Proceedings of the 2000 Winter Simulation Conference
J. A. Joines, R. R. Barton, K. Kang, and P. A. Fishwick, eds.

AN ENHANCED TWO-STAGE SELECTION PROCEDURE

E. Jack Chen
W. David Kelton

Department of Quantitative Analysis and Operations Management
University of Cincinnati
Cincinnati, OH 45221, U.S.A.

ABSTRACT

This paper discusses implementation of a two-stage procedure to determine the simulation run length for selecting the best of k designs. We purpose an *Enhanced Two-Stage Selection* (ETSS) procedure. The number of additional replications at the second stage for each design is determined by both the variances of the sample means and the differences of the sample means of alternative designs. We show that the ETSS procedure gives valid selections with significantly reduced simulation replications compared to Rinott's procedure. An experimental performance evaluation demonstrates the validity of the ETSS procedure.

1 INTRODUCTION

Discrete event simulations are widely used to compare alternative system designs or operating policies. When evaluating k alternative system designs, we would like to select the best of these k designs and to control the probability that the selected design really is the best one. Let μ_i denote the expected response of design i. Our goal is to find the design with the smallest expected response $\mu^* = \min_{i=1,2,\ldots,k}(\mu_i)$. If the goal is to select a design with the biggest expected response, the procedure can be modified easily to accommodate that. We achieve this goal by using a class of ranking and selection (R&S) procedures.

Many R&S procedures are directly or indirectly based on Dudewicz and Dalal (1975) or Rinott's (1978) indifference-zone-selection procedure. Unfortunately, their procedures require the simulation output sequence to be independent and normally distributed. When the simulation output are sample averages, by the Central Limit Theorem (CLT), the normality assumption is typically valid provided that the sample size is large enough. The independence assumption, though, requires more attention. Moreover, many indifference-zone-selection procedures determine the number of additional replications based on a conservative *least favorable configuration* (LFC) assumption (Rinott 1978); see Section 2.5. Some new approaches (Chen et al. 1999) incorporate first-stage sample mean information in determining the number of additional replications. They suggest that an average case analysis may lead to a significant reduction in computing effort, relative to indifference-zone-selection procedures. For an overview of existing methods of ranking and selection see Law and Kelton (2000), or Bechhofer, Santner, and Goldsman (1995).

Generally speaking, we can improve the efficiency of R&S procedures with a *pre-selection* (Goldsman et al. 1999). The pre-selection approach is a screening device that attempts to select a (random-size) subset of the k alternative designs that contains the best one. The inferior designs will be excluded from further consideration, reducing the overall simulation time. After we cut down a large number of alternatives into a more manageable number, we then carry out a higher accuracy selection to make the fine-tuned choice of the best. In some extreme cases, the subset may contain only one design, i.e., we select outright the best design. We propose an Enhanced Two-Stage Selection (ETSS) procedure that is efficient in allocating the required simulation replications at the second stage.

This paper derives an enhanced two-stage selection procedure that utilizes the information of both the means and variances obtained from the first stage to determine the number of additional replications. In Section 2 we recall Rinott's (1978) procedure that serves as reference for comparison. We provide background necessary to understand our proposed procedure. In Section 3 we present our methodologies and proposed procedure for the ranking and selection. In Section 4, we show our empirical-experiment results. The new procedure compares favorably with Rinott's (1978) procedure. In Section 5, we give concluding remarks.

2 BACKGROUND

To facilitate what follows we define some notation:

N_i:	the number of replications (or batches) for design i,
σ_i^2:	the variance of design i,

X_{ij}: the average of the observations from the j^{th} replication of the i^{th} design, or the j^{th} batch mean of the i^{th} design of a terminating or steady-state simulation.

μ_i: the mean performance measure for design i, i.e., $\mu_i = E(X_{ij})$,

$\hat{\mu}_i$: the sample mean performance measure for design i, i.e., $\sum_{j=1}^{N_i} X_{ij}/N_i$,

$S_i^2(N_i)$: the sample variance of design i with N_i replications.

2.1 Indifference-Zone-Selection Procedure

Let μ_{i_l} be the l^{th} smallest of the μ_i's, so that $\mu_{i_1} \leq \mu_{i_2} \leq \ldots \leq \mu_{i_k}$. Our goal is to select a design with the smallest expected response, μ_{i_1}. Let "CS" denote this event of "correct selection." In stochastic simulation, CS typically cannot be guaranteed with certainty. Furthermore, if μ_{i_1} and μ_{i_2} are very close together, we might not care if we mistakenly choose design i_2, whose expected response is μ_{i_2}. This "practically significant" difference d^* (a positive real number) is called the *indifference zone* in the statistical literature. Therefore, we want a procedure that avoids making a large number of replications to resolve this unimportant difference. That is, we want P(CS) $\geq P^*$ provided that $\mu_{i_2} - \mu_{i_1} \geq d^*$, where the minimal CS probability P^* and the "indifference" amount d^* are both specified by the user.

2.2 Two-Stage Rinott Procedure

The two-stage procedure of Rinott (1978) has been widely applied. Let n_0 be the number of initial replications from each design. The first-stage sample means $\hat{\mu}_i = \sum_{j=1}^{n_0} X_{ij}/n_0$, and marginal sample variances

$$S_i^2(n_0) = \frac{\sum_{j=1}^{n_0}(X_{ij} - \hat{\mu}_i)^2}{n_0 - 1}, \text{ for } i = 1, 2, \ldots, k$$

are computed. Based on the number of initial replications n_0 and the sample variance estimates $S_i^2(n_0)$ obtained from the first stage, the number of additional simulation replications for each design in the second stage is $N_i - n_0$, where

$$N_i = \max(n_0, \lceil (hS_i(n_0)/d^*)^2 \rceil), \text{ for } i = 1, 2, \ldots, k, \quad (1)$$

where $\lceil z \rceil$ is the smallest integer that is greater than or equal to the real number z, and h (which depends on k, P^*, and n_0) is a constant that solves Rinott's (1978) integral (h can also be found from the tables in Wilcox, 1984). We then compute the overall sample means $\bar{\hat{\mu}}_i = \sum_{j=1}^{N_i} X_{ij}/N_i$, and select the design with the smallest $\bar{\hat{\mu}}_i$ as the best one. Basically, the computing budget is allocated proportional to the estimated sample variances.

2.3 Subset Selection

Goldsman et al. (1999) describe a procedure to pre-select the subset of designs so that inferior designs can be excluded from further simulation.

Design i_l, whose sample mean $\hat{\mu}_{i_l} = \min_{i=1,2,\ldots,k}(\hat{\mu}_i)$, is automatically included in the subset. For all $i \neq i_l$, let

$$t = t_{1-(P^*)^{\frac{1}{k-1}}, n_0 - 1},$$

$$S_{i,i_l}^2 = \frac{\sum_{j=1}^{n_0}(X_{ij} - X_{i_l j} - (\hat{\mu}_i - \hat{\mu}_{i_l}))^2}{n_0 - 1},$$

and

$$W_{i,i_l} = tS_{i,i_l}/\sqrt{n_0}.$$

Then design i will be included in the subset only if $\hat{\mu}_i - \hat{\mu}_{i_l} \leq (W_{i,i_l} - d^*)^+$, where $(a)^+ = \max(0, a)$, and $t_{1-\alpha,n}$ is the $1-\alpha$ quantile of the t distribution with n degrees of freedom (df).

2.4 Optimal Computing Budget Allocation (OCBA)

Chen et al. (1999) point out that a drawback of Rinott's procedure is that only the information of variances is used when determining the additional replications at the second stage. They propose an OCBA that utilizes the information of both the means and variances obtained from the first stage. Their procedure is based on a fixed total computing budget $T = \sum_{i=1}^{k} N_i$ and attempts to maximize P(CS).

They use the *Approximate Probability of Correct Selection*(APCS) as a lower bound on the P(CS). That is

$$\text{P(CS)} \geq 1 - \sum_{l=2}^{k} \text{P}(\hat{\mu}_{i_1} > \hat{\mu}_{i_l}).$$

The right hand side of the above equation is the $APCS$. They show that for a fixed number of replications, the $APCS$ can be asymptotically maximized when

$$\frac{N_{i_1}}{N_{i_2}} \to \frac{\sigma_{i_1}}{\sigma_{i_2}} \left[\sum_{l=2}^{k} \left(\frac{\hat{\delta}_{i_2,i_1}^2}{\hat{\delta}_{i_l,i_1}^2} \right) \right]^{1/2}, \quad (2)$$

$$\frac{N_{i_l}}{N_{i_2}} \to \left(\frac{\sigma_{i_l}/\hat{\delta}_{i_l,i_1}}{\sigma_{i_2}/\hat{\delta}_{i_2,i_1}} \right)^2 \text{ for } l = 3, 4, \ldots, k, \quad (3)$$

where $\hat{\delta}_{i_l,i_1} = \hat{\mu}_{i_l} - \hat{\mu}_{i_1}$, and σ_i is the standard deviation of the response of design i. However, in reality the value

of σ_i is unknown, so the standard deviation of the sample responses $S_i(n_0)$ will be used.

