Proceedings of the 2008 Winter Simulation Conference
S. J. Mason, R. R. Hill, L. Mönch, O. Rose, T. Jefferson, J. W. Fowler eds.

RESTRICTED SUBSET SELECTION

E. Jack Chen

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

ABSTRACT

This paper develops procedures for selecting a set of normal populations with unknown means and unknown variances in order that the final subset of selected populations satisfies the following requirements: with probability at least P^*, the selected subset will contain a population or "only and all" of those populations whose mean lies less than the distance d^* from the smallest mean. The size of the selected subset is random, however, at most m populations will finally be chosen. A restricted subset attempts to exclude populations that are deviated more than d^* from the smallest mean. Here P^*, d^*, and m are users specified parameters. The procedure can be used when the unknown variances across populations are unequal. An experimental performance evaluation demonstrates the validity and efficiency of these restricted subset selection procedures.

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 $\mu^* = \min_{1 \le i \le k} \mu_i$. If the system with the biggest 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. Most R&S procedures, e.g. Dudewicz and Dalal (1975) or Rinott's (1978) indifference-zone selection procedures, have focused on identifying the best system. Nevertheless, Koening and Law (1985) have developed a two-stage indifference-zone procedure to select a subset of size m containing the v best of k systems; where ($1 \le v \le m < k$). For a general overview of R&S, see Bechhofer et al. (1995) or Swisher et al. (2003).

Chen et al. (2008) 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 the need to provide a subset of the m best systems; instead of only the best system. They indicate 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-m systems, i.e. the selected m best systems. If the selected subset contains poor systems, the convergence rate of simulation-optimization procedures (that invoking the R&S process) can be negatively affected. Consequently, they extend OCBA (Optimal Computing Budget Allocation, Chen et al. 2000) to maximize the probability of correctly selecting the top-m systems with a given computing budget, denoted OCBA-m.

Another approach of selecting a subset of good systems is to select a restricted subset, which attempts to exclude populations that are deviated more than d^* from the best. The size of the selected subset is random, however, at most m populations will finally be chosen. The goal is to provide a statistical guarantee of correctly selecting a restricted subset and has no computing-budget constraint. Extending the work of Gupta and Santner (1973), Sullivan and Wilson (1989) have developed a two-stage restricted subset selection procedure, denoted V_E, that determines a subset of maximum size m that contains at least one system that is within a pre-specified amount of the best. In this paper, we use different approaches to derive restricted subset selection procedures. Furthermore, the proposed restricted subset selection procedures have the capability to determine a subset of maximum seize m that contains only and all of the m system(s) that are within a pre-specified amount of the best.

Selection procedures that are developed based on the least favorable configuration (LFC, see Section 2.3) are conservative and become inefficient when the number of systems is large. Newer approaches, e.g. Chen et al. (2000) and Chen (2004), take into account the difference of sample means, which can significantly increase the efficiency of

978-1-4244-2708-6/08/$25.00 ©2008 IEEE

selection procedures. Hence, we incorporate this approach into subset-selection procedures as well.

The paper is organized as follows. In Section 2, we provide the background necessary to understand the proposed procedure. In Section 3, we present our methodologies and the proposed procedures for selecting a subset containing up to m d^*-near-best systems, see Section 2.1. In Section 4, we show our empirical-experiment results. In Section 5, we give concluding remarks.

2 BACKGROUND

First, some notations:

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

Let μ_{i_l} be the l^{th} smallest of the μ_i's, so that $\mu_{i_1} \le \mu_{i_2} \le \ldots \le \mu_{i_k}$. Our goal is to select a subset of systems with the smallest expected response $\mu_{i_l} \le \mu_{i_1} + d^*$. Let CS denote the event of "correct selection." In a stochastic simulation, a CS can never be guaranteed with certainty. The possibility of CS, denoted by P(CS), is a random variable depending on sample sizes and becomes higher as the sample sizes increase. Moreover, in practice, if the difference between μ_{i_1} and μ_j is very small, we might not care if we mistakenly choose system j, whose expected response is $\mu_j > \mu_{i_1}$. 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(CS) $\ge P^*$ provided that $\mu_j - \mu_{i_1} \ge d^*$, where the minimal CS probability P^* and the "indifference" amount d^* are both specified by the users.

Configurations satisfying $\mu_j - \mu_{i_1} \ge d^*$ are said to be in the *preference zone* for a correct selection; configurations satisfying $\mu_j - \mu_{i_1} < d^*$ are said to be in the indifference zone. Formally, we say a system i is *d^*-near-best* or *a good design* if μ_i is within a specified amount d^* of the smallest mean.

