Asymptotic Approximation of the Probability of Correctly Selecting the Best Systems

Mahmoud H. Alrefaei[1,a)], Mohammad H. Almomani[2,b)] and Sarah N. Alabed Alhadi[1,c)]

[1]*Jordan University of Science and Technology, Department of Mathematics and Statistics, Irbid, Jordan*
[2]*The Hashemite University, Department of Mathematics, Zarqa, Jordan*

[a)]Corresponding author: alrefaei@just.edu.jo
[b)]mhmomani@hu.edu.jo
[c)]snalabdelhadi13@sci.just.edu.jo

Abstract. Consider the problem of selecting the best stochastic system or the best m systems among a finite but large alternative systems. If a limited computational budget is available to be distributed among the different alternatives, then instead of distributing these computations evenly, the optimal computing budget allocation (OCBA) can be used to distribute this budget in a smart way so as to maximize the probability of correct selection (PCS). However, the OCBA does not tell how large is the PCS. In this paper, we present a procedure that resembles the OCBA, but it gives an approximation of PCS. Thus the user can stop the simulation whenever a precision level is reached.

INTRODUCTION

Consider the problem of locating the best m systems among a very large number of alternative simulated systems. This problem appears in many aspects of real life such as scheduling systems, manufacturing systems, telecommunication systems, etc. The Ordinal Optimization (OO) was proposed to relax the problem to select a good enough solution instead of simulating each alternative accurately. The optimal computing budget allocation (OCBA) was then proposed to enhance the OO. OCBA assumes that there are available computational budget, and the question is how to distribute them among the competing alternative systems in order to maximize the probability of correct selection (P(CS)) instead of distributing them evenly. However, in all the previous literature, the OCBA do not approximate the probability of correct selection. In this paper, we discuss how to approximate P(CS) based on the OCBA distribution, the user then can chose whether to continue the simulation until a certain level of precision is achieved or not.

Each sample of the performance value requires one simulation run, therefore for large scale problems, large number of samples are needed which is very time consuming and maybe impossible. In this situation, one would change the objective to find a good enough solution rather than estimating accurately the performance value of each system.

Suppose we have an available computing budget (computing time) to be distributed among the different alternative systems in order to maximize the probability of correctly selecting the best m systems. To achieve this goal [1] proposed the Optimal Computing Budget Allocation ($OCBA$) procedure that gives a large number of simulation samples to the systems that have more impact on identifying the best systems, whereas it gives a limited simulation sample for those systems that have little impact on the final solution, see [2],[3], [4], and [5].

There exists a large literature on assessing $P(CS)$ based on classical statistical models, [6] and [4]). However, most of these approaches are only suitable for problems with a small number of systems. [5] introduced an estimation technique that approximates $P(CS)$ for ordinal comparison when the number of systems is large based on a Bayesian model. more details about MOCBA can be found in [7], [8], [9], [10].

The main objective of this paper is to provide an approximation of the probability of correctly selecting the best m systems among k different systems subject to a constraint on the total available budget.

1st International Conference on Mathematical and Related Sciences (ICMRS 2018)
AIP Conf. Proc. 1991, 020022-1–020022-5; https://doi.org/10.1063/1.5047895
Published by AIP Publishing. 978-0-7354-1707-6/$30.00

Optimal Computing Budget Allocation for Selecting the Best m

Consider the problem of selecting a set S_m that contains the best m systems among k systems. We will discuss how to allocate the available budget among the alternative systems in order to improve the probability of correct selection.

The problem can be formulated as maximizing the probability of correctly selecting the best m systems; $P(CS_m)$ subject to a constraint on the total number of available samples, see [11], [12] . In mathematical notation, the problem is formulated as follows:

$$\max_{N_1,\dots,N_k} \; P(CS_m)$$
$$\text{s.t.} \; \sum_{i=1}^{k} N_i = T \tag{1}$$

where T is the total number of simulation replications (budget), k is the total number of systems and N_i is the number of simulation replications allocated to system i. Assume that Y_{ij} is the j^{th} simulation sample of system i for estimating the mean Y_i , where $j = 1, \dots, N_i$ and $\overline{Y_i} = 1/N_i \sum_{j=1}^{N_i} Y_{ij}$ is the estimated mean. If we use multiple replications method or batch means in simulation, then by the Central Limit Theorem, $\bar{Y}_i$ is normally distributed with mean Y_i and variance σ_i^2/N_i. In practice σ_i^2 is unknown, so we estimate it using the sample variances s_i^2 for Y_{ij}.

Suppose that we are interested in selecting a set S_m that contains the m systems with the smallest means. Let $\bar{Y}_{i_r}$ be the r-th smallest (order statistic) of $\{\bar{Y}_1, \bar{Y}_2, \dots, \bar{Y}_n\}$, i.e. $\bar{Y}_{i_1} \leq \bar{Y}_{i_2} \leq \dots \leq \bar{Y}_{i_n}$. Then, the selected subset is given by $S_m = \{i_1, i_2, \dots, i_m\}$. The correct selection is that S_m contains the actual m smallest means, i.e. $CS_m = \{\max_{i \in S_m} Y_i \leq \min_{i \notin S_m} Y_i\}$, where Y_i is the mean of system i.

Let $\hat{Y}_i$ be independent samples of $\bar{Y}_i$ obtained by multiple replications method of simulation, then the probability of correct selection can be approximated by [11]

$$\begin{aligned}
P(CS_m) &= P\{\hat{Y}_i \leq \hat{Y}_j\}, \forall\, i \in S_m,\; j \notin S_m \\
&\geq P\{\hat{Y}_i \leq c \text{ and } \hat{Y}_j \geq c\}, i \in S_m,\; j \notin S_m \\
&= \prod_{i \in S_m} P\{\hat{Y}_i \leq c\} \prod_{i \notin S_m} P\{\hat{Y}_i \geq c\} = APCS_m
\end{aligned}$$

where c is a constant between Y_m and Y_{m+1}.

Let $\alpha_i = (Y_i - c)/\sigma_i$ then for $i \in S_m$, $\alpha_i < 0$. Since $\hat{Y}_i$ is normally distributed with mean Y_i and variance σ_i^2/N_i, for large N_i, by ([13] (Chapter 5 Section 3)) we get:

$$\begin{aligned}
P(\hat{Y}_i \leq c) &= P\left(\frac{\hat{Y}_i - Y_i}{\sigma_i/\sqrt{N_i}} \leq \frac{c - Y_i}{\sigma_i/\sqrt{N_i}}\right) \\
&= 1 - \left(1 - P\left(\frac{\hat{Y}_i - Y_i}{\sigma_i/\sqrt{N_i}} \leq -\alpha_i\sqrt{N_i}\right)\right) \\
&\geq 1 + \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{N_i}\alpha_i} \exp\left(\frac{-\alpha_i^2 N_i}{2}\right)
\end{aligned}$$

Similarly for $i \notin S_m$, $\alpha_i > 0$, therefore

$$\begin{aligned}
P(\hat{Y}_i \geq c) &= P\left(\frac{\hat{Y}_i - Y_i}{\sigma_i/\sqrt{N_i}} \geq \frac{c - Y_i}{\sigma_i/\sqrt{N_i}}\right) \\
&= P\left(\frac{\hat{Y}_i - Yi}{\sigma_i/\sqrt{N_i}} \leq \frac{Y_i - c}{\sigma_i/\sqrt{N_i}}\right) \\
&= 1 - \left(1 - P\left(\frac{\hat{Y}_i - Yi}{\sigma_i/\sqrt{N_i}} \leq \frac{Y_i - c}{\sigma_i/\sqrt{N_i}}\right)\right) \\
&\geq 1 - \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{N_i}\alpha_i} \exp\left(\frac{-\alpha_i^2 N_i}{2}\right)
\end{aligned}$$

The probability of correct selection then becomes

$$
\begin{aligned}
APCS_m &= \prod_{i\in S_m} P\{\hat{Y}_i \leq c\} \prod_{j\neq S_m} P\{\hat{Y}_j \geq c\} \\
&\geq \prod_{i\in S_m}\left[1+\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{N_i}\alpha_i}\exp\left(\frac{-\alpha_i^2 N_i}{2}\right)\right]\prod_{i\notin S_m}\left[1-\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{N_i}\alpha_i}\exp\left(\frac{-\alpha_i^2 N_i}{2}\right)\right] \\
&= EAPCS_m
\end{aligned}
$$

where EAPCS denotes the estimated approximate values of the probability of correct selection.

The optimization problem (1) then becomes,

$$
\max_{N_1,N_2,\ldots,N_k} \prod_{i\in S_m}\left[1+\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{N_i}\alpha_i}\exp\left(\frac{-\alpha_i^2 N_i}{2}\right)\right]\prod_{i\notin S_m}\left[1-\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{N_i}\alpha_i}\exp\left(\frac{-\alpha_i^2 N_i}{2}\right)\right] \quad (2)
$$

$$
\text{subject to} \sum_{i=1}^{k} N_i = T, \qquad N_i \in \mathbb{N}, \qquad i = 1,2,...,k.
$$

Using the Lagrangian relaxation function of the optimization problem (2) and after some calculations we conclude the following theorem.

Theorem 1 *Given a total number of simulation samples T to be allocated to k competing systems in order to select a set of the best m systems whose performance is depicted by normal random variables with means $Y_1, Y_2, \ldots, Y_k$ and finite variances $\sigma_1^2, \sigma_2^2, \ldots, \sigma_k^2$ respectively. As $T \longrightarrow \infty$, the approximate probability of correct selection can be asymptotically maximized when*

$$
\frac{N_i}{\sigma_i^2/\delta_i^2} = \frac{N_j}{\sigma_j^2/\delta_j^2} \quad (3)
$$

where N_i is the number of samples allocated to system i, $\delta_i = Y_i - c$, and c is a constant satisfies $\max_{j\in S_m} Y_j \leq c \leq \min_{j\notin S_m} Y_j$.

If we fix an index s say, then we know that

$$
\alpha_s^2 N_s = \alpha_j^2 N_j,\, j \neq s \quad (4)
$$

where $\alpha_j = \delta_j/\sigma_j$. therefore,

$$
N_j = \frac{\alpha_s^2}{\alpha_j^2} N_s
$$

Since $\sum_{j\in\Theta} N_j = T$, we get

$$
\sum_{j\in\Theta} \frac{\alpha_s^2}{\alpha_j^2} N_s = T
$$

For $j \neq s$, let $D_j = \alpha_s^2/\alpha_j^2$ then $D_s = \sum_{j\neq s} \frac{\alpha_s^2}{\alpha_j^2}$ and

$$
\begin{aligned}
N_s &= T/D_s \quad (5)\\
N_j &= D_j N_s,\, j\in\Theta,\, j\neq s \quad (6)
\end{aligned}
$$

If $s \in S_m$, then $\alpha_s < 0$, $\alpha_j < 0$ for $j \in S_m$ and $\alpha_j > 0$ for $j \notin S_m$. Using (4) we get t $\sqrt{N_j}\alpha_j = \sqrt{N_s}\alpha_s,\, j \in S_m$ and $\sqrt{N_j}\alpha_j = -\sqrt{N_s}\alpha_s,\, j \notin S_m$. Therefore, using (2), the optimal value of the probability of correct election $P(CS)$, can be approximated by

$$
\begin{aligned}
AP(CS^*) &= \prod_{i\in S_m}\left[1+\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{N_i}\alpha_i}\exp\left(\frac{-\alpha_i^2 N_i}{2}\right)\right]\prod_{i\notin S_m}\left[1-\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{N_i}\alpha_i}\exp\left(\frac{-\alpha_i^2 N_i}{2}\right)\right] \\
&= \left[1+\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{N_s}\alpha_s}\exp\left(\frac{-\alpha_s^2 N_s}{2}\right)\right]^k \quad (7)
\end{aligned}
$$

Similarly, if $s \notin S_m$, then $\alpha_s > 0$, and

$$AP(CS^*) = \left[1 - \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{N_s}\alpha_s} \exp\left(\frac{-\alpha_s^2 N_s}{2}\right)\right]^k \tag{8}$$

Since the means and variances are unknown, then one can use their estimates so far to estimate the value of c and the approximated value of $P(CS^*)$.