2.5 Motivation

Rinott's procedure is a conservative procedure that obtains $\text{P(CS)} \geq P^*$ by assuming $\mu_{i_l} = \mu_{i_1} + d^*$ for $l = 2, 3, \ldots, k$, i.e., the LFC. In reality, we rarely encounter the LFC. The OCBA modifies Rinott's procedure and takes into consideration the difference between mean estimators $\hat{\delta}_{i_l,i_1} = \hat{\mu}_{i_l} - \hat{\mu}_{i_1}$ for $l = 2, 3, \ldots, k$. However, OCBA is a fixed-sample-size procedure; it attempts to maximize the P(CS) with a fixed total computing budget. Moreover, the achieved P(CS) cannot be estimated until the end of the procedure. If the achieved P(CS) is less than desired then the simulation needs to be restarted with a bigger simulation budget. We propose an ETSS procedure, which also takes into consideration the difference between mean estimators $\hat{\delta}_{i_l,i_1}$ for $l = 2, 3, \ldots, k$ when estimating the required replications at the second stage. However, the ETSS attempts to minimize the total number of second-stage replications with a given P^*. The approach of the OCBA and the ETSS are completely different. However, the final allocation strategies are very similar. The ETSS provides some insights into the optimization results of the OCBA.

3 METHODOLOGIES

In this section we introduce the methodologies we used in our procedures. To meet the i.i.d. samples requirement in the case of steady-state simulation , one can use the popular output-analysis method: *batch means*, for example Chen and Kelton (2000a,b). We incorporate subset pre-selection and the two-stage Rinott procedure to form the Enhanced Two-Stage Selection (ETSS) procedure.

3.1 Enhanced Two-Stage Selection Procedure

Using the QI procedure, we will simulate n_0 replications for each design at the first stage. We then perform the subset pre-selection at the end of the first stage. Therefore, stage two will evaluate only $k' \leq k$ designs, where k' is the number of designs included in the subset.

In Rinott's procedure, only one positive real number h is used to estimate additional simulation replications at the second stage, see Section (2.2). That is, only one variable h is used in Rinott's integral for all k designs based on the LFC. Without loss of generality, assuming $\mu_{i_l} - \mu_{i_1} \geq d^*$, for $i_l \neq i_1$. If we let f and F denote the density and distribution function, respectively, of the t distribution with $n_0 - 1$ df, the equation

$$\text{P(CS)} \geq P^* = \int_{-\infty}^{\infty} F\,(h+t)^{k-1}\, f(t)dt \qquad (4)$$

is based on the LFC, $\text{P}(\hat{\mu}_{i_1} < \hat{\mu}_{i_l} | \mu_{i_1} + d^* = \mu_{i_l}) = (P^*)^{1/(k-1)}$ for $l = 2, 3 \ldots, k$, see Law and Kelton (2000), pp. 575-576. However, we seldom encounter the LFC in practice, i.e., $\mu_{i_1} + d^* < \mu_{i_l}$ for some i_l. Thus, with all other things being equal

$$\text{P}(\hat{\mu}_{i_1} < \hat{\mu}_{i_l} | \mu_{i_1} + d^* < \mu_{i_l}) > (P^*)^{1/(k-1)}.$$

Because the computation of N_i is based on the LFC, it is larger than necessary when $\mu_{i_1} + d^* < \mu_{i_l}$ for some i_l.

To eliminate this deficiency, we propose an enhanced simulation replications allocation algorithm. Based on the equation

$$\text{P(CS)} = \Pi_{l=2}^{k} \text{P}(\hat{\mu}_{i_1} < \hat{\mu}_{i_l} | \mu_{i_1} + d^* \leq \mu_{i_l}),$$

it can be shown that

$$\text{P(CS)} = \int_{-\infty}^{\infty} \Pi_{l=2}^{k} F\left(\frac{\mu_{i_l} - \mu_{i_1}}{d^*/h} + t\right) f(t)dt. \qquad (5)$$

For $l = 2, 3, \ldots, k$, we will set

$$P(\hat{\mu}_{i_1} < \hat{\mu}_{i_l} | \delta_{i_l,i_1} = \mu_{i_l} - \mu_{i_1}) = (P^*)^{1/(k-1)},$$

then

$$P^* = \Pi_{l=2}^{k} P(\hat{\mu}_{i_1} < \hat{\mu}_{i_l} | \delta_{i_l,i_1} = \mu_{i_l} - \mu_{i_1}).$$

Note that in contrast to the LFC, δ_{i_l,i_1} may be larger than d^*. In essence, we propose using different h_i values for different designs i. Let

$$r_i = \max(\mu_i - \mu_{i_1}, d^*)/d^*, \qquad (6)$$

and

$$h_i = h/r_i, \qquad (7)$$

then equation (5) becomes

$$\text{P(CS)} = \int_{-\infty}^{\infty} \Pi_{l=2}^{k} F\left(r_{i_l} h_{i_l} + t\right) f(t)dt.$$

Therefore,

$$\text{P(CS)} \;=\; \int_{-\infty}^{\infty} F\,(h+t)^{k-1}\, f(t)dt. \qquad (8)$$

Here h is the same as Rinott's procedure. We would like to point out that this integral is an approximation suggested in Rinott (1978), which includes some comparison between

the integral he proposed with this approximation. Solve h from the above equation with P(CS) set to P^*, and let

$$N_i = \max(n_0, \lceil (h_i S_i(n_0)/d^*)^2 \rceil), \text{ for } i = 1, 2, \ldots, k. \quad (9)$$

In practice, however, exact value of r_i cannot be computed because μ_i is unknown, so

$$\hat{r}_i = \max(\hat{\mu}_i - \hat{\mu}_{i_1}, d^*)/d^*$$

will be used.