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 the 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 sample variances

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

for $i = 1, 2, \ldots, k$ are computed. Based on the number of the 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

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

where $\lceil z \rceil$ is the smallest integer that is greater than or equal to the real number z, and h_1 (which depends on k, n_0, and P^*) is a constant that can be found from the tables in Koenig and Law (1985) or Law and Kelton (2000).

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

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

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 a t distribution with $n_0 - 1$ degrees of freedom (df). 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 $i = 1, 2, \ldots, k$. 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 $(\tilde{X}_i - \mu_i)/(d^*/h_1)$ have a t distribution with $n_0 - 1$ df (see Dudewicz and Dala 1975).

2.3 Selecting a Subset

In this section, we derive the required sample sizes for selecting a subset of size m that contains the $v(\leq m)$ best of k systems. Note that both m and v are users specified parameters. Under the LFC $\mu_{i_1} = \mu_{i_2} = \ldots = \mu_{i_v}$ and $\mu_{i_v} + d^* = \mu_{i_{v+1}} = \ldots = \mu_k$, the procedure guarantees with probability P^* that the selected subset of size m containing systems i_l for $l = 1, 2, \ldots, v$.

Let T be a random variable with pdf f and cdf F. Hogg and Craig (1995, p. 198) show that the distribution of the u^{th} order statistics of m observations of T is

$$g_{m,u}(t_u) = \beta(F(t_u); u, m-u+1) f(t_u),$$

where

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

is the beta distribution with shape parameters a and b and $\Gamma(a)$ is the gamma function. Note that $\Gamma(a) = (a-1)!$ for any positive integer a.

Let

$$G_{m,i}(t_i) = \int_{-\infty}^{t_i} g_{m,i}(t)dt.$$

Then

$$g_{m,m}(t_m) = m[F(t_m)]^{m-1} f(t_m) \text{ and } G_{m,m}(t_m) = [F(t_m)]^m.$$

Let $\tilde{X}_v$ be the largest weighted sample mean from $\tilde{X}_{i_l}$ for $l = 1, 2, \ldots, v$ and let μ_v be its unknown true mean. Let $\tilde{X}_u$ be the u^{th} $(u = m - v + 1)$ smallest weighted sample mean from $\tilde{X}_{i_l}$ for $l = v+1, v+2, \ldots, k$ and let μ_u be its unknown true mean. We can write

$$\begin{aligned}
& \text{P(CS)} \\
= \; & \text{P}[\tilde{X}_v < \tilde{X}_u] \\
= \; & \text{P}\left[\frac{\tilde{X}_v - \mu_v}{d^*/h} \leq \frac{\tilde{X}_u - \mu_u}{d^*/h} + \frac{\mu_u - \mu_v}{d^*/h}\right] \\
= \; & \text{P}\left[T_v \leq T_u + \frac{\mu_u - \mu_v}{d^*/h}\right] \\
= \; & \int_{-\infty}^{\infty} G_{v,v}\left(t_u + \frac{\mu_u - \mu_v}{d^*/h}\right) g_{k-v,m-v+1}(t_u)dt_u \\
\geq \; & \int_{-\infty}^{\infty} [F(t_u + h)]^v g_{k-v,m-v+1}(t_u)dt_u \\
= \; & \frac{(k-v)!}{(m-v)!(k-m-1)!} \\
& \int_{-\infty}^{\infty} [F(t+h)]^v [F(t)]^{m-v} [F(-t)]^{k-m-1} f(t)dt.
\end{aligned}$$

In this case, f and F are functions of the t distribution with $n_0 - 1$ df. The fourth equality follows since t_{i_l} for $l = 1, 2, \ldots, v$ are independent. The inequality follows since $\mu_u - \mu_v \geq d^*$ under the indifference-zone approach. We equate the right-hand side to P^* and solve for h (which depends on k, m, v, n_0, and P^*). Since the value of h is determined such that under the LFC $\text{P}[T_v \leq T_u + h] = P^*$. Let $\Upsilon = T_v - T_u$, then $\text{P}[\Upsilon \leq h] = P^*$. That is, under the LFC the value of h is the P^* quantile of the distribution of Υ. See Koenig and Law (1985) and Chen (2007) for more information on subset selection.

Hence, the required sample sizes for selecting a subset of size m containing the v best systems are

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

After simulating N_i samples for systems $i = 1, 2, \ldots, k$, the procedure returns the m systems that have the smallest weighted sample means. If $m = v = 1$, then the goal is to choose the best system. If $m > v = 1$, we are interested in choosing a subset of size m containing the best. If $m = v > 1$, we are interested in choosing the m best systems.

2.4 Effect of the Indifference Amount

In testing the null hypothesis $\text{H}_0 : \mu_u \leq \mu_i$, for us to reject the null hypothesis and conclude with confidence level $1 - \alpha$ that $\mu_u > \mu_i$ is the same as the lower endpoint of the one-tailed $1 - \alpha$ CI (confidence interval) is positive, i.e. $\bar{X}_u - \bar{X}_i - w_{iu} > 0$, where w_{iu} denotes the half-width of the one-tailed $1 - \alpha$ CI. For detail on duality of hypothesis test and CI, see Rice (1995). The half-width w_{iu} depends on the sample sizes and becomes smaller as the sample sizes become large. This implies the sample sizes (N_u and N_i) should be large enough so that $w_{iu} < \bar{X}_u - \bar{X}_i$. By symmetry of the normal distribution $\text{P}[\bar{X}_u - \bar{X}_i \geq (\mu_u - \mu_i) - w_{iu}] \geq 1 - \alpha$. To obtain $\text{P}[\bar{X}_u - \bar{X}_i > 0] \geq 1 - \alpha$, the sample size should be large enough so that $w_{iu} < d_{iu} = \mu_u - \mu_i$.

Let $\hat{d}_{iu} = \bar{X}_u - \bar{X}_i$. Procedures developed based on the LFC achieve $w_{iu} < d^*$ and consequently the one-tailed $1 - \alpha$ CI of d_{iu} $CI1 = (\hat{d}_{iu} - d^*, \infty]$. Whereas procedures that take into account sample means attempt to achieve $w_{iu} < d_{iu}$ and $CI2 = (\hat{d}_{iu} - d_{iu}, \infty] \approx (0, \infty]$. Hence, the allocated sample sizes are just large enough for us to conclude $\mu_i < \mu_u$ (provided $\mu_i + d^* \leq \mu_u$) with a desired confidence but no

more than necessary. The indifference amount in (1) and (2) corresponds to the targeted CI half width. If the targeted CI half-width is $d = \max(d^*, \mu_u - \mu_i)$, then d (instead of d^*) should be used in those equations. In practice, the true means μ_u and μ_i are unknown, we need to use the sample means $\bar{X}_u$ and $\bar{X}_i$ to estimate d. See Chen (2004) for detail.

To take into account the effect of the indifference amount, Chen (2007) proposes the following for subset selection. Sort the sample means such that $\bar{X}_{b_1} \le \bar{X}_{b_2} \le \ldots \le \bar{X}_{b_k}$. 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. Compute

$$d_{b_l} = \begin{cases} \max(d^*, L(\bar{X}_{b_{m+1}}) - \bar{X}_{b_l}) & 1 \le l \le v \\ \max(d^*, \bar{X}_{b_l} - U(\bar{X}_{b_v})) & v+1 \le l \le k. \end{cases} \quad (3)$$

Then the required sample sizes are estimated by

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

Furthermore, the problem of maximizing the P(CS) given a total 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 subset-selection 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

$$d_{b_l} = \begin{cases} \bar{X}_{b_{m+1}} - \bar{X}_{b_l} & 1 \le l \le v \\ \bar{X}_{b_l} - \bar{X}_{b_v} & v+1 \le l \le k. \end{cases} \quad (4)$$

Following the strategy of Chen et al. (2008), we allocate the sample sizes for each system such that

$$\frac{N_i}{N_j} = \left(\frac{S_i(n_i)/d_i}{S_j(n_j)/d_j} \right)^2, i, j \in \{1, 2, \ldots, k\}, \text{ and } i \ne j,$$

where n_i and n_j are the current sample sizes of systems i and j, respectively. Note that in OCBA-m, $d_i = \bar{X}_i - \mu_0$, where $\mu_0 = (\bar{X}_{b_v} + \bar{X}_{b_{v+1}})/2$. Our experimental results indicate that the sample sizes derived based on (4) generally perform better than based on $d_i = \bar{X}_i - \mu_0$.

3 METHODOLOGIES

In this section, we extend the subset selection procedure to select a restricted subset. As with most selection procedure, the proposed selection procedures require the input data to be independent and identically distributed (iid) normal. However, the variance can be different across systems. Many performance measures of interest are taken over some average of a sample path or a batch of samples. Thus, many applications tend to have a normally distributed simulation output. If the non-normality of the samples is a concern, users can use batch means to "manufacture" samples that appeared to be iid normal, as determined by the tests of independence and normality (see, for example, Chen and Kelton 2007). In the selection procedures described next, the sampling operations can be carried out independently across systems. Hence, one can deploy the selection procedure in a parallel and distributed environment.

3.1 Comparisons With a Control

We use the approach of Chen (2006) to derive the procedure. Suppose the sample sizes are N_i and N_c, respectively, for systems i and c, where system c is a control. The test at confidence level $1-\alpha$ of $H_0 : \mu_i < \mu_c + d^*$ against the alternative $H_1 : \mu_i \ge \mu_c + d^*$ is based on the test statistic

$$Z = \frac{\bar{X}_i - \bar{X}_c - d^*}{\sqrt{\sigma_i^2/N_i + \sigma_c^2/N_c}}.$$

The acceptance region for this test is $Z \le z_{1-\alpha}$, where $z_{1-\alpha}$ is the $1-\alpha$ quantile of the standard normal distribution. That is,

$$\bar{X}_i - \bar{X}_c - d^* \le z_{1-\alpha} \sqrt{\sigma_i^2/N_i + \sigma_c^2/N_c}.$$

Let Φ denote the standard normal cdf. If $\mu_i - \mu_c > d^*$, the probability of committing a Type II error, i.e. concluding that the null hypothesis is true when in fact it is false, is

$$\begin{aligned} \beta &= \mathrm{P}[\bar{X}_i - \bar{X}_c - d^* \le z_{1-\alpha} \sqrt{\sigma_i^2/N_i + \sigma_c^2/N_c}] \\ &\le \Phi(z_{1-\alpha}). \end{aligned}$$

The inequality holds because Z has a standard normal distribution when $\mu_i - \mu_c = d^*$.

To control both the Type I error, i.e. we don't accept the null hypothesis when it is true, and the Type II error, we will use the test statistic

$$Z = \frac{\bar{X}_i - \bar{X}_c}{\sqrt{\sigma_i^2/N_i + \sigma_c^2/N_c}}.$$

The acceptance region for this test is still $Z \le z_{1-\alpha}$ or

$$\bar{X}_i - \bar{X}_c \le z_{1-\alpha} \sqrt{\sigma_i^2/N_i + \sigma_c^2/N_c}.$$

Note that with this test statistics the probability of accepting the null hypothesis that $\mu_i < \mu_c + d^*$ will be less than $1-\alpha$ when $\mu_c < \mu_i < \mu_c + d^*$ and will decrease as μ_i deviates more from μ_c. If $\mu_i - \mu_c > d^*$, the probability that the test

statistic falls in the acceptance region is

$$\beta \leq \Phi(z_{1-\alpha} - \frac{d^*}{\sqrt{\sigma_i^2/N_i + \sigma_c^2/N_c}}).$$

For fixed d^* and α, β can be evaluated as a function of sample sizes N_c and N_i. For more detail, see Rice (1995). Suppose we want to limit the probability of β, the sample sizes N_c and N_i should be large enough such that

$$z_{1-\alpha} - \frac{d^*}{\sqrt{\sigma_i^2/N_i + \sigma_c^2/N_c}} \leq z_\beta.$$

Let $\alpha = \beta = 1 - P^*$. Then

$$2z_{1-\alpha} \leq \frac{d^*}{\sqrt{\sigma_i^2/N_i + \sigma_c^2/N_c}}.$$

Hence, the CI half-width

$$w'_{ci} = z_{1-\alpha}\sqrt{\sigma_i^2/N_i + \sigma_c^2/N_c} \leq d^*/2.$$

That is, to control both Type I and Type II errors, the allocated sample sizes should be large enough such that the $1-\alpha$ one-tailed CI half width is less than $H = d^*/2$ when the difference between the two systems is at least d^*. Recall that the indifference amount in (1) and (2) corresponding to the targeted CI half width. Consequently, the required sample sizes for selecting a restricted subset (up to size m) in a two-stage procedure are

$$N_i = \max(n_0 + 1, \lceil (hS_i(n_0)/H)^2 \rceil), \text{ for } i = 1, 2, \ldots, k. \quad (5)$$

Note that the constant h here is the same as in (2). We then generate additional $N_i - n_0$ samples for system i in the second stage and compute the weighted sample mean $\tilde{X}_i$. However, $H = d^*/2$ instead of d^* should be used when computing the weight W_{i1}. Select system b_l if and only if $\tilde{X}_{b_l} \leq \min(\tilde{X}_{b_m}, \tilde{X}_{b_1} + d^*/2)$. We denote this two-stage restricted subset selection procedure RSS.

3.2 The Rationale

In this section, we provide the rationale of using H instead of d^* in (5) to determined the required sample sizes of restricted subset selection.

If the constant h is obtained with n_0, v, m, k, and P^*, and under the LFC (i.e. $\mu_v + d^* = \mu_u$), then the sample sizes allocated by (2) ensures that

$$\begin{aligned} \text{P(CS)} &= \text{P}[\tilde{X}_v < \tilde{X}_u] \\ &= \text{P}[\tilde{X}_v - \tilde{X}_u - d^* < \mu_v - \mu_u] \\ &\geq \text{P}[\tilde{X}_v - \tilde{X}_u - w_{vu} < \mu_v - \mu_u] \\ &= P^* \end{aligned}$$

The inequality follows because we *infer* (from the arguments in Section 2.4) that the sample sizes allocated by (2) achieves the one-tailed P^* CI half width $w_{vu} \leq d^*$. The last equality follows because the property of CI half width. Note that if $w_{vu} > d^*$, then P(CS) $< P^*$. It is our conjecture that

$$w_{vu} = \frac{h}{\sqrt{2}}\sqrt{\frac{S_v^2(n_0)}{N_v} + \frac{S_u^2(n_0)}{N_u}}.$$

If we allocate the sample sizes by (5), then $w_{vu} \leq d^*/2$. Let $\tilde{X}_{b_1} (\leq \tilde{X}_v)$ denote the smallest weighted sample mean. Then

$$\begin{aligned} & \text{P}[\tilde{X}_{b_1} + d^*/2 < \tilde{X}_u] \\ \geq\ & \text{P}[\tilde{X}_v + d^*/2 < \tilde{X}_u] \\ \geq\ & \text{P}[\tilde{X}_v - w_{vu} + d^* < \tilde{X}_u] \\ \geq\ & \text{P}[\tilde{X}_v - \tilde{X}_u - w_{cu} < \mu_v - \mu_u] \\ =\ & P^* \end{aligned}$$

Hence,

$$\text{P}[\tilde{X}_u < \tilde{X}_{b_1} + d^*/2] \leq 1 - P^*.$$

That is, there is no more $1-P^*$ probability that system u (i.e. the system whose weighted sample means is the u^{th} smallest among system i_l for $l = v+1, v+2, \ldots, k$) will be included in the restricted subset.

Under the configuration that $\mu_{i_l} = \mu_{i_1} + d^*$ for $l = 2, 3, \ldots, k$ (i.e. the LFC and $v = 1$) and $v < m$, the probability of $\tilde{X}_{b_{m+1}} < \tilde{X}_{b_1} + d^*/2$ will be less than $1 - P^*$. However, the probability of $\tilde{X}_{b_l} < \tilde{X}_{b_1} + d^*/2$ for $l = 2, 3, \ldots, m$ will be greater than $1 - P^*$. Note that under this configuration the RSS procedure will allocate more samples than those allocated by the V_E procedure of Sullivan and Wilson (1989). Under this configuration, the subset determined by V_E will on average contain $(m-1)/2$ non-d^*-near-best systems (Sullivan and Wilson 1989). Since precision become higher as sample sizes become larger and RSS allocates more samples than V_E for each system, the subset determined by RSS should on average contain no more than $(m-1)/2$ non-d^*-near-best systems under the same configuration. Under the LFC and $v = m$, the probability of $\tilde{X}_{b_{m+1}} < \tilde{X}_{b_1} + d^*/2$ will be less than $1 - P^*$. That is, there is at least P^* probability that all the selected systems in the restricted subset are d^*-near-best. However, the size of the restricted subset may be less than m, i.e. some of the d^*-near-best systems may not be selected.

Note that the value of v is an user specified parameter via the constant of h. As the value of v increases so will the value of h and the sample sizes; consequently, the average

number of the non-d^*-near-best systems included in the subset will decrease. In the special case that $v = m$, the P(CS) is the probability that only and all of the v d^*-near-best system(s) are included in the subset. For example, if $\mu_{i_1} + d^* < \mu_{i_2}$ and the constant h is obtained with $v = m(> 1)$ and P^*, then the procedure guarantees that the final subset contains only system i_1 with probability P^*. On the other hand, under the LFC (i.e. $\mu_{i_l} + d^* = \mu_{i_{v+1}}$, for $l = 1,2,\ldots,v$) the procedure guarantees that the final subset contains system i_l for $l = 1,2,\ldots,v$ with probability P^*.

3.3 Sequential Procedure of Selecting a Restricted Subset

We now present a cost-effective sequential approach to select a restricted subset of size up to m. We denote this SRSS (Sequential Restricted Subset Selection) procedure. To improve the efficiency of the procedure, we take into account the difference of sample means when computing the required sample sizes.

Sequential Restricted 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 number $t = 0$, and $N_{1,t} = N_{2,t} = \cdots = N_{k,t} = n_0$. Specify the value of the indifference amount d^* and the required precision P^*.
2. Calculate the sample means and sample variances. Rank the sample means such that $\bar{X}_{b_1} \leq \bar{X}_{b_2} \leq \ldots \leq \bar{X}_{b_k}$.
3. Calculate the required sample size $N_{b_l,t+1} = \max(n_0 + 1, \lceil (2hS_{b_l}(N_{b_l,t})/d_{b_l})^2 \rceil)$, for $l = 1,2,\ldots,k$. Here d_{b_l} is computed according to (3).
4. If $N_{i,t+1} \leq N_{i,t}$, for $i = 1,2,\ldots,k$, go to step 6.
5. Simulate additional $\lceil (N_{i,t+1} - N_{i,t})^+/2 \rceil$ samples for system i. Here $(x)^+ = \max(0,x)$. Set $t = t+1$. Go to step 2.
6. Select system b_l iff $\bar{X}_{b_l} \leq \min(\bar{X}_{b_m}, \bar{X}_{b_1} + d^*/2)$.

Note that sample means $\bar{X}_i$, instead of weighted sample means $\tilde{X}_i$, are used to determine the subset. This is because there are more than two stages of sample means and we can no longer use the approach of Dudewicz and Dalal (1975) to compute the weighted sample means. While $\bar{X}_i$ for $i = 1,2,\ldots,k$ are still t-distributed, they have different degrees of freedom. Hence, this sequential procedure is a heuristic. Nevertheless, our empirical studies indicate that this procedure performs well in terms of P(CS) and sample sizes. The critical value h depends on k, m, v, n_0, and P^*. Even though the sample sizes for each system change at each iteration, we use the initial value of h through all iterations. This simplifies the programming effort and provides conservative estimates of the sample sizes.

Let us consider the steps between taking additional samples, i.e. steps 2 through 5, be one iteration. 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. For example, two-stage procedures allocate all the required samples at the second stage based on the information obtained at the first stage. At the other extreme, some selection procedures, e.g. Kim and Nelson (2001) take only one additional sample from each system that is still under consideration at each iteration. While these procedures generally require smaller sample sizes, they require many iterations and the associated overhead.

We propose to use the sample size allocation strategy of Chen and Kelton (2005) 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

$$\delta_{i,t+1} = \lceil (N_{i,t+1} - N_{i,t})^+/2 \rceil. \quad (6)$$

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 $(X_{i1}, X_{i2}, \ldots, X_{iN_{i,t}})$.

The SRSS 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 (6). 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 becomes less aggressive to avoid allocating more samples than necessary to non-promising systems.

4 EMPIRICAL EXPERIMENTS

In this section, we present some empirical results of performing restricted subset selection using the RSS and SRSS procedures. We chose the first-stage sample size to be $n_0 = 20$. The number of systems under consideration is $k = 10$. The indifference amount d^* is set to 1. The targeted size of the subset m is set to 5. The minimal P(CS) of P^* is set to 0.90 and 0.95. Furthermore, we select the m

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).

We tested the following two configurations:

- Setting 1: $\mu_1 = 0$, and $\mu_j = d^*$, for $j = 2,3,\ldots,k$.
- Setting 2: $\mu_i = 0$, for $i = 1,2,\ldots,5$, and $\mu_j = d^*$, for $j = 6,7,\ldots,k$.

In setting 1, a correct selection means the selected subset contains system 1. In setting 2, a correct selection means none of the non-best systems (i.e, systems 6 through 10) are selected. Note that the size of the subset may be less than m. We did not perform any experiments of the non-LFC, since it has been shown that procedures that take into account the difference of sample means can significantly reduce the required sample sizes. Interested readers can see Chen (2007) for some experimental results.

The setting of variances is as follows.

- Equal Variances: $\sigma_i^2 = 6^2$, for $i = 1,2,\ldots,k$.
- Increasing Variances: $\sigma_i^2 = i^2$, for $i = 1,2,\ldots,k$.
- Decreasing Variances: $\sigma_i^2 = (11-i)^2$, for $i = 1,2,\ldots,k$.

We perform 10,000 independent experiments to obtain the actual P(CS). The number of times the selected subset contains the desired systems is counted among the 10,000 independent experiments. The observed correct selection proportion, $PCS1$, is then obtained by dividing this number by 10,000. We also compute $PCS2$, the proportion that the selected subset contains only and all of the v d^*-near-best system(s), i.e. in setting 1 it contains only system 1; and in setting 2 it contains systems 1 through 5. We list the results ($PCS1$, $PCS2$, the average final size of the subset $\hat{m}$, the average sample size of each simulation run T, i.e. $T = \sum_{R=1}^{10000}\sum_{i=1}^{k} N_{R,i}/10000$, $N_{R,i}$ is the total number of replications or batches for design i in the R^{th} independent run, and the standard error of T) for RSS and SRSS. Furthermore, the $Iter$ row lists the average number of iterations of SRSS.

Table 1 lists the results of setting 1 where system 1 is the only best system. The $v = 1$ and $v = 5$ columns list the results when the critical constant h is obtained with ($n_0 = 20$, $k = 10$, $m = 5$) $v = 1$ and $v = 5$, respectively. The allocated sample size of restricted subset selection is four times the size of unrestricted subset selection. Recall that $H = d^*/2$ is used in (5) whereas d^* is used in (2). When h is obtained with $v = 1$, i.e. the selected subset contains the best system, the average size of the subset is 2 and the observed $PCS1$'s are close to 1. The selected subset contains only system 1 (i.e. the best system) about 0.55 fraction of the times. When h is obtained with $v = 5$, i.e. the selected subset contains five d^*-near-best systems, the average size of the subset is close to 1. The procedures correctly determine that alternatives other than system 1 are not d^*-near-best. Furthermore, the observed $PCS1$'s are all 1, i.e. the selected subset always contains system 1. The subset contains only system 1 about 0.97 fraction of the times.

Table 2 lists the results of setting 2 where systems 1 through 5 are the best systems. The observed P(CS)'s are greater than the nominal values. When h is obtained with $v = 1$, the average size of the subset is 3.45. When h is obtained with $v = 5$, the average size of the subset is 4.95. The procedures incorrectly determine some of the best systems as non-d^*-near-best. When h is obtained with $v = 1$, the selected subset contains systems 1 through 5 about 0.21 (i.e. (0.1254+0.2888+0.1355+0.3128)/4) fraction of the times. On the other hand, when h is obtained with $v = 5$, the selected subset contains systems 1 through 5 about 0.94 (i.e. (0.9132+0.9570+0.9323+0.9749)/4) fraction of the times. The sequentialized procedure SRSS has better performance than the two-stage procedure RSS in terms of P(CS) and sample sizes. Moreover, the number of iterations initiated by the SRSS procedure is small.

Tables 3 and 4 list the experimental results of the increasing variances configuration. Tables 5 and 6 list the experimental results of the decreasing variances configuration. These results are generally similar to the experiments when the variances are equal.

5 CONCLUSIONS

We have presented two restricted-subset-selection procedures that provide effective means for screening a large set of systems. These procedures are versatile, easy to apply and can be incorporated with other procedures to select promising systems from large alternatives for a follow-up processing.

Selection procedures that are developed based on the LFC are conservative. Newer approaches utilize both the means and variances from earlier stages; the marginal computational effort is minimal, yet the achieved efficiency improvement is significant. The sequential procedure preserves the simple structure of indifference-zone selection while being more efficient in situations where many alternative systems are non-competitive. We strongly recommend users use the sequentialized version of the selection procedures since they performs better than two-stage procedures in terms of sample size and probability of correct selection.

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.

Buchholz, P. and A. Thümmler. 2005. Enhancing Evolutionary Algorithms with Statistical Selection Procedures for Simulation Optimization. In *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, M. Fu, and L. H. Lee. 2008. Efficient Simulation Budget Allocation for Selecting an Optimal Subset. *INFORMS Journal on Computing*. To Appear.

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. 2006. Comparison With a Standard Via All-Pairwise Comparisons. *Journal of Discrete Event Dynamic Systems* 16(3): 385-403.

Chen, E. J. 2007. Indifference-Zone Subset Selection Procedures: Using Sample Means to Improve Efficiency. *Proceedings of the 2007 Winter Simulation Conference*, ed. S. G. Henderson, B. Biller, M.-H. Hsieh, J. Shortle, J. D. Tew, and R. R. Barton, 535-543. 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. A Procedure for Generating Batch-Means Confidence Interval for Simulation: Checking Independence and Normality. *Simulation: Transactions of The Society for Modeling and Simulation International* 83: 683-694.

Dudewicz, E. J. and S. R. Dalal. 1975. Allocation of Observations in Ranking and Selection with Unequal Variances. *Sankhya* B37: 28–78.

Gupta, S. S. and T. J. Santner. 1973. On Selection and Ranking Procedures - a Restricted Subset Selection Rule. *Proceedings of the 39th Session of the International Statistical Institute*, 478-486.

Hogg, R. V., and A. T. Craig. 1995. *Introduction to Mathematical Statics*. 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.

Rice, J. A. 1995. *Mathematical Statistics and Data Analysis,* 2nd ed., Duxbury Press, Belmont, California.

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 BIOGRAPHIES

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>.

Table 1: Setting 1 and Equal Variances

		$P^*=0.90$		$P^*=0.95$	
		$v=1$	$v=5$	$v=1$	$v=5$
	$PCS1$	0.9813	1.0	0.9973	1.0
R	$PCS2$	0.4633	0.9580	0.5979	0.9795
S	$\hat{m}$	2.12	1.06	1.80	1.03
S	T	2967	20947	4986	26214
	$std(T)$	303	2163	514	2675
	$PCS1$	0.9866	1.0	0.9992	1.0
S	$PCS2$	0.4886	0.9753	0.6314	0.9912
R	$\hat{m}$	2.10	1.03	1.74	1.01
S	T	2664	20864	4749	26215
S	$std(T)$	310	248	351	261
	$Iter$	10	13	11	14

Table 2: Setting 2 and Equal Variances

		$P^*=0.90$		$P^*=0.95$	
		$v=1$	$v=5$	$v=1$	$v=5$
	$PCS1$	0.9331	1.0	0.9761	1.0
R	$PCS2$	0.1254	0.9132	0.2888	0.9570
S	$\hat{m}$	3.11	4.89	3.70	4.95
S	T	2967	20947	4986	26214
	$std(T)$	303	2163	514	2675
	$PCS1$	0.9456	1.0	0.9845	1.0
S	$PCS2$	0.1335	0.9323	0.3128	0.9749
R	$\hat{m}$	3.18	4.92	3.78	4.97
S	T	2570	20850	4633	26206
S	$std(T)$	265	264	321	266
	$Iter$	10	13	11	14

Table 3: Setting 1 and Increasing Variances

		$P^*=0.90$		$P^*=0.95$	
		$v=1$	$v=5$	$v=1$	$v=5$
	$PCS1$	0.9953	1.0	0.9998	1.0
R	$PCS2$	0.4656	0.9650	0.6160	0.9828
S	$\hat{m}$	1.94	1.04	1.67	1.02
S	T	3191	22346	5343	28109
	$std(T)$	425	3030	721	3770
	$PCS1$	0.9986	1.0	0.9999	1.0
S	$PCS2$	0.4814	0.9707	0.6374	0.9882
R	$\hat{m}$	1.93	1.04	1.66	1.02
S	T	2745	22349	4997	28050
S	$std(T)$	401	226	475	251
	Iteration	10	14	11	14

Table 4: Setting 2 and Increasing Variances

		$P^*=0.90$		$P^*=0.95$	
		$v=1$	$v=5$	$v=1$	$v=5$
	$PCS1$	0.9282	0.9999	0.9759	1.0
R	$PCS2$	0.1617	0.9124	0.3173	0.9566
S	$\hat{m}$	3.32	4.89	3.78	4.95
S	T	3191	22346	5343	28109
	$std(T)$	425	3030	721	3770
	$PCS1$	0.9456	1.0	0.9871	1.0
S	$PCS2$	0.1667	0.9256	0.3352	0.9729
R	$\hat{m}$	3.31	4.91	3.83	4.97
S	T	2533	22294	4708	28022
S	$std(T)$	452	318	582	296
	Iteration	10	14	11	14

Table 5: Setting 1 and Decreasing Variances

		$P^*=0.90$		$P^*=0.95$	
		$v=1$	$v=5$	$v=1$	$v=5$
	$PCS1$	0.9817	1.0	0.9975	1.0
R	$PCS2$	0.4833	0.9607	0.6068	0.9812
S	$\hat{m}$	2.08	1.05	1.79	1.02
S	T	3175	22389	5342	28027
	$std(T)$	426	3023	711	3782
	$PCS1$	0.9842	1.0	0.9993	1.0
S	$PCS2$	0.4871	0.9769	0.6340	0.9897
R	$\hat{m}$	2.10	1.03	1.73	1.01
S	T	2940	22251	5149	27989
S	$std(T)$	296	407	315	385
	Iteration	10	14	11	14

Table 6: Setting 2 and Decreasing Variances

		$P^*=0.90$		$P^*=0.95$	
		$v=1$	$v=5$	$v=1$	$v=5$
	$PCS1$	0.9462	0.9999	0.9762	1.0
R	$PCS2$	0.1331	0.9131	0.2813	0.9573
S	$\hat{m}$	3.14	4.89	3.67	4.95
S	T	3175	22389	5342	28027
	$std(T)$	426	3023	711	3782
	$PCS1$	0.9458	1.0	0.9831	1.0
S	$PCS1$	0.1336	0.9348	0.3087	0.9760
R	$\hat{m}$	3.18	4.91	3.76	4.97
S	T	3026	22294	5217	28027
S	$std(T)$	193	312	169	292
	Iteration	11	14	11	14