Proceedings of the 2002 Winter Simulation Conference
E. Yücesan, C.-H. Chen, J. L. Snowdon, and J. M. Charnes, eds.

A CONSERVATIVE ADJUSTMENT TO THE ETSS PROCEDURE

E. Jack Chen

BASF Corporation
3000 Continental Drive - North
Mount Olive, NJ 07828-1234, 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 designs. The *Enhanced Two-Stage Selection* (ETSS) 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 a conservative adjustment with the ETSS procedure to increase the probability of correct selection. We show how the adjustment allocates more simulation replications or batches to more promising designs at the second stage. An experimental performance evaluation demonstrates the efficiency of the adjustment.

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 design as the best and control the probability that the selected design really is the best. Let μ_i denote the expected response of design i. Our goal is to find the design with the smallest expected response $\mu^* = \min_{1\le i\le k} \mu_i$. If the design 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. 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. However, these indifference-zone selection procedures determine the number of additional replications based on a conservative *least favorable configuration* (LFC) assumption 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 (2000a) incorporate first-stage sample mean information in determining the number of additional replications. In an average case analysis, both procedures are more efficient in allocating sample sizes than Rinott's procedure. There are several new approaches aiming to improve the efficency of R&S procedures; Berger and Deely (1994), and Chick (1997) use a Bayesian framework for constructing ranking and selection procedures. For an overview of existing methods of R&S see Law and Kelton (2000) or Goldsman and Nelson (2001).

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 other uncontrollable factors. We propose adding a conservative adjustment to the ETSS procedure to increase the possibility of correct selection.

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

$S_i^2(N_i)$: the sample variance of design 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 $\mu_{i_1} \leq \mu_{i_2} \leq \ldots \leq \mu_{i_k}$. Our goal is to select a design with the smallest expected response μ_{i_1}. However, in practice, if the difference between μ_{i_1} and μ_{i_2} is very small, we might not care if we mistakenly choose design i_2, whose expected response is μ_{i_2}. The "practically significant" difference d^* (a positive real number) between the best and a satisfactory design 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) $\geq P^*$ 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 The Two-Stage Rinott Procedure

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

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

for $i = 1, 2, \ldots, k$ 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 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 where h (which depends on k, P^*, and n_0) is a constant which solves Rinott's (1978) integral (h can be calculated by the FORTRAN program *rinott* in Bechhofer et al. (1995), or can be found from the tables in Wilcox (1984) or Bechhofer et al. (1995)). We then compute the overall sample means $\bar{X}_i = \sum_{j=1}^{N_i} X_{ij}/N_i$, and select the design with the smallest $\bar{X}_i$. Basically, the computing budget is allocated proportionally to the estimated sample variances. Moreover, the derivation of this procedure is based on the LFC (i.e., assuming $\mu_{i_l} = \mu_{i_1} + d^*$, for all $l = 2, 3, \ldots, k$). However, in reality, we rarely encounter the LFC; therefore, this procedure is consequently conservative.

2.3 An Enhanced Two-Stage Selection (ETSS) Procedure

Chen and Kelton (2000a) propose an ETSS procedure that takes into account not only the sample variances, but also the difference of sample means across designs. The ETSS is derived with the assumption that we know the true means; however, the true means are estimated by sample means in practice. Thus, the ETSS procedure is a heuristic approach and does not guarantee P(CS) $\geq P^*$. The ETSS procedure uses fewer simulation replications or batches than Rinott's procedure. Moreover, while the observed P(CS)'s of the ETSS procedure are slightly lower than Rinott's procedure, they are still generally higher than the specified P^*.

We can improve the efficiency of R&S procedures with a *pre-selection*. 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. Chen (2001) points out that the ETSS procedure self-includes an intrinsic subset selection process, which means that a subset pre-selection does not actually need to be performed. That is, design i having the total required sample size $N_i = n_0$ is excluded from further simulation. In other words, based on n_0 samples from the first stage, the ETSS procedure has reached certain confidence on the probability $\Pr[\bar{X}_i > \bar{X}_b]$, for $i \neq b$, where $\bar{X}_b$ is the smallest of the $\bar{X}_i$'s in the current stage, i.e., $\bar{X}_b = \min_{1 \leq i \leq k} \bar{X}_i$. Thus, we remove the subset selection step from the ETSS procedure listed in Chen and Kelton (2000a). Let

$$\hat{d}_i = \max(d^*, \bar{X}_i - \bar{X}_b). \quad (2)$$

The ETSS procedure computes the number of required simulation replications or batches for each design based on the following formula

$$N_i = \max(n_0, \lceil (hS_i(n_0)/\hat{d}_i)^2 \rceil), \text{ for } i = 1, 2, \ldots, k. \quad (3)$$

In Chen and Kelton (2000a), they set $r_i = \hat{d}_i/d^*$, $h_i = h/r_i$, and

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

Although the sample sizes stay the same, we use equations (2) and (3) because they are simpler and easier to interpret.