The difference between equations (9) and (1) is that h_i instead of h is used. The information of the difference of the sample means between alternative designs is embedded in the value of h_i through r_i of equation (6), thus, equation (9) utilizes the information of both the means and variances. We would like to point out that the derivation of this N_i is based entirely on equalities, thus, it is optimal. However, the true value of r_i is unknown and $\hat{r}_i$ will be used. Because $\hat{r}_i$ is a random variable, P(CS)=P^* may be not true for this heuristic approach. To be conservative, a constant $0 < c < 1$ can be used so that $\hat{r}_i = \max(c(\hat{\mu}_{i_l} - \hat{\mu}_{i_1}), d^*)/d^*$, or the $(1-\alpha)$ confidence limits of μ_i can be used to compute $\hat{r}_i$, i.e., $\hat{r}_i = \max(L(\hat{\mu}_i) - \hat{\mu}_{i_1}, d^*)/d^*$, where $L(\hat{\mu}_i)$ are the $1-\alpha$ lower confidence limits of μ_i. On the LFC equation (4) becomes P(CS) = P^*, therefore, the N_i's determined by equation (9) will be exactly the same as if they are obtained from Rinott's procedure. However, if $\hat{\mu}_{i_1} + d^* < \hat{\mu}_i$, then less computational budget will be allocated for design i than that allocated according to Rinott's procedure, because $h_i < h$. In cases that $\hat{\mu}_{i_1} + d^* > \hat{\mu}_i$, $\hat{r}_i = 1$ is used because the difference between these two sample means is less than the indifference amount; it will be considered as correct selection if this μ_i is selected. Therefore, it is not necessary to make a large number of replications to resolve this insignificant difference.

Interestingly enough, if we compute the ratio N_{i_l}/N_{i_2} with N_i estimated from equation (9) and assuming $\hat{\delta}_{i_2,i_1} \geq d^*$, we get

$$\frac{N_{i_l}}{N_{i_2}} = \left(\frac{S_{i_l}(n_0)/\hat{\delta}_{i_l,i_1}}{S_{i_2}(n_0)/\hat{\delta}_{i_2,i_1}} \right)^2 \text{ for } l = 2, 3, \ldots, k.$$

This is the same as equation (3) in Section (2.4). Therefore, for $l = 2, 3, \ldots, k$, the ratio N_{i_l}/N_{i_2} will be the same as that in OCBA. On the other hand, the ratio

$$\frac{N_{i_1}}{N_{i_2}} = \hat{r}_{i_2}^2 (S_{i_1}(n_0)/S_{i_2}(n_0))^2.$$

This is different from that of OCBA, i.e., equation(2), because OCBA does not use the indifference parameter d^* and OCBA attempts to maximize P(CS) with a fixed total simulation budget, while we try to minimize total simulation budget with given P^*. If the user is not able to provide the indifference amount d^*, then $d^* = \hat{\delta}_{i_2,i_1}$ will be used in the ETSS procedure. Therefore, if the user attempts to maximize P(CS) with a given simulation budget then OCBA can be used. On the hand, the ETSS can be used not only when the user attempts to minimize simulation budget with a given P^*, but also when the user attempts to maximize P(CS) with a given simulation budget. For example, if ETSS determines the total number of simulation replications $T = \sum_{i=1}^{i=k} N_i$ and the available budget is $A < T$, then we will use $N_i = N_i \times A/T$ replications for design i.

The ETSS algorithm:

1. Simulate n_0 replications.
2. Perform subset pre-selection according to procedures developed by Goldsman et al. (1999), see Section (2.3).
3. If there is only one element in the subset, go to step 6.
4. For each design i in the subset, compute the needed additional replications $N_i - n_0$. Here N_i will be computed according to equation (9).
5. Simulate $N_i - n_0$ additional replications for each design i in the subset.
6. Return the values i_1 and $\bar{\hat{\mu}}_{i_1}$.

Because the information of the means is used in computing additional replications, ETSS has better performance than the performance of Rinott procedure.

3.2 Properties of ETSS

Goldsman and Schmeiser (1997) list some properties that a good estimator should posses. We use these properties to assess the desirability of our algorithm. The followings describe the performance of our algorithm under each property.

- Statistical performance. If we obtain reasonably accurate estimates of the differences of the mean responses and variances from the first stage, then the P(CS) from our ETSS should be close to the pre-specified level P^*.
- Ease of computation. Our algorithm involves only little more than $O(N)$ operations, N is the number of observations. The enhancement, calculation of $\hat{r}_i$, added to Two-Stage Rinott procedure is computationally inexpensive.
- Parsimonious storage requirements. Our data storage is k, one for the each sample means.

- Ease of understanding. Our algorithm is a straightforward enhancement of Rinott's procedure.
- Numerical stability. The limits of machine precision is the limits of our algorithm precision.
- User-specified parameters. We require only two parameters: the desired probability of correct selection P^*, and the indifference amount d^*.
- Amenability for use in algorithms. Our algorithm can be incorporated with other procedures easily.

4 EMPIRICAL EXPERIMENTS

In this section we present some empirical results obtained from simulations using Rinott's procedure and the ETSS procedure proposed in this paper. The purpose of the experiments was not so much to test the methods thoroughly, but rather to demonstrate the interdependence between the variances and the differences of the sample means and simulation run lengths, and the validity of our methods.

In this experiment, we focus on the validity and performance of the ETSS at the second stage. Therefore, step 2 of the ETSS was skipped. Moreover, we assume that there is more than one element in the subset. Furthermore, the input data X_{ij} for step 4 will be i.i.d. normal because they are batch means, and the difference of $X_{i_1 j}$ and $X_{i_k j}$ will be small because they are in the pre-selection subset.

4.1 Experiment 1: Equal Variances

There are ten alternative designs in the pre-selection subset. Suppose $X_{ij} \sim \mathcal{N}(i, 6^2)$, $i = 1, 2, \ldots, 10$, where $\mathcal{N}(\mu, \sigma^2)$ denotes the normal distribution with mean μ and variance σ^2. We want to select a design with the minimum mean. It is obvious that design 1 is the actual best design. The indifference amount d^* is set to 0.90 for all cases. We compare the required simulation replications of Rinott's procedure and our ETSS. Furthermore, 10,000 independent experiments are performed to obtain the actual P(CS), we count the number of times we successfully selected the true best design (design 1 in this example) in those 10,000 independent experiments. P(CS) is then obtained by dividing this number by 10,000 representing the correct selection percentage. We use two different initial replications $n_0 =$ 20, and 30.

Table 1: P(CS) and Sample Sizes for Experiment 1

	$P^* = 0.90$		$P^* = 0.95$	
Procedure	P(CS)	avg s.r.	P(CS)	avg s.r.
Rinott(20)	98.94%	4284	99.73%	6720
ETSS(20)	98.08%	1335	98.91%	1929
Reduction		2949		4791
Rinott(30)	99.45%	5020	99.69%	6343
ETSS(30)	98.94%	1433	99.27%	1760
Reduction		3587		4583

The results of our experiment 1 are summarized in Table 1. The *P(CS)* column lists the percentage of correct selection. The *avg s.r.* column lists the average of the total simulation replications ($\sum_{i=1}^{k} N_i$) used in each procedure. The *Rinott(20)* row lists the results of Rinott's procedure with initial replications $n_0 = 20$. The *ETSS(20)* row lists the results of ETSS procedure with initial replications $n_0 =$ 20. The *Reduction* row lists the reduction of simulation replications achieved by the ETSS procedure. Note that the P(CS) are all larger than the specified $P^* = 0.90$ and $P^* =$ 0.95. This is an indication that both selection procedures are conservative, which is expected for Rinott's procedure because it is based on the LFC. On the other hand, this is a nice result for the ETSS procedure because its derivation is not based on the LFC. ETSS is more efficient than Rinott's procedure because when determining additional simulation replications, ETSS exploits the information of both the sample means and variances. The sample means can provide valuable information of relative differences among different designs.

Table 2 lists the detailed simulation replications used for each design under different selection procedures with $P^* = 0.95$ and $n_0 = 20$. The *Rinott* column lists the average simulation replications for each design under Rinott's procedure. The *ETSS* column lists the average simulation replications for each design under ETSS procedure. The *Reduction* column lists the reduction of simulation replications, and the *R%* column lists the percentage of the reduction of the number of simulation replications. We would like to point that Rinott's procedure will be the same as equal allocation for additional simulation replications in this settings, i.e., the variances are equal for all designs. Our experimental results confirm that. On the other hand, in the ETSS the number of additional simulation replications decreases as the differences $\hat{\delta}_{i,i_1}$ increase. This makes good sense because as the difference of $\mu_i - \mu_{i_1} (> 0)$ increases, it is more likely that $\hat{\mu}_i > \hat{\mu}_{i_1}$. In another words, as the observed $\hat{\delta}_{i,i_1} = \hat{\mu}_i - \hat{\mu}_{i_1} (> 0)$ increases, it is less likely that $\mu_i < \mu_{i_1}$. The ratio of the average number of simulation replications allocated for design 10 and design 1

Table 2: Detailed Sample Sizes for $P^* = 0.95$ and $n_0 = 20$ of Experiment 1

Design	Rinott	ETSS	Reduction	R%
1	706	700	6	0.85%
2	664	437	227	34.19%
3	669	313	356	53.21%
4	672	199	473	70.39%
5	666	112	554	83.18%
6	672	61	611	90.92%
7	670	36	634	94.63%
8	666	25	641	96.26%
9	664	21	643	96.84%
10	665	20	645	96.99%

of Rinott's procedure is 0.942 (665/706), which is close to the theoretical value 1 ($((S_{10}(n_0)/S_1(n_0))^2 = (6/6)^2$). On the other hand, this ratio is only 0.0286 (20/700) under the ETSS procedure. This is where ETSS can significantly improve the efficiency of Two-Stage Rinott procedure.