The $OCBA_m$ Procedure for Selecting m Good Enough Systems

The idea of $OCBA_m$ algorithm is to distribute the available budget on the alternative systems based on their impact on the final solution. So it starts by giving each alternative an initial sample t_0 in order to get estimates of their objective values. Then a small increment Δ will be added to the budget and distribute them based on equation (5). we continue this way until the budget is consumed. Note that the selection of initial sample size and the increment is important, because if t_0 is large, then it means we spend more computation on non necessary alternatives. If t_0 and Δ are too small, then the estimated objective values and variances are not accurate which leads to incorrect distribution of the budget.

Numerical Example

Consider 100 normally distributed alternatives with means given by $0.1, 0.2, \ldots, 10.0$, and assume that the variance is 1.0 for all systems. Suppose that we are interested in selecting a subset of 10 best alternatives. Assume that $c = 1.05$, then the approximation of the probability of correct selection $P(CS)$ is given by Figure 1.

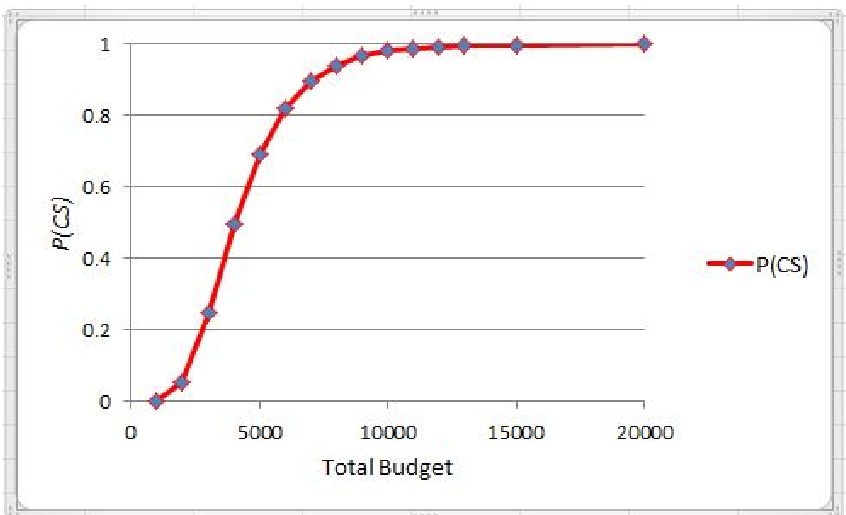

FIGURE 1. The approximate probability of correct selection as a function of the total budget T

It is clear that only 1,000 samples are needed to get the P(CS) very close to unity. Now, the procedure described above is used to estimate P(CS) of selecting the best 10 alternatives among the 100 alternatives given in the above example. The initial sample size is given by $t_0 = 10$ and the increment is given by $\Delta = 20$ per stage until the total budget exceeds $N = 20,000$ samples. Figure 2 gives the average P(CS) over 10 replications.

It is clear that the estimate of P(CS) larger than the approximate values of the $P(CS)$.

Conclusion

In this paper, we have considered the problem of selecting the best m systems among a finite but large alternative systems. The alternative systems usually represent the expected performance of some complex stochastic system.

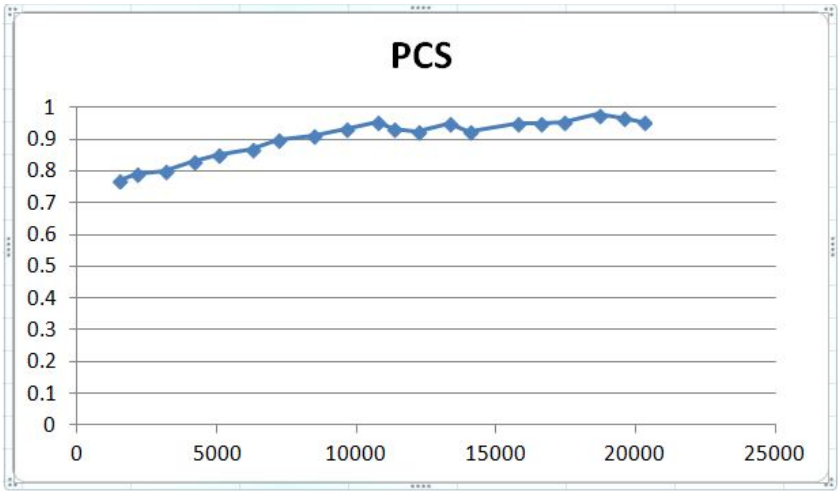

FIGURE 2. The estimated probability of correct selection as a function of T

Therefore no exact formula of the objective function is available, and the objective values have to be estimated using simulation. If the number of alternative systems is large, then simulating each system accurately may be difficult and in some cases is infeasible. Assume that there are limited computational budget to be distributed among the competent systems, then instead of distribute them evenly, we discussed how to distribute the available computational budget to the alternative systems in order to maximize the probability of correct selection, so the alternatives that have more impact on the final solution will be given more simulation samples. We also discussed how to approximate the probability of correct selection. We have tested the procedure on a generic example. The simulation shows a better performance than the approximated probability of correct selection.

REFERENCES

[1] C. H. Chen, E. Yücesan, and S. Chick, Discrete Events Dynamic Systems **10**, 251–270 (2000).

[2] C. H. Chen, C. D. Wu, and L. Dai, IEEE Transaction on Robotics and Automation **15(1)**, 44–56 (1999).

[3] H. C. Chen, C. H. Chen, L. Dai, and E. Yücesan, "New development of optimal computing budget allocation for discrete event simulation," in *Proceedings of the 1997 Winter Simulation Conference*, edited by D. H. W. S. Andrad.ttir, K. J. Healy and B. L. Nelson (Institute of Electrical and Electronics Engineers, Inc., Piscataway, New Jersey, 1997), pp. 334–341.

[4] J. Banks, *Handbook of simulation* (John Wiley, 1998).

[5] C. H. Chen, IEEE Transactions on Automatic Control **41**, 1227–1231 (1996).

[6] D. Goldsman and B. L. Nelson, "Ranking, selection and multiple comparisons in computer simulation," in *Proceedings of the 1994 Winter Simulation Conference*, edited by J. D. Tew, S. Manivannan, D. A. Sadowski, and A. F. Seila (Institute of Electrical and Electronics Engineers, Inc., Piscataway, New Jersey, 1994), pp. 192–199.

[7] M. H. Alrefaei and M. H. Almomani, Journal of the Franklin Institute **344**, 495–506 (2007).

[8] M. H. Almomani and M. H. Alrefaei, Journal of Applied Probability and Statistics **6(1&2)**, 13–27 (2012).

[9] M. H. Almomani and M. H. Alrefaei, Asia-Pacific Journal of Operational Research **33** (2016), 10.1142/S0217595916500093.

[10] M. Al-Salem, M. H. Almomani, M. H. Alrefaei, and A. Diabat, Simulation Modelling Practice and Theory **71**, 149–159 (2017).

[11] C. H. Chen, D. He, M. Fu, and L. H. Lee, Informs Journal on Computing **20(4)**, 579–595 (2008).

[12] C. H. Chen, D. He, M. Fu, and L. H. Lee, "Efficient selection of an optimal subset for optimization under uncertainty," in *Proceedings of the 2007 INFORMS Simulation Society Research Workshop* (2007), pp. 1–5.

[13] S. Ross, *A first Course in Probability* (Macmillan, 1994).