

INDIFFERENCE-ZONE SUBSET SELECTION PROCEDURES: USING SAMPLE MEANS TO IMPROVE EFFICIENCY

E. Jack Chen

BASF Corporation
333 Mount Hope Avenue
Rockaway, New Jersey 07866, U.S.A.

ABSTRACT

Two-stage selection procedures have been widely studied and applied in determining the required sample size (i.e., the number of replications or batches) for selecting the best of k systems and for selecting a subset. The *Enhanced Two-Stage Selection* procedure is a heuristic two-stage selection procedure that takes into account not only the variance of samples, but also the difference of sample means when determining the sample sizes. This paper discusses the use of the same technique to select a subset of size m that contains at least c of the v best of k systems. Numerical experiments indicate that the proposed sample size allocation strategy is superior to other methods in the literature.

1 INTRODUCTION

Discrete-event simulation has been widely used to compare alternative system designs or operating policies. When evaluating k alternative system designs, we select one or more systems as the best and control the probability that the selected systems really are the best. Let μ_i denote the expected response of system i . Our goal is to find the system with the smallest expected response $\mu^* = \min_{1 \leq i \leq k} \mu_i$. If the system with the largest expected response is desired, just replace min with max in the formula. We achieve this goal by using a class of ranking and selection (R&S) procedures. However, efficiency is still a key concern for using simulation to solve R&S problems.

Many R&S procedures are directly or indirectly developed based on Dudewicz and Dalal (1975) or Rinott's (1978) indifference-zone selection procedures and have focused on identifying the best system. Nevertheless, Koenig and Law (1985) have developed a two-stage indifference-zone procedure to select a subset of size m that contains the v best of k systems; where $(1 \leq v \leq m < k)$. Sullivan and Wilson (1989)

have developed a two-stage subset selection procedure that determines a subset of maximum size m containing at least one of the systems that are within a pre-specified amount of the best. Let μ_{i_l} be the l^{th} smallest of the μ_i 's, so that $\mu_{i_1} \leq \mu_{i_2} \leq \dots \leq \mu_{i_k}$. These indifference-zone selection procedures determine the number of additional replications based on a conservative *least favorable configuration* (LFC, i.e., assuming $\mu_{i_1} + d^* = \dots = \mu_{i_v} + d^* = \mu_{i_{v+1}} = \dots = \mu_{i_k}$ and do not take into account the value of sample means; see Section 2.2. Some new approaches including (Chen et al. 2000) and (Chen and Kelton 2000, 2005) incorporate first-stage sample mean information in determining the number of additional replications. In an average case analysis, these procedures are more efficient in allocating sample sizes than procedures that are derived based on the LFC. There are several new approaches aiming to improve the efficiency of R&S procedures; Chick and Inoue (2001) use a Bayesian framework for constructing ranking and selection procedures. Kim and Nelson (2001) eliminates inferior systems based on whether a partial sum is within a continuation region. For an overview of existing methods of R&S see Bechhofer et al. (1995), Swisher et al. (2003) or Branke et al. (2005).

Chen et al. (2006) point out that R&S procedures have been incorporated with other simulation procedures to make statistically valid inferences (e.g., Buchholz and Thümmler 2005) and consequently the need to provide a subset of the v best systems (i.e., the top- v systems); instead of only the best system. They point out that the overall efficiency of certain types of simulation-optimization algorithms (e.g., evolutionary population-based search) depends highly on the quality of the selected top- v systems. If the selected subset contains poor systems, the convergence rate of simulation-optimization procedures (that invoke the R&S process) can be negatively affected.

Mahamunulu (1976) considers a generalized version of selection problem. The goal is to select a subset of size

m that contains at least c of the v best of k systems, where $\max(1, m + v + 1 - k) \leq c \leq \min(m, v)$ and $\max(m, v) \leq k - 1$. If $c = v = m = 1$, then the problem is to choose the best system. When $c = v < m$, we are interested in choosing a subset of size m that contains the v best; Koenig and Law (1985) have addressed this case. If $1 < c = v = m$, we are interested in choosing the m best systems, i.e., the case considered by Chen et al. (2006). We propose applying the techniques of Chen and Kelton (2000, 2005) to the generalized selection procedure.

The paper is organized as follows. In Section 2, we provide the background necessary to understand our proposed procedures. In Section 3, we present our methodologies and proposed procedures for selecting a subset of size m that contains at least c of the v best of k systems. In Section 4, we show our empirical-experiment results. In Section 5, we give concluding remarks.

2 BACKGROUND

In this section, we introduce the necessary notation and background:

- X_{ij} : the independent and normally distributed observations from the j^{th} replication or batch of the i^{th} system,
- N_i : the total number of replications or batches for system i ,
- n_i : the intermediate number of replications or batches for system i ,
- μ_i : the expected performance measure for system i , i.e., $\mu_i = E(X_{ij})$,
- $\bar{X}_i(n_i)$: the sample mean performance measure for system i with n_i samples, i.e., $\sum_{j=1}^{n_i} X_{ij}/n_i$,
- \bar{X}_i : the sample mean performance measure for system i shorthand for $\bar{X}_i(n_i)$,
- σ_i^2 : the variance of the observed performance measure of system i from one replication or batch, i.e., $\sigma_i^2 = \text{Var}(X_{ij})$,
- $S_i^2(n_i)$: the sample variance of system i with n_i replications or batches, i.e., $S_i^2(n_i) = \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2 / (n_i - 1)$.

2.1 Indifference-Zone Selection Procedures

Our goal is to select a system with the smallest expected response μ_{i_1} . Let CS denote the event of ‘‘correct selection.’’ In a stochastic simulation, a CS can never be guaranteed with certainty. The probability of CS, denoted by $P(\text{CS})$, is a random variable depending on sample sizes and other uncontrollable factors. Moreover, in practice, if the difference between μ_{i_1} and μ_{i_2} is very

small, we might not care if we mistakenly choose system i_2 , whose expected response is μ_{i_2} . The ‘‘practically significant’’ difference d^* (a positive real number) between the best and a satisfactory system is called the indifference zone in the statistical literature, and it represents the smallest difference about which we care. Therefore, we want a procedure that avoids making a large number of replications or batches to resolve differences less than d^* . That means we want $P(\text{CS}) \geq P^*$ provided that $\mu_{i_{v+1}} - \mu_{i_v} \geq d^*$, where the minimal CS probability P^* and the ‘‘indifference’’ amount d^* are both specified by the user.

2.2 Two-Stage Selection Procedures

The two-stage procedure of Dudewicz and Dalal (1975) to select the best of k systems has been widely studied and applied. Let n_0 be the number of initial replications or batches. The first-stage sample means

$$\bar{X}_i^{(1)} = \frac{1}{n_0} \sum_{j=1}^{n_0} X_{ij},$$

and marginal sample variances

$$S_i^2(n_0) = \frac{\sum_{j=1}^{n_0} (X_{ij} - \bar{X}_i^{(1)})^2}{n_0 - 1},$$

for $i = 1, 2, \dots, k$ are computed. Based on the number of initial replications or batches n_0 and the sample variance estimate $S_i^2(n_0)$ obtained from the first stage, the number of additional simulation replications or batches for each system in the second stage is $N_i - n_0$, where

$$N_i = \max(n_0 + 1, \lceil (h_1 S_i(n_0)/d^*)^2 \rceil), \text{ for } i = 1, 2, \dots, k. \quad (1)$$

Here $\lceil z \rceil$ is the smallest integer that is greater than or equal to the real number z and h_1 (which depends on k , P^* , and n_0) is a constant that can be found from the tables in Law and Kelton (2000).

The derivation of the procedure is based on that to select the best of k systems

$$P(\text{CS}) \geq \int_{-\infty}^{\infty} [F(t + h_1)]^{k-1} f(t) dt$$

and we equate the right-hand side to P^* to solve for h_1 . Here f and F , respectively, denote the probability density function (pdf) and the cumulative distribution function (cdf) of the t distribution with $n_0 - 1$ degrees of freedom (d.f.).

We then compute the second-stage sample means

$$\bar{X}_i^{(2)} = \frac{1}{N_i - n_0} \sum_{j=n_0+1}^{N_i} X_{ij}.$$

Define the weights

$$W_{i1} = \frac{n_0}{N_i} \left[1 + \sqrt{1 - \frac{N_i}{n_0} \left(1 - \frac{(N_i - n_0)(d^*)^2}{h_1^2 S_i^2(n_0)} \right)} \right]$$

and $W_{i2} = 1 - W_{i1}$, for $i = 1, 2, \dots, k$. Compute the weighted sample means

$$\tilde{X}_i = W_{i1} \bar{X}_i^{(1)} + W_{i2} \bar{X}_i^{(2)}$$

and select the system with the smallest \tilde{X}_i . Note that the expression for W_{i1} was chosen to make $(\tilde{X}_i - \mu_i)/(d^*/h)$ have a t distribution with $n_0 - 1$ d.f. (see Dudewicz and Dala 1975). Koenig and Law (1985) extended the procedure to select a subset of size m that contains the v best of k systems. They show that

$$P(\text{CS}) \geq (k - m) \binom{k - v}{m - v} \times \int_{-\infty}^{\infty} [F(t + h)]^v [F(t)]^{m-v} [F(-t)]^{k-m-1} f(t) dt. \quad (2)$$

See Section 3.1 for an explanation of this formula. The required sample sizes are computed by (1) except that different critical constant h is used. We equate the right-hand side to P^* and solve for the constant h (depends on k, v, m, n_0 , and P^*), which can be found from the tables in Koenig and Law (1985) or Law and Kelton (2000).

The derivation of these procedures are based on the LFC. However, in reality, we rarely encounter the LFC; therefore, these procedures are consequently conservative. Basically, the computing budget is allocated proportionally to the estimated sample variances.

2.3 An Enhanced Two-Stage Selection (ETSS) Procedure

Based on Rinott's (1978) procedure, Chen and Kelton (2000) propose an ETSS procedure that takes into account not only the sample variances, but also the difference of sample means across systems when determining the sample sizes. The ETSS is derived with the assumption that the true means are known; however, the true means are estimated by sample means in practice. Thus, the ETSS procedure is a heuristic approach and

does not guarantee $P(\text{CS}) \geq P^*$. Chen (2004) proposes using a conservative adjustment to increase $P(\text{CS})$.

Let h_r be the critical constant that solves Rinott's integral; let $d_i = \max(d^*, \mu_i - \mu_{i_1})$ and

$$N_i = \max(n_0, \lceil (h_r S_i(n_0)/d_i)^2 \rceil), \quad i = 1, 2, \dots, k. \quad (3)$$

Then the allocated number of required simulation replications or batches for each system are large enough to achieve the precision guarantee. Since the true means are unknown, d_i 's need to be estimated by the sample means. Let \bar{X}_{b_l} be the l^{th} smallest of the sample means \bar{X}_i 's, so that $\bar{X}_{b_1} \leq \bar{X}_{b_2} \leq \dots \leq \bar{X}_{b_k}$. Then

$$\hat{d}_i = \max(d^*, \bar{X}_i - \bar{X}_{b_1}). \quad (4)$$

The ETSS procedure computes the number of required simulation replications or batches for each system using

$$N_i = \max(n_0, \lceil (h_r S_i(n_0)/\hat{d}_i)^2 \rceil), \quad i = 1, 2, \dots, k. \quad (5)$$

Besides the difference in the critical constant h_r , the difference between (5) and (1) is that \hat{d}_i is being used instead of d^* . This makes sense when our objective is to achieve $P[\bar{X}_{b_l} < \bar{X}_{i_l}, \text{ for } l = 2, 3, \dots, k] \geq P^*$, i.e., to find the good systems. If the objective of the simulation experiments is to estimate the differences of the expected responses, we can use different experimental designs and different procedures to obtain more precise estimates. The differences in the sample means are embedded in \hat{d}_i ; consequently, this procedure will allocate fewer replications or batches to the less promising system i , whose sample mean $\bar{X}_i \gg \bar{X}_{b_1}$.

Even though under the LFC the observed $P(\text{CS})$ of the ETSS procedure is likely to be less than the minimal CS probability P^* , the ETSS procedure can significantly reduce the computational effort in more typical cases that are far from the LFC. Furthermore, in many cases the consequence of making a mistake is not too costly when the chosen one is close to $\mu_{i_1} + d^*$.

3 METHODOLOGIES

In this section we present the rationale of applying the technique of ETSS to the generalized subset selection procedure. We use a new approach to derive the lower bound of subset selection $P(\text{CS})$. Like other selection procedures, the proposed procedures assume input data are independent and identically distributed (i.i.d.) normal and allow unknown and unequal variances across systems. If non-normality of the input data is a concern, users can use batch means (see Law and Kelton 2000, Chen and Kelton 2007) to obtain sample means that are essentially i.i.d. normal.

3.1 Assessing the Probability of Correct Selection

Let w_{i_l} denote the half-width of the one-tailed $(P^*)^{1/(k-1)}$ confidence interval (c.i.) of $d_{i_l} = \mu_{i_l} - \mu_{i_1}$, for $l = 2, 3, \dots, k$. Chen (2004) points out that procedures developed based on the LFC achieve $w_{i_l} < d^*$ and consequently the one-tailed $1 - \alpha$ c.i. of d_{i_l} $CI1 = (\hat{d}_{i_l} - d^*, \infty]$. Whereas procedures that take into account sample means attempt to achieve $w_{i_l} < d_{i_l}$ and $CI2 = (\hat{d}_{i_l} - d_{i_l}, \infty] \approx (0, \infty]$. Hence, the allocated sample sizes are just large enough to conclude $\mu_{i_1} < \mu_{i_l}$ (provided $\mu_{i_1} + d^* \leq \mu_{i_l}$) with a desired confidence but no more than necessary. That is, to obtain $P[\bar{X}_{i_1} < \bar{X}_{i_l}] \geq 1 - \alpha$, the sample size should be large enough so that the one-tailed $1 - \alpha$ c.i. half-width $w_{i_l} < d_{i_l}$. In practice, d_{i_l} needs to be estimated by (4). It is straightforward to use \hat{d}_i instead of d^* in (1) to compute the required sample size when the goal is to select the best.

The derivation of (2) is based on the fact that for $i = 1, 2, \dots, k$,

$$T_i = \frac{\tilde{X}_i(N_i) - \mu_i}{d^*/h}$$

has a t distribution with $n_0 - 1$ d.f., where h depends on c, v, m, k, n_0 , and P^* . Note that T_i 's are independent. If f and F , respectively, are the pdf and the cdf of the random variable T . It is known that the distribution of the u^{th} order statistic of n observations of T is

$$g_u(t_u) = \beta(F(t_u); u, n - u + 1)f(t_u),$$

where

$$\beta(x; a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}$$

is the pdf of the beta distribution with shape parameters a and b ; see Hogg and Craig (1995 p. 198). Note that $\Gamma(a) = (a-1)!$ for any positive integer a . In our case, f and F are functions of the t distribution with $n_0 - 1$ d.f.

We now assess the P(CS) of this selection problem. Let Y_c be the c^{th} smallest sample mean from $\tilde{X}_{i_l}(N_{i_l})$ for $l = 1, 2, \dots, v$ and let μ_c be its unknown true mean. Let Y_d be the d^{th} ($d = m - c + 1$) smallest sample mean from $\tilde{X}_{i_l}(N_{i_l})$ for $l = v + 1, v + 2, \dots, k$ and let μ_d be its unknown true mean. Then

$$\begin{aligned} P(\text{CS}) &= P[Y_c < Y_d] \\ &= P\left[\frac{\tilde{X}_{i_c}(N_{i_c}) - \mu_{i_c}}{d^*/h} \leq \frac{\tilde{X}_{i_{v+d}}(N_{i_{v+d}}) - \mu_{i_{v+d}}}{d^*/h} + \frac{\mu_{i_{v+d}} - \mu_{i_c}}{d^*/h}\right] \end{aligned}$$

$$\begin{aligned} &= P\left[T_{i_c} \leq T_{i_{v+d}} + \frac{\mu_{i_{v+d}} - \mu_{i_c}}{d^*/h}\right] \\ &\geq P[T_{i_c} \leq T_{i_{v+d}} + h]. \end{aligned}$$

The inequality follows since $\mu_{i_{v+d}} - \mu_{i_c} = \mu_{i_{m+v-c+1}} - \mu_{i_c} \geq \mu_{i_{v+1}} - \mu_{i_v} \geq d^*$. Furthermore, if $d_{i_{v+d}}$ is used instead of d^* in the above equations, then we obtain strict equality.

Note that $T_{i_c} \sim \beta(F(T_{i_c}); c, v - c + 1)f(T_{i_c})$; and $T_{i_{v+d}} \sim \beta(F(T_{i_{v+d}}); m - c + 1, k - v - m + c)f(T_{i_{v+d}})$. Hence,

$$\begin{aligned} P(\text{CS}) &\geq \int_{-\infty}^{\infty} \int_{-\infty}^{x+h} \beta(F(y); c, v - c + 1)f(y) \beta(F(x); m - c + 1, k - v - m + c)f(x) dy dx \\ &= \frac{v!}{(c-1)!(v-c)!} \frac{(k-v)!}{(m-c)!(k-v-m+c-1)!} \times \\ &\int_{-\infty}^{\infty} \int_{-\infty}^{x+h} [F(y)]^{c-1} [F(-y)]^{v-c} f(y) [F(x)]^{m-c} [F(-x)]^{k-v-m+c-1} f(x) dy dx. \end{aligned}$$

We equate the right-hand side to P^* to solve for h . In the case that $c = v$, the above equation is simplified to (2). See Appendix for the bound of the probability of correct selection when the goal is to select the largest μ_i 's.

Under the LFC, the inequality in above equations becomes strict equality and $P(\text{CS}) = P^*$. In the non-LFC, the allocated sample sizes should be large enough to distinguish systems i_v and i_{m+1} . Let

$$d_{i_l} = \begin{cases} \mu_{i_{m+1}} - \mu_{i_l} & 1 \leq l \leq v \\ \mu_{i_l} - \mu_{i_v} & v + 1 \leq l \leq k. \end{cases}$$

The sample size

$$N_{i_l} = \max(n_0 + 1, \lceil (hS_{i_l}(n_0)/d_{i_l})^2 \rceil), \text{ for } l = 1, 2, \dots, k.$$

Then the allocated sample sizes are large enough to provide the precision guarantee.

Again, since the true means are unknown, d_{i_l} 's need to be estimated by the sample means. Moreover, we incorporate the indifference amount in determining the required sample sizes. Let

$$d_{b_l} = \begin{cases} \max(d^*, \bar{X}_{b_{m+1}} - \bar{X}_{b_l}) & 1 \leq l \leq v \\ \max(d^*, \bar{X}_{b_l} - \bar{X}_{b_v}) & v + 1 \leq l \leq k. \end{cases}$$

The required sample sizes are estimated by

$$N_{b_l} = \max(n_0 + 1, \lceil (hS_{b_l}(n_0)/d_{b_l})^2 \rceil), l = 1, \dots, k. \quad (6)$$

We denote this the ETSS procedure.

Furthermore, conservative adjustments can be used to increase the P(CS). Let $U(\bar{X}_{b_v})$ and $L(\bar{X}_{b_{m+1}})$ be the upper and lower P^* confidence limits of μ_{b_v} and $\mu_{b_{m+1}}$, respectively. Then

$$d_{b_l} = \begin{cases} \max(d^*, L(\bar{X}_{b_{m+1}}) - \bar{X}_{b_l}) & 1 \leq l \leq v \\ \max(d^*, \bar{X}_{b_l} - U(\bar{X}_{b_v})) & v + 1 \leq l \leq k. \end{cases} \quad (7)$$

We denote this the AT (Adjusted ETSS) procedure.

3.2 Discussion

Chen et al. (2006) develop a procedure (denoted OCBA- m) to select a subset of size m that contains the m best systems, i.e., the optimal subset of size m , with a fixed sample size and they aim to maximize the P(CS). In their setting, they do not incorporate the indifference-zone strategy. Our previous empirical studies indicate that when the goal is to maximize the P(CS) with a fixed sample size, the sample size allocation strategy that takes into account the indifference amount does not perform better.

The problem of maximizing the P(CS) given a sample size is the dual of minimizing the sample size given the P(CS). Consequently, the solutions to both problems are the same. To adapt the proposed procedure to maximum P(CS) instead of minimizing the sample size, we can set the indifference amount to $\bar{X}_{b_{v+1}} - \bar{X}_{b_v}$. Note that in this case the indifference amount is a random variable. Then

$$d_{b_l} = \begin{cases} \bar{X}_{b_{m+1}} - \bar{X}_{b_l} & 1 \leq l \leq v \\ \bar{X}_{b_l} - \bar{X}_{b_v} & v + 1 \leq l \leq k. \end{cases}$$

Following the strategy of Chen et al. (2006), we allocate the sample sizes for each system such that

$$\frac{N_i}{N_j} = \left(\frac{S_i(n_0)/d_i}{S_j(n_0)/d_j} \right)^2, i, j \in \{1, 2, \dots, k\}, \text{ and } i \neq j. \quad (8)$$

Note that in OCBA- m , $d_i = \bar{X}_i - c$, where $c = (\bar{X}_{b_v} + \bar{X}_{b_{v+1}})/2$.

3.3 Sequential Procedure of Selecting a Subset

We now present a cost-effective sequential approach to select a subset of size m that contains at least c of the v best systems from k alternatives. We denote this SRS (Sequential Ranking and Selection) procedure.

Generic Sequential Ranking and Selection Procedure:

1. Let $N_{i,t}$ be the sample size allocated for system i and $\bar{X}_{i,t}$ be the sample mean of system i at the t^{th} iteration. Simulate n_0 samples for all systems. Set the iteration index $t = 0$, and $N_{1,t} = N_{2,t} = \dots = N_{k,t} = n_0$. Specify the value of the indifference amount d^* and the required precision P^* .
2. Calculate sample means and sample variances. Rank the sample means and obtain the index b_l for $l = 1, 2, \dots, k$.
3. Calculate the new sample size $N_{b_l,t+1} = \max(n_0 + 1, \lceil (hS_{b_l}(N_{b_l,t})/d_{b_l})^2 \rceil)$, for $l = 1, 2, \dots, k$. Here d_{b_l} is computed according to (7).
4. If $N_{i,t+1} \leq N_{i,t}$, for $i = 1, 2, \dots, k$, go to step 6.
5. Simulate additional $\delta_{i,t+1} = \lceil (N_{i,t+1} - N_{i,t})^+ / 2 \rceil$ samples for all systems. Here $(X)^+ = \max(X, 0)$. Set $N_{i,t+1} = N_{i,t} + \delta_{i,t+1}$ and $t = t + 1$. Go to step 2.
6. Return the values b_l and $\bar{X}_{b_l}(N_{b_l})$, for $l = 1, 2, \dots, m$.

We can reduce the number of iterations with a larger incremental sample size for system i at the t^{th} iteration, but we run the risk of allocating more samples than necessary to non-promising systems. We propose to compute the incremental sample size dynamically with all the information obtained up to the current iteration. The additional sample size for alternative i at iteration $t + 1$ is

$$\delta_{i,t+1} = \lceil (N_{i,t+1} - N_{i,t})^+ / 2 \rceil. \quad (9)$$

We use the equation $S_i^2(r) = (\sum_j^r X_{ij}^2 / r - \bar{X}_i^2) r / (r - 1)$ to compute the variance estimator, so we are only required to store the triple $(N_{i,t}, \sum_{t=1}^{N_{i,t}} X_{it}, \sum_{t=1}^{N_{i,t}} X_{it}^2)$ instead of the entire sequences $(X_{i1}, X_{i2}, \dots, X_{iN_{i,t}})$.

The SRS procedure is able to estimate the required sample size for each system to obtain the specified P^* based on information obtained up to the current stage, so we are able to allocate incremental sample sizes intelligently. The procedure allocates a large incremental sample size at the first iteration and reduces the incremental sample sizes approximately by half at later iterations; see (9). This makes sense because there should be room for more aggressive budget allocation at early iterations, which reduces the number of iterations and the overhead in computing sample means and sample variances. Furthermore, we don't run the risk of allocating more samples than necessary because we have estimated the required sample size for each system. As

the procedure proceeds, the incremental sample size allocation become less aggressive to avoid allocating more samples than necessary to non-promising systems.

Furthermore, the factor of 1/2 in (9) insures that the allocated sample sizes satisfy approximately the ratios of (8). Assuming $(hS_i(n_0)/d_i)^2 - n_0 > 0$ at the first iteration, then the sample size for each alternative i $N_{i,1}$ is a little more than half of N_i obtained from (6). Therefore, the ratios of (8) should be roughly the same when N_i is replaced by $N_{i,1}$. On the other hand, if the procedure stops when the subset contains only the estimated m best systems, then sample size $N_{i,t}$ should be close to N_i . Thus, the ratios should also be roughly the same. Consequently, the required ratios of sample size for each alternative to maximize P(CS) is maintained approximately at all iterations. The critical value h depends on c, v, m, k, n_0 , and P^* . Even though the sample sizes for each system will change at each iteration, we can use the initial value of h through all iterations. This simplifies the programming effort and provides conservative estimates of the sample sizes.

3.4 Selecting an Optimal Subset Given a Fixed Computing Budget

We also implement the procedure to select the m best systems given a fixed computing budget. The procedure is basically the same as OCBA- m except that the allocated sample sizes for each system are different. Furthermore, the incremental sample size at each iteration is computed dynamically. A smaller incremental sample size generally has better performance in terms of P(CS), but will result in a larger number of iterations and consequently longer runtime. Let T be the allocated total number of simulation replications (budget) and let $\Omega_{t+1} = T - \sum_{i=1}^k N_{i,t}$ be the available computing budget at the $(t+1)^{th}$ iteration. We set the incremental computing budget at iteration $t+1$

$$\Delta_{t+1} = \min(\Omega_{t+1}, \max([\Omega_{t+1}/2], 5k)).$$

Sequential Optimal Subset Selection Procedure:

1. Let $N_{i,t}$ be the sample size allocated for system i and $\bar{X}_{i,t}$ be the sample mean of system i at the t^{th} iteration. Simulate n_0 samples for all systems. Set the iteration index $t = 0$, and $N_{1,t} = N_{2,t} = \dots = N_{k,t} = n_0$. Specify the allowed total sample size T .
2. Calculate sample means and sample variances. Rank the sample means and obtain the index b_l for $l = 1, 2, \dots, k$.
3. If $\sum_{i=1}^k N_{i,t} = T$, go to step 6.

4. Increase the sample sizes by Δ_{t+1} and calculate the new sample sizes, $N_{1,t+1}, N_{2,t+1}, \dots, N_{k,t+1}$ according to (8).
5. Simulate additional $\delta_{i,t+1} = (N_{i,t+1} - N_{i,t})^+$ samples for all systems. Set $t = t + 1$. Go to step 2.
6. Return the values b_l and $\bar{X}_{b_l}(N_{b_l})$, for $l = 1, 2, \dots, m$.

4 EMPIRICAL EXPERIMENTS

In this section we present some empirical results obtained from simulations using the KL (Koenig and Law 1985), ETSS, AT (ETSS with a conservative adjustment), and SRS (sequentialized AT) procedures. In order to compare our procedures with other known procedures, we use the similar experiments of Chen et al. (2006). Furthermore, we select the m ($= c = v$) best systems based on the sample means (i.e., \bar{X}_i 's) instead of the weighted sample means (i.e., \tilde{X}_i 's).

4.1 Experiment 1: Equal Variances

There are ten alternative systems under consideration. Suppose $X_{ij} \sim \mathcal{N}(i, 6^2)$, $i = 1, 2, \dots, 10$, where $\mathcal{N}(\mu, \sigma^2)$ denotes the normal distribution with mean μ and variance σ^2 . We want to select three systems with the smallest means, i.e., $m = c = v = 3$. It is obvious that systems 1, 2, and 3 are the best three systems. The indifference amount d^* is set to 1.0 in all cases. We experimented with two initial sample sizes $n_0 = 20$ and 40. We compare the actual P(CS) of KL, ETSS, AT, and SRS procedures.

Furthermore, 100,000 independent experiments are performed to obtain the actual P(CS). The number of times we successfully selected the true three best systems (systems 1, 2, and 3 in this example) is counted among the 100,000 independent experiments. P(CS), the correct selection proportion, is then obtained by dividing this number by 100,000.

Table 1 lists the results of Experiment 1. The $P(CS)$ column lists the proportion of correct selection. The T column lists the average of the total simulation replications ($T = \sum_{R=1}^{100000} \sum_{i=1}^k N_{R,i} / 100000$, $N_{R,i}$ is the total number of replications or batches for system i in the R^{th} independent run) used in each procedure. The $KL(20)$, $ETSS(20)$, $AT(20)$, and $SRS(20)$ rows list the results of the procedures with initial replications $n_0 = 20$. Note that the observed P(CS)'s are all higher than the specified $P^* = 0.90$ and $P^* = 0.95$. AT has better coverage than the ETSS procedure with a larger total number of replications. The SRS has better coverage than AT with a smaller total number of replications. Because the variance of the sample mean is larger with

Table 1: P(CS) and Sample Sizes of Experiment 1.

| P^* Procedure | 0.90 | | 0.95 | |
|--------------------|--------|------|--------|------|
| | P(CS) | T | P(CS) | T |
| KL(20) | .99243 | 4955 | .99692 | 6273 |
| ETSS(20) | .94425 | 1445 | .95691 | 1812 |
| AT(20) | .98676 | 2704 | .99264 | 3411 |
| SRS(20) | .99560 | 2078 | .99844 | 2819 |
| KL(40) | .99249 | 4615 | .99652 | 5771 |
| ETSS(40) | .97295 | 1394 | .96980 | 2634 |
| AT(40) | .99107 | 2254 | .99544 | 2788 |
| SRS(40) | .99435 | 1860 | .99777 | 2439 |

Table 2: P(CS) and Sample Sizes of Experiment 2.

| P^* Procedure | 0.90 | | 0.95 | |
|--------------------|--------|------|--------|------|
| | P(CS) | T | P(CS) | T |
| KL(20) | .99288 | 5307 | .99654 | 6712 |
| ETSS(20) | .98196 | 808 | .98811 | 1006 |
| AT(20) | .99095 | 1277 | .99622 | 1803 |
| SRS(20) | .99473 | 988 | .99808 | 1335 |
| KL(40) | .99264 | 4964 | .99708 | 6194 |
| ETSS(40) | .99163 | 747 | .99535 | 882 |
| AT(40) | .99263 | 992 | .99660 | 1309 |
| SRS(40) | .99399 | 873 | .99737 | 1102 |

a smaller initial sample size n_0 , the adjustment yields more improvement in P(CS) when n_0 is small.

4.2 Experiment 2: Increasing Variances

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

The results are listed in Table 2. Because good systems have smaller variances than systems in Experiment 1, the total simulation replications are smaller than Experiment 1 for ETSS and its extensions. In this setting we are more confident of the best selections at the first stage. Therefore, less simulation replications are needed to obtain the desired confidence. For KL's procedure, the simulation replications allocation is based entirely on the variances; thus, $N_i > N_j$ when $S_i(n_0) > S_j(n_0)$. Consequently, KL allocates more simulation replications than Experiment 1. All other procedures take into consideration the difference of sample means; therefore, some $N_i < N_j$ even though $S_i(n_0) > S_j(n_0)$. Procedures that take into account the sample means have the most significant reduction in the number of replications or batches in this setting, i.e., the inferior alternatives have the largest variances. All observed P(CS)'s are greater than the specified nominal level.

4.3 Experiment 3: Decreasing Variances

This is another variation of Experiment 1. All settings are preserved except that the variance of each system decreases as the mean increases. Namely, $X_{ij} \sim \mathcal{N}(i, (11 - i)^2)$, $i = 1, 2, \dots, 10$.

The results are listed in Table 3. Because good systems have larger variances than Experiment 1, the total simulation replications are larger than Experiment 1. In this setting we have less confidence of the good selection at the first stage. Therefore, more simulation replications are needed to obtain the desired confidence. All procedures allocate smaller additional simulation replications for systems with inferior systems in this setting, i.e., the variances decrease as the sample means increase. Once again, AT has better coverage than ETSS with additional replications or batches. Since inferior systems have smaller variances, we are confident to exclude those systems from further simulations at early iterations.