4.2 Experiment 2: Increasing Variances

This is a variation of experiment 1. All settings are preserved except that the variance of each design increases as the mean increases. Namely, $X_{ij} \sim \mathcal{N}(i, (6 + i/2)^2)$, $i = 1, 2, \ldots, 10$.

The results are listed in Tables 3 and 4. Because most designs have larger variances than those in experiment 1, the total number of simulation replications is more than in experiment 1. We are less confident of the best selection with these settings. Therefore, more simulation replications are needed to obtain the desired confidence. Rinott's procedure will allocate more additional simulation replications for designs with larger variances and the simulation replications allocation is based entirely on the variances, thus, $N_i > N_j$ when $S_i > S_j$. Even though ETSS also allocates more additional simulation replications for designs with larger variances, it takes into account the difference of the sample means. Therefore, it is not always true that $N_i > N_j$ when $S_i > S_j$. In fact, δ_{i,i_1} has such a big influence in h_i that the additional simulation replications decrease as the variances increase. The ratio of the average simulation replications allocated for design 10 and design 1 of Rinott's procedure is 2.77 (2042/738), which is smaller than the theoretical value 3.36 ($(11/6)^2$). This ratio is only 0.057 (42/736) under the ETSS procedure. ETSS is most effective in this setting, achieving about 80% reduction in the number of simulation replications.

4.3 Experiment 3: Decreasing Variances

This is another variation of experiment 1. All settings are preserved except that the variance of each design decreases as the mean increases. Namely, $X_{ij} \sim \mathcal{N}(i, (6 - i/2)^2)$, $i = 1, 2, \ldots, 10$.

The results are listed in Tables 5 and 6. Because most designs have smaller variances than those in experiment 1, the total number of simulation replications is less than in experiment 1. We are more confident of the best selection with these settings. Therefore, fewer simulation replications are needed to obtain the desired confidence. Rinott's procedure allocates fewer additional simulation replications for designs with inferior designs in this setting, i.e., as the sample means increase the variances decrease. The ratio of the average number of simulation replications allocated for design 10 and design 1 of Rinott's procedure is 0.0623 (43/690), which is larger than the theoretical value 0.027 ($(1/6)^2$). This ratio is 0.029 (20/686) under the ETSS procedure. ETSS achieves less improvement in this setting. However, the percentage reduction of the number of simulation replications is still about 50%.

Table 3: P(CS) and Sample Sizes for Eperiment 2

	$P^* = 0.90$		$P^* = 0.95$	
Procedure	P(CS)	avg s.r.	P(CS)	avg s.r.
Rinott(20)	99.46%	10300	99.70%	13031
ETSS(20)	98.28%	2108	98.73%	2701
Reduction		8192		10330
Rinott(30)	99.42%	9754	99.81%	12287
ETSS(30)	98.62%	1835	99.08%	2279
Reduction		7919		10008

Table 4: Detailed Sample Sizes for $P^* = 0.95$ and $n_0 = 20$ of Experiment 2

Design	Rinott	ETSS	Reduction	R%
1	738	736	2	0.27%
2	780	508	272	34.87%
3	900	428	472	52.44%
4	1039	333	706	67.95%
5	1176	243	933	79.34%
6	1333	162	1171	87.85%
7	1505	112	1393	92.56%
8	1670	78	1592	95.33%
9	1842	53	1789	97.12%
10	2042	42	2000	97.94%

Table 5: P(CS) and Sample Sizes for Experiment 3

	$P^* = 0.90$		$P^* = 0.95$	
Procedure	P(CS)	avg s.r.	P(CS)	avg s.r.
Rinott(20)	99.41%	2386	99.71%	3033
ETSS(20)	98.64%	1232	98.95%	1525
Reduction		1154		1508
Rinott(30)	99.34%	2269	99.78%	3018
ETSS(30)	98.79%	1200	99.38%	1459
Reduction		1069		1559

Table 6: Detailed Sample Sizes for $P^* = 0.95$ and $n_0 = 20$ of Experiment 3

Design	Rinott	ETSS	Reduction	R%
1	690	686	4	0.58%
2	565	369	196	34.69%
3	463	216	247	53.35%
4	376	102	274	72.87%
5	298	44	254	85.23%
6	231	24	207	89.61%
7	169	20	149	88.17%
8	117	20	97	82.91%
9	76	20	56	73.68%
10	43	20	23	53.49%

4.4 Experiment 4: Dependent Data and Equal Variances

This is another variation of experiment 1 and is designed to check the robustness of both procedures when the assumption of independence is violated. All settings are preserved except that data are correlated. The *first-order auto-regressive* (AR(1)) process, generated by the recurrence relation

$$X_{ij} = \mu_i + \varphi(X_{ij-1} - \mu_i) + \epsilon_j \text{ for } j = 1, 2, \ldots,$$

where

$$E(\epsilon_j) = 0, \quad E(\epsilon_j \epsilon_k) = \begin{cases} \sigma^2 & \text{if } j = k \\ 0 & \text{otherwise} \end{cases},$$

$$0 < \varphi < 1,$$

and X_{i0} is specified to some random variate x_{i0} drawn from the steady-state distribution. The AR(1) process shares many characteristics observed in simulation output processes, including first- and second-order nonstationarity and autocorrelations that decline exponentially with increasing lag. If we make the additional assumption that the ϵ_j's are normally distributed, since we have already assumed that they are uncorrelated, they will now be independent as well, i.e., the ϵ_j's are i.i.d. $\mathcal{N}(0, \sigma^2)$. It can be shown that X_{ij} has asymptotically a $\mathcal{N}(\mu_i, \frac{\sigma^2}{1-\varphi^2})$ distribution.

If we set $\mu_i = i$, $\varphi = 0.5$, and $\sigma = 5^2$, then $X_{ij} \sim \mathcal{N}(i, 100/3)$, $i = 1, 2, \ldots, 10$. Table 7 contains the simulation results for the two selection procedures. We can see that the relative performances of different procedures are very similar with what we saw in previous experiments, except that, in general, smaller sample sizes are used in this experiment. This is due to a smaller variance ($100/3 < 6^2$) of the input sequence X_{ij}'s and most probably dependence within alternatives. Both procedures underestimated the variance when data are not independent, so the simulation replications are not large enough and the P(CS) is not as good as that in previous experiments. However, ETSS is still more efficient than Rinott's procedure in term of simulation replications. Table 8 lists the detailed simulation replications used for each design under different selection procedures with $P^* = 0.95$ and $n_0 = 20$.

Table 7: P(CS) and Sample Sizes for Experiment 4

	$P^* = 0.90$		$P^* = 0.95$	
Procedure	P(CS)	avg s.r.	P(CS)	avg s.r.
Rinott(20)	92.06%	4568	94.43%	5804
ETSS(20)	90.49%	1499	92.75%	1858
Reduction		3069		3946
Rinott(30)	91.39%	4342	93.83%	5474
ETSS(30)	89.50%	1429	92.29%	1759
Reduction		2913		3715

Table 8: Detailed Sample Sizes for $P^* = 0.95$ and $n_0 = 20$ of Experiment 4

Design	Rinott	ETSS	Reduction	R%
1	591	593	-2	-0.34%
2	577	366	211	36.57%
3	580	289	291	50.17%
4	580	209	371	63.97%
5	580	137	443	76.38%
6	578	93	485	83.91%
7	581	59	522	89.85%
8	577	42	535	92.72%
9	577	34	543	94.11%
10	578	31	547	94.64%

5 CONCLUDING REMARKS

Many two-stage indifference-selection procedures ignore a large amount of first-stage sampling information. We have presented a simulation-replication-allocation algorithm that utilizes both the means and variances from the first stage. Even though the approach of OCBA and ETSS are different, the computing budget allocation strategy are very similar. However, OCBA is a fixed-sample-size procedure and the achieved P(CS) cannot be computed until the end of the procedure. On the other hand, ETSS attempts to minimize the computing budget with a given minimal probability of correct selection P^*. Moreover, ETSS provides some insight into the computing budget allocation strategy of OCBA.

At the end of the first stage of the ETSS, inferior designs are excluded from further consideration. Therefore, little effort is expended on simulating inferior designs. Instead, most effort will be allocated to achieve higher accuracy of more promising designs. The second stage of our ETSS is a straightforward enhancement to Rinott's procedure and is very easy to implement. The required number of replications at the second stage for each design is computed based on both the variances of sample means among the same design and the differences of sample means between alternative designs. The marginal computation effort required for our ETSS is minimal, yet the achieved efficiency improvement is significant. This simulation-replication-allocation technique can also be applied to other ranking-and-selection criteria such as selecting the m best of k designs, see Koenig and Law (1985), or multiple comparison with the best, see Matejcik and Nelson (1995).

The results from our empirical experiments show that the ETSS procedure is a powerful tool for selecting the best design out of k alternatives. The main advantage of the ETSS is that the algorithm determines the number of additional simulation replications based on both the means and variances and thus significantly improves the efficiency

of R&S procedures. The simplicity of this method should make it attractive to simulation practitioners or software developers.

We are not able to give a definitive recommendation on the choice of the number of initial replications n_0 at the first stage. The "optimal" choice of n_0 depends on the variances of X_{ij}'s, which are unknown. If n_0 is chosen too small, overestimated variances from the first stage might result in excessively large second-stage replications. Moreover, inaccurate mean estimates in the first stage may cause the final P(CS) to be smaller than desired. Because the ETSS procedure relies heavily on the accuracy of the variance estimates and the mean estimates from the first stage to compute the required simulation replications. In general we would recommend that n_0 be large enough to obtain reasonable variance estimates. Especially when N_i is much larger than n_0, there is little to lose with a larger n_0.

To avoid the risk of relying too much on the variance estimates from only one stage, multistage and sequential selection procedures have been developed, see Law and Kelton (2000). Because the value of h in equation (8) depends on the number of replications, which will be different for different designs at the end of the second-stage, it may not be easy to modify the ETSS into a sequential procedure.

REFERENCES

Bechhofer, R. E., T. J. Santner and D. Goldsman. 1995. *Design and analysis of experiments for statistical selection, screening and multiple comparisons*. New York: John Wiley & Sons, Inc.

Chen, H. C., C. H. Chen, J. Lin and E. Yücesan. 1999. An asymptotic allocation for simultaneous simulation experiments. In *Proceedings of the 1999 Winter Simulation Conference*, ed. P.A. Farrington, H.B. Nembhard, D.T. Sturrock, and G.W. Evans, 359–366. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.

Chen, E. J., and W. D. Kelton. 2000a. Mean estimation based on ϕ-mixing sequences. In *Proceedings of the* 33^{rd} *Annual Simulation Symposium*.

Chen, E. J., and W. D. Kelton. 2000b. A stopping procedure based on phi-mixing conditions. In *Proceedings of the 2000 Winter Simulation Conference*, ed. J. A. Joines, R. Barton, P. Fishwick, and K. Kang. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.

Dudewicz, E. J. and S. R. Dalal. 1975. Allocation of observations in ranking and selection with unequal variances. *Sankhya* B37: 28–78.

Goldsman D., B. L. Nelson, T. Opicka, and A. A. B. Pritsker. 1999. A ranking and selection project: Experiences from a university-industry collaboration. In *Proceedings of the 1999 Winter Simulation Conference*, ed. P.A. Farrington, H.B. Nembhard, D.T. Sturrock, and G.W. Evans, 83–92. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.

Goldsman D., and B. W. Schmeiser. 1997. Computational efficiency of batching methods. In *Proceedings of the 1997 Winter Simulation Conference*, ed. S. Andradóttir, K.J. Healy, D.H. Withers, and B.L. Nelson, 202–207. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.

Koenig, L. W., and A. M. Law. 1985. A procedure for selecting a subset of size m containing the l best of k independent normal populations, with applications to simulation. *Commun. Statist. - Simulation and Computation* 14: 719–734.

Law, A. M., and W. D. Kelton. 2000. *Simulation modeling and analysis*. 3rd ed. New York: McGraw-Hill.

Matejcik, F. J. and B. L. Nelson. 1995. Two-Stage multiple comparisons with the best for computer simulation. *Operations Research* 43: 633–640.

Rinott, Y., 1978. On two-stage selection procedures and related probability inequalities. *Communications in Statistics* A7: 799-811.

Wilcox, R. R., 1984. A table for rinott's selection procedure. *Journal of Quality Technology* 16(2): 97-100.

AUTHOR BIOGRAPHIES

E. JACK CHEN is a Senior Staff Specialist with BASF Corporation. He received an M.S. in computer science from Syracuse University, an M.B.A. from Northern Kentucky University, and a Ph.D. in Quantitative Analysis from the University of Cincinnati. His research interests are in the area of computer simulation, statistical data analysis and stochastic processes. His email and web addresses are <chenj@econqa.cba.uc.edu> and <http://www.econqa.cba.uc.edu/~chenj>.

W. DAVID KELTON is a Professor in the Department of Quantitative Analysis and Operations Management at the University of Cincinnati. He received a B.A. in mathematics from the University of Wisconsin-Madison, an M.S. in mathematics from Ohio University, and M.S. and Ph.D. degrees in industrial engineering from Wisconsin. His research interests and publications are in the probabilistic and statistical aspects of simulation, applications of simulation, and stochastic models. He is co-author of *Simulation Modeling and Analysis* (3d ed., 2000, with Averill M. Law), and *Simulation With Arena* (1998, with Randall P. Sadowski and Deborah A. Sadowski), both published by McGraw-Hill. Currently, he serves as Editor-in-Chief of the *INFORMS Journal on Computing*, and has been Simulation Area Editor for *Operations Research*, the *INFORMS Journal on Computing*, and *IIE Transactions*, as well as Associate Editor for *Operations Research*, the *Journal of Manufacturing Systems*, and *Simulation*. From

1991 to 1999 he was the INFORMS co-representative to the Winter Simulation Conference Board of Directors and was Board Chair for 1998. In 1987 he was Program Chair for the WSC, and in 1991 was General Chair. His email and web addresses are <david.kelton@uc.edu> and <http://www.econqa.cba.uc.edu/~keltond>.