA problem often increases exponentially as the sample size increases. Table 1 provides a summary of literature based on SAA with fixed sample size.

2.2. SAA with variable sample size

Later studies use variable sample size in SAA to solve stochastic optimization problems. Homem-De-Mello (2003) first provides a variable-sample framework to solve a discrete stochastic optimization problem. The author shows that the sample size must grow at a certain rate to ensure convergence. Royset (2013) proposes a closed-loop feedback optimal-control model to adaptively select sample sizes in variable sample average approximation (VSAA) algorithm to solve smooth stochastic programs (SSP). Although the method results in a sample size selection policy that appears to be robust to changing problem instances, it is not applicable to stochastic optimization problems with integer restrictions. Optimization problems that require integer solutions in their decision variables involve non-smooth functions which are not easily convertible to smooth functions for the application of the SAA method. Pasupathy (2010) determines a balance choice of sample sizes and error tolerances in variable samples method of SAA where sample size refers to a measure of problem-generation effort and error tolerance is a measure of solution quality. Note that all the literature discussed above investigated problems in which SAA occurs only in the objective function.

Another stream of research investigates problems in which SAA occurs in the constraints (e.g., Branda, 2014 and Zhang et al., 2012). It is important to note that if the approximation occurs in the objective function then the major challenge relies on obtaining solutions that converge to the original problem. However, in cases where the approximation occurs in the constraints, i.e., in chance-constraints, one needs to ensure that the feasibility region of the approximating problem coincides with that of the original problem (Branda, 2014). Branda (2014) estimates the rate of convergence and sample size for lower bounds in SAA which ensures that the feasible solutions of the SAA are feasible for the original problem. Zhang et al. (2012) develop a method for stochastic programs with complementary constraints where the equilibrium constraints can be replaced with smooth functions. Bastin et al. (2006) implement variable sample size technique to estimate choice probabilities in solving unconstrained mixed logit models. Byrd et al. (2012) develop a varying sample size based methodology to solve large scale machine learning problems. Similarly, a number of other related literatures such as Deng and Ferris (2009), Krejić and Krklec (2013), Krejic and Jerinkic (2014), Bastin et al. (2006), Byrd et al. (2012) and Bastin (2004) study SAA with variable sample size to tackle different optimization problems with or without constraints. The literature for SAA with variable sample sizes is summarized in Table 2. Note that all the methods discussed