4.4 Experiment 4: the LFC

In this experiment, we test the procedure under the LFC since the minimum P(CS) occurs in this setting. Namely, $X_{ij} \sim \mathcal{N}(0, 6^2)$, $i = 1, 2, 3$; $X_{ij} \sim \mathcal{N}(d^*, 6^2)$, $i = 4, 5, \dots, 10$. The indifference amount is set to 1.

The results are listed in Table 4. As expected, the observed P(CS)'s of ETSS are smaller than the specified P^* under the LFC, especially with smaller n_0 . The conservative adjustment significantly increases the P(CS)'s; however, they are still slightly smaller than the nominal values. Three of the observed P(CS)'s of KL are also smaller than the nominal values. We believe this is because of the stochastic nature of the experiments. The SRS performs well even under the LFC; achieves the highest observed P(CS)'s while allocates less samples than KL.

Table 3: P(CS) and Sample Sizes of Experiment 3.

| P^* Procedure | 0.90 | | 0.95 | |
|--------------------|--------|------|--------|------|
| | P(CS) | T | P(CS) | T |
| KL(20) | .99217 | 5306 | .99654 | 6711 |
| ETSS(20) | .90940 | 2101 | .92840 | 2626 |
| AT(20) | .98113 | 3955 | .98768 | 4992 |
| SRS(20) | .99521 | 3145 | .99831 | 4177 |
| KL(40) | .99288 | 4965 | .99706 | 6192 |
| ETSS(40) | .95481 | 2124 | .96497 | 2605 |
| AT(40) | .98805 | 3536 | .99345 | 4381 |
| SRS(40) | .99431 | 2892 | .99771 | 3780 |

Table 4: P(CS) and Sample Sizes of Experiment 4.

| P^* Procedure | 0.90 | | 0.95 | |
|--------------------|--------|------|--------|------|
| | P(CS) | T | P(CS) | T |
| KL(20) | .89906 | 4955 | .95036 | 6273 |
| ETSS(20) | .63686 | 3111 | .70969 | 3931 |
| AT(20) | .87156 | 4626 | .93952 | 6057 |
| SRS(20) | .93286 | 4890 | .97304 | 6232 |
| KL(40) | .89870 | 4615 | .94939 | 5771 |
| ETSS(40) | .78479 | 3430 | .84791 | 4283 |
| AT(40) | .89223 | 4427 | .94744 | 5654 |
| SRS(40) | .91453 | 4550 | .96060 | 5730 |

4.5 Experiment 5: Maximizing the P(CS)

We compare the performance of the SRS procedure with OCBA- m to select the top- m systems with a fixed computing budget. The three settings we tested are as defined in experiments 1 through 3 and we vary the computing budget T between 200 and 7000 for these two procedures. In order to correctly compare these procedures' performance only in the sample-sizes allocation strategy. The incremental sample size of SRS at each iteration is set to 50 instead of using the dynamic-incremental-sample-size rule described in Section 3.4. The detail of the results are not listed here. However, in general SRS performs better than OCBA- m . In the increasing variance setting, OCBA- m performs slightly better than SRS; if there is a difference in the performance. On the other hand, SRS performs significantly better in the other two settings.

5 CONCLUSIONS

Many two-stage indifference-zone selection procedures ignore a large amount of first-stage sampling information. The ETSS procedure utilizes both the means and variances from the first stage. Hence, the marginal computational effort required for the ETSS procedure is minimized, yet the achieved efficiency improvement is significant. Even though ETSS is a heuristic procedure, derived with the assumption that the true means are known, it does have strong basis.

From our experiments, we notice that ETSS often makes a wrong selection when a non-best alternative has the smallest sample mean at the first stage. We recommend using a conservative adjustment, which is computed dynamically according to the variance of the best alternative at the first stage, to ETSS to increase P(CS). The adjustment effectively allocates additional replications or batches to more promising alternatives. We have more confidence in the mean estimator when its variance is small resulting in a smaller adjustment.

On the other hand, we have less confidence in the mean estimator when its variance is large resulting in a larger adjustment.

Moreover, since the quality of the first-stage sample means have great influence in the performance of the ETSS procedure, we recommend using a larger first-stage sample size for the ETSS procedure. Note that with the same minimal required probability of correct selection P^* , the total sample sizes using $n_0 = 40$ may be less than using $n_0 = 20$ and achieve higher P(CS) at the same time. One drawback of two-stage selection procedures is that they rely heavily on the information from only one stage. To eliminate this drawback, we have sequentialized the AT procedure.

Our experimental results show that ETSS and its extensions are powerful tools for selecting a subset of size m that contains at least c of the v best systems out of k alternatives. The main advantage of the procedures is that the algorithm determines the number of additional simulation replications based on both the means and variances, which significantly improves the efficiency of R&S procedures. We strongly recommend using the sequentialized version of the selection procedure since it performs better than two-stage procedures in terms of sample size and probability of correct selection. The simplicity of this method should make it attractive to simulation practitioners or software developers.

A APPENDIX

When the goal is to select the largest μ_i 's, the P(CS) is equal to the probability that the $(m - c + 1)^{\text{th}}$ largest of $(\bar{X}_{i_1}, \bar{X}_{i_2}, \dots, \bar{X}_{i_{k-v}})$ is less than the c^{th} largest of $(\bar{X}_{i_{k-v+1}}, \bar{X}_{i_{k-v+2}}, \dots, \bar{X}_{i_k})$. This is equal to the probability that the $(k - v - m + c)^{\text{th}}$ smallest of $(\bar{X}_{i_1}, \bar{X}_{i_2}, \dots, \bar{X}_{i_{k-v}})$ is less than the $(v - c + 1)^{\text{th}}$ smallest of $(\bar{X}_{i_{k-v+1}}, \bar{X}_{i_{k-v+2}}, \dots, \bar{X}_{i_k})$.

It is helpful to regard $c' = k - v - m + c$, $v' = k - v$, and $m' = k - m$. This is consistent with the fact that the graphs of the pdf of $\beta(x; a, b)$ and $\beta(x; b, a)$ are mirror images of each other. Hence,

$$\begin{aligned}
 & \text{P(CS)} \\
 & \geq \int_{-\infty}^{\infty} \int_{-\infty}^{x+h} \beta(F(y); k - v - m + c, m - c + 1) \times \\
 & \quad f(y)\beta(F(x); v - c + 1, c)f(x)dydx \\
 & = \frac{(k - v)!}{(k - v - m + c - 1)!(m - c)!} \frac{v!}{(v - c)!(c - 1)!} \times \\
 & \quad \int_{-\infty}^{\infty} \int_{-\infty}^{x+h} [F(y)]^{k-v-m+c-1} [F(-y)]^{m-c} f(y) \\
 & \quad [F(x)]^{v-c} [F(-x)]^{c-1} f(x)dydx.
 \end{aligned}$$

Furthermore,

$$d_{i_l} = \begin{cases} \mu_{k-v+1} - \mu_{i_l} & 1 \leq l \leq k-v \\ \mu_{i_l} - \mu_{i_{k-m}} & k-v+1 \leq l \leq k. \end{cases}$$

REFERENCES

- Bechhofer, R. E., T. J. Santner, and D. M. Goldsman. 1995. *Design and Analysis of Experiments for Statistical Selection, Screening and Multiple Comparisons*. New York: John Wiley & Sons, Inc.
- Branke, J., S. Chick, and C. Schmidt. 2005. Selecting a Selection Procedure. Submitted to *Management Science*.
- Buchholz, P. and A. Thümmler. 2005. Enhancing Evolutionary Algorithms with Statistical Selection Procedures for Simulation Optimization. *Proceedings of the 2005 Winter Simulation Conference*, ed. M. E. Kuhl, N. M. Steiger, F. B. Armstrong, and J. A. Joines, 842-852. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.
- Chen, C. H., D. He, and M. Fu. 2006. Efficient Simulation Budget Allocation for Selecting an Optimal Subset. Submitted to *INFORMS Journal on Computing*.
- Chen, C. H., J. Lin, E. Yücesan, and S. E. Chick. 2000. Simulation Budget Allocation for Further Enhancing the Efficiency of Ordinal Optimization. *Journal of Discrete Event Dynamic Systems* 10(3): 251-270.
- Chen, E. J. 2004. Using Ordinal Optimization Approach to Improve Efficiency of Selection Procedures. *Journal of Discrete Event Dynamic Systems* 14(2): 153-170.
- Chen, E. J. and W. D. Kelton. 2000. An Enhanced Two-Stage Selection Procedure. *Proceedings of the 2000 Winter Simulation Conference*, ed. J.A. Joines, R. Barton, P. Fishwick, and K. Kang, 727-735. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.
- Chen, E. J. and W. D. Kelton. 2005. Sequential Selection Procedures: Using Sample Means to Improve Efficiency. *European Journal of Operational Research* 166(1): 133-153.
- Chen, E. J. and W. D. Kelton. 2007. Confidence-Interval Estimation Using Quasi-Independent Sequences. *IIE Transactions*. Forthcoming.
- Chick, S. E. and K. Inoue. 2001. New Two-Stage and Sequential Procedures for Selecting the Best Simulated System. *Operations Research* 49: 1609-1624.
- Dudewicz, E. J. and S. R. Dalal. 1975. Allocation of Observations in Ranking and Selection with Unequal Variances. *Sankhya* B37: 28-78.
- Hogg, R. V. and A. T. Craig. 1995. *Introduction to Mathematical Statistics*. 5th ed. New Jersey: Prentice Hall.
- Kim, S.-H. and B. L. Nelson. 2001. A Fully Sequential Procedure for Indifference-Zone Selection in Simulation. *ACM TOMACS* 11(3): 251-273.
- 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. *Communications in Statistics - Simulation and Communication* B14: 719-734.
- Law, A. M. and W. D. Kelton. 2000. *Simulation Modeling and Analysis*. 3rd ed. New York: McGraw-Hill.
- Mahamunulu, D. M. 1967. Some Fixed-Sample Ranking and Selection Problems. *Ann. Math. Statist* 38: 1079-1091.
- Rinott, Y. 1978. On Two-stage Selection Procedures and Related Probability Inequalities. *Communications in Statistics* A7: 799-811.
- Sullivan, D. W. and J. R. Wilson. 1989. Restricted Subset Selection Procedures for Simulation. *Operations Research* 37: 52-71.
- Swisher, J. R., S. H. Jacobson, and E. Yücesan. 2003. Discrete-Event Simulation Optimization using Ranking, Selection, and Multiple Comparison Procedures: A Survey. *ACM Transactions on Modeling and Computer Simulation* 13(2): 134-154.

AUTHOR BIOGRAPHY

E. JACK CHEN is a Senior Staff Specialist with BASF Corporation. He received a Ph.D. degree from University of Cincinnati. His research interests are in the area of computer simulation. His email address is e.jack.chen@basf.com.