The ETSS Algorithm:

1. Simulate n_0 replications.
2. For each design i, compute the needed additional replications $N_i - n_0$. Here N_i will be computed according to equation (3).

3. Simulate $N_i - n_0$ additional replications for each design i.
4. Return the values b and $\bar{X}_b$.

The difference between equation (3) and (1) is that $\hat{d}_i$ is being used instead of d^*. This makes sense when our objective is to achieve $\Pr[\bar{X}_{i_1} < \bar{X}_{i_l},\ \text{for } l = 2, 3, \ldots, k] \geq P^*$, i.e., to find the good designs. 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 design i, whose sample mean $\bar{X}_i >> \bar{X}_b$.

If $\bar{X}_{i_l} - \bar{X}_b \geq d^*$ and $N_{i_l} > n_0$, then for $l = 2, 3, \ldots, k$, the ratio

$$\frac{N_{i_l}}{N_{i_2}} = \left(\frac{d_{i_2}}{d_{i_l}}\right)^2 \left(\frac{S_{i_l}(n_0)}{S_{i_2}(n_0)}\right)^2 \tag{4}$$

is the same as that in the Optimal Computing Budget Allocation (OCBA) (Chen et al. 2000). On the other hand, the ratio

$$\frac{N_{i_1}}{N_{i_2}} = \left(\frac{d_{i_2}}{d^*}\right)^2 \left(\frac{S_{i_1}(n_0)}{S_{i_2}(n_0)}\right)^2 \tag{5}$$

is different from that of the OCBA. This is because the OCBA does not use the indifference parameter d^*.

3 METHODOLOGIES

In this section we present the basis of adding a conservative adjustment to the ETSS procedure to improve the probability of correct selection. As with most two-stage selection procedures, independent and identically distributed (i.i.d.) normal input data are required. If the input data are not i.i.d. normal, users can use batch means (see Chen and Kelton 2000b) to obtain sample means that are essentially i.i.d. normal.

3.1 Adjustment of the Difference of Sample Means

From our previous experiments, we notice that the first-stage sample mean of the best design is often not the smallest (i.e. $\bar{X}_{i_1} \neq \bar{X}_b$ or similarly $i_1 \neq b$) when the procedure makes an incorrect selection. This is because ETSS uses the smallest sample mean from the first stage as a control to compute simulation replications or batches for the second stage. The ETSS procedure uses $\bar{X}_b$, an estimator of μ_{i_1}, as a reference point to estimate the control distance. When the best design has an unusually large sample mean or a non-best design has an unusually small sample mean in the first stage, it often results in a smaller than necessary second stage sample size for the unknown best design in order to balance the unusual sample mean in the first stage. Consequently, the overall sample mean of the best design is not the smallest.

Since ETSS uses not only the variances but also the differences of sample means in the first stage, the ETSS procedure generally requires higher precision in the first stage than Rinott's procedure. Chen and Kelton (2000a) suggest using a constant $0 < c < 1$ so that $\hat{d}_i = \max(d^*, c(\bar{X}_{i_l} - \hat{\mu}_b))$, or $L(\bar{X}_i)$, the two-tailed lower $(1-\alpha)$ confidence limit of μ_i, to compute $d_i = \max(d^*, L(\bar{X}_i) - \hat{\mu}_b)$. The drawback of the first adjustment is that N_i's ($i \neq b$) are increased by the same ratio, and the second adjustment is somewhat cumbersome. We propose a new adjustment to the ETSS procedure that takes into consideration the randomness of $\bar{X}_b$ and allocates more replications or batches to more promising alternatives.

Let

$$d_i' = \max(d^*, \bar{X}_i - \bar{X}_b - ad^*), \tag{6}$$

where $a > 0$. The difference between equations (2) and (6) is that the difference in sample means from the first stage is adjusted with the amount ad^* in equation (6). Moreover,

$$N_i' = \max(n_0, \lceil (hS_i(n_0)/d_i')^2 \rceil), \ \text{ for } i = 1, 2, \ldots, k. \tag{7}$$

The difference among equations (1), (3), and (7) lies in the different values of d^*, d_i, and d_i' that are used respectively. With this adjustment, we take into account the randomness of $\bar{X}_b$ and allocate more simulation replications or batches to more promising designs.

If $\bar{X}_i = \bar{X}_b + nd^*$, then

$$\frac{d_i'}{d_i} = \begin{cases} 1 & 0 < n \leq 1 \\ 1/n & 1 < n \leq 1 + a \\ 1 - a/n & 1 + a < n. \end{cases}$$

Thus, if N_i' and $N_i > n_0$, then

$$\frac{N_i'}{N_i} = \begin{cases} 1 & 0 < n \leq 1 \\ n^2 & 1 < n \leq 1 + a \\ (n/(n-a))^2 & 1 + a < n. \end{cases}$$

Therefore, design i with the first stage sample mean $\bar{X}_i$, such that $d^* < \bar{X}_i - \bar{X}_b \leq (1+a)d^*$, will have significantly increased simulation replications or batches with this adjustment. The number of replications for that particular design is increased by $n^2 - 1$ times, where $n = (\bar{X}_i - \bar{X}_b)/d^*$. On the other hand, if $(1+a)d^* < \bar{X}_i - \bar{X}_b$, the extra simulation replications or batches allocated with this adjustment is very minimal. If $\bar{X}_i - \bar{X}_b \leq d^*$, then there are no changes in simulation replications or batches.

If a is a very large number (i.e., approaching ∞), then the values of d_i' will always be equal to d^*. The allocation of simulation replications or batches will be the same as the Rinott's procedure. Since the variance σ_b has great influence on the sample mean $\bar{X}_b$, we set $a = S_b(\bar{X}_b)/d^*$, where $S_b(\bar{X}_b) = S_b(n_0)/\sqrt{n_0}$ is the standard deviation of the sample mean of design b. That is, we will use $\bar{X}_b + S_b(\bar{X}_b)$ as the reference point. If the procedure is to find the maximum, then $\bar{X}_m - S_m(\bar{X}_m)$ should be used, where $\bar{X}_m = \max_{1\le i\le k} \bar{X}_i$. Even though this is somewhat arbitrary, we know that $\Pr[\mu_b \le \bar{X}_b + S_b(\bar{X}_b)] \approx .84$. We use this equation as an approximation and make the inference that $\Pr[\mu_{i_1} \le \bar{X}_b + S_b(\bar{X}_b)] \approx .84$. In this setting, $\hat{d}_i = \max(d^*, \bar{X}_i - U(\bar{X}_b))$, where $U(\bar{X}_b)$ is the upper one-tailed 0.84 confidence limit of μ_b.

We would like to point out that the purpose of R&S procedures is not to estimate μ_{i_1}, it is to select design i_1 such that $\mu_{i_1} = \min_{1\le i\le k} \mu_i$. However, since ETSS uses $\bar{X}_b$ as a reference point, we would prefer to have some confidence in $\bar{X}_b$. If the variance of the sample of the best alternative in the first stage is small, we will have more confidence in the mean estimator. Therefore, the adjustment made will be small. On the other hand, if the variance is large, we are less confident with this mean estimator; and consequently, the adjustment made will be large.

3.2 Some Intuition of Common Random Numbers

Let P_I(CS) denote the probability of correct selection with independent sampling, P_C(CS) denote P(CS) with Common Random Numbers (CRNs), event E_l, for $l = 2, 3, \ldots, k$, denote $\bar{X}_{i_l} - \bar{X}_{i_1} > 0$, $\Pr_I(E_l)$ and $\Pr_C(E_l)$ denote the probability of event E_l with independent sampling and with CRNS, respectively. With independent sampling across alternatives, E_l's are positively correlated and by Slepian's inequality (Tong 1980)

$$\begin{aligned} P_I(\mathrm{CS}) &= \Pr_I[E_2 \text{ and } E_3 \text{ and } \ldots \text{ and } E_k] \\ &= \Pr_I[E_2]\ldots\Pr_I[E_k|E_2, E_3, \ldots, E_{k-1}] \\ &\ge \prod_{l=2}^{k} \Pr_I[E_l]. \end{aligned}$$

The equality holds for $k = 2$, for $k > 2$ the equation holds with strict inequality.

Furthermore, it is known that in some cases

$$\Pr_C[E_2]\ldots\Pr_C[E_k|E_2, E_3, \ldots, E_{k-1}] \le \prod_{l=2}^{k} \Pr_C[E_l].$$

However,

$$\prod_{l=2}^{k} \Pr_I[E_l] \le \prod_{l=2}^{k} \Pr_C[E_l],$$

and it is possible that the following inequality still holds

$$\begin{aligned} &\Pr_C[E_2]\cdots\Pr_C[E_k|E_2, E_3, \ldots, E_{k-1}] \\ &\qquad \ge \Pr_I[E_2]\cdots\Pr_I[E_k|E_2, E_3, \ldots, E_{k-1}]. \end{aligned}$$

That is, $\Pr_C$(CS) is still larger than or equal to $\Pr_I$(CS).

The ETSS procedure uses the difference in sample means and the variance of samples to compute the required sample size for each design. When we use CRNs with the ETSS procedure, we reduce the variance of the difference of sample means and improve the precision of pairwise comparisons. Consequently, the P(CS) of the procedure may also be improved. Since the proposed procedure is based on the ETSS procedure, we recommend using CRNs with the adjusted ETSS procedure.

4 EMPIRICAL EXPERIMENTS

In this section we present some empirical results obtained from simulations using the Rinott, ETSS, and AT_a (ETSS with adjustment a) with $a = 0.5$, $S_b(\bar{X}_b)/d^*$, and 2. We also run experiments using CRNs with $a = S_b(\bar{X}_b)/d^*$.

4.1 Experiment 1 Equal Variances

There are ten alternative designs under consideration. 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 best design. The indifference amount d^* is set to 0.90 for all cases. We compare the actual P(CS) of Rinott's procedure, ETSS, and AT_a procedures. AT_S indicates $a = S_b(\bar{X}_b)/d^*$. AT_C uses CRNs with $a = S_b(\bar{X}_b)/d^*$. We use two different initial replications $n_0 = 20$, and 30. The variance of the best design is 6^2. Therefore, $S_b(\bar{X}_b)/d^* \approx (6/\sqrt{n_0})/0.9$, and when $n_0 = 20$ and 30, $a \approx 1.49$ and 1.22 respectively. Furthermore, 10,000 independent experiments are performed to obtain the actual P(CS). The number of times we successfully selected the true best design (design 1 in this example) is counted among the 10,000 independent experiments. P(CS), the correct selection percentage, is then obtained by dividing this number by 10,000.

Table 1 lists the results of Experiment 1. The *P(CS)* column lists the percentage of correct selection. The $\overline{T}$ column lists the average of the total simulation replications ($\overline{T} = \sum_{R=1}^{10000}\sum_{i=1}^{k} N_{R,i}/10000$, $N_{R,i}$ is the total number of replications or batches for design i in the R^{th} independent run) used in each procedure. The *Rinott(20), ETSS(20)* and *AT_a(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_a has better coverage than the ETSS procedure with a slightly larger total number of replications. Moreover, both

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

	$P^* = 0.90$		$P^* = 0.95$	
Procedure	P(CS)	$\overline{T}$	P(CS)	$\overline{T}$
Rinott(20)	99.34%	5256	99.74%	6669
ETSS(20)	95.08%	1197	95.73%	1502
$\mathrm{AT}_{.5}$(20)	96.63%	1391	97.63%	1763
AT_S(20)	98.07%	1736	98.79%	2174
AT_2(20)	98.81%	2047	99.51%	2578
AT_C(20)	100.00%	1937	100.00%	2441
Rinott(30)	99.42%	5001	99.76%	6316
ETSS(30)	96.89%	1234	97.66%	1497
$\mathrm{AT}_{.5}$(30)	97.90%	1405	98.60%	1738
AT_S(30)	98.68%	1645	99.42%	2040
AT_2(30)	99.23%	2037	99.65%	2556
AT_C(30)	100.00%	1773	100.00%	2199

the P(CS) and the total number of replications increase as a increases. Although the total number of simulation replications is significantly smaller, the observed P(CS)'s are very close to Rinott's procedure when $a = 2$.

Because the variance of the sample mean is larger with a smaller initial sample size n_0, the adjustment yields more improvement in P(CS) when n_0 is small. Furthermore, using CRNs with adjusted ETSS and two-stage selection procedures generally improves P(CS). With CRNs, the number of replications or batches is more than independent sampling across alternatives, and effectively allocating more samples to more promising alternatives since the frequency with which the best design has the smallest sample mean in the first stage is higher than independent sampling, i.e., $\Pr_C[\bar{X}_{i_1} = \bar{X}_b] \geq \Pr_I[\bar{X}_{i_1} = \bar{X}_b]$. When comparing the results of AT_S and AT_C, we increase the number of replications or batches only of the best three alternatives and significantly reduce the number of replications or batches of inferior alternatives.

Tables 2 and 3 list the detailed simulation replications allocated for each design under different selection procedures with $n_0 = 20$. We did not list the results of $n_0 = 30$ because they are similar. The *Rinott*, *ETSS* and AT_a columns list the average simulation replications for each design under the respective procedure. We would like to point out that Rinott's procedure will be the same as the equal allocation for additional simulation replications in this settings, i.e., the variances are equal for all designs. Our experimental results confirm this observation. On the other hand, in the ETSS and AT_a procedures, the number of additional simulation replications decