An Adequate Choice of Initial Sample Size for Selection Approach

Mohammad H. Almomani, and Rosmanjawati Abdul Rahman

Abstract—In this paper, we consider the effect of the initial sample size on the performance of a sequential approach that used in selecting a good enough simulated system, when the number of alternatives is very large. We implement a sequential approach on $M/M/1$ queuing system under some parameter settings, with a different choice of the initial sample sizes to explore the impacts on the performance of this approach. The results show that the choice of the initial sample size does affect the performance of our selection approach.

Keywords—Ranking and Selection, Ordinal Optimization, Optimal Computing Budget Allocation, Subset Selection, Indifference-Zone, Initial Sample Size.

I. INTRODUCTION

We consider optimizing the expected performance of a complex stochastic system that cannot be evaluated exactly, but has to be estimated using simulation. Our goal is to solve the following optimization problem

$$\min_{\theta \in \Theta} J(\theta)$$  \hspace{1cm} (1)

where the feasible solution set $\Theta$ is a finite, huge and has no structure. Meanwhile, $J$ is the expected performance measure, $L$ is a deterministic function depends on $\theta$ and $\xi$, and we can write $J(\theta) = E[L(\theta, \xi)]$. $\theta$ is a vector that representing the system design parameters, and $\xi$ represents all the random effect of the system. If we simulate the system to get estimate of $E[L(\theta, \xi)]$, then the confidence interval of this estimate cannot be improved faster than $1/\sqrt{k}$ where $k$ is the number of samples used to get estimates of $J(\theta)$. This rate maybe good for some problems with a small number of alternatives but it is not good enough for the class of complex simulation which we consider in this paper. Thus, one could compromise the objective to get a good enough solution rather than doing extensive simulation.

Ranking and Selection (R&S) procedures, are used to select the best system or a subset that contain the best systems when the number of alternatives is small, see Kim and Nelson [1]. The problem arise for a large scale problems since it needs a huge computational time. In this situation, we would compromise our objective to finding good systems rather than estimating accurately the performance value for these systems. The idea lies in Ordinal Optimization (OO) procedure, that proposed by Ho et al. [2].

In many selection procedures, sample size in the first stage $n_0$ play an important role to the performance of these procedures. In fact, the initial sample size $n_0$ cannot be too small since we might get a poor estimates for the sample mean and variances. On the other hand, $n_0$ cannot be too large, because in the first stage there exist many noncritical system and by giving a large number of sample will result in losing a large number of samples and also wasting computation time. However, Chen et al. [3] and Chen et al. [4] suggested that a $n_0$ should be between 10 and 20 as a good choice for the initial sample size. Unfortunately, there is no clear formula to calculate an appropriate value of the initial sample size $n_0$ for the selection procedures, when the number of alternative is large.

In this paper, we study the effects of the initial sample size $n_0$ on the performance of one of the selection procedures; a sequential approach Almomani and Abdul Rahman [5]. We also consider a heuristic approach to selecting a good simulated system with high probability when the number of alternative system is huge. This approach consists of four stages; in the first stage we use the OO procedure to select randomly a subset that overlaps with the set of the actual best $m\%$ systems with high probability from the feasible solution set $\Theta$. In the second stage, we use Optimal Computing Budget Allocation (OCBA) technique to allocate the available computing budget in a way that maximizes the probability of correct selection. This will follow by a Subset Selection (SS) procedure to get a smaller subset that contains the best system from the subset that is selected before. In the final stage, we use the Indifference-Zone (IZ) procedure to select the best system among the survivors in the previous stage. This approach are applied to $M/M/1$ queuing system with a different choice of the initial sample size $n_0$ to know the effect of the $n_0$ on the selection approach performance.

This paper is organized as follows; In the next section we give a background about $OO$, $OCBA$, $SS$, and $IZ$ procedures. In Section 3, we present our sequential approach. Section 4, includes the $M/M/1$ queuing system example. Finally, in Section 5, we give some concluding remarks.

II. BACKGROUND

A. Ordinal Optimization (OO)

The $OO$ procedure has emerged as an efficient technique for simulation and optimization. The aim of this procedure is to find good systems, rather than estimating the performance value of these systems accurately. The $OO$ procedure has been proposed by Ho et al. [2].

Suppose that the Correct Selection (CS) is to select a subset $G$ of $g$ systems from the feasible solution set $\Theta$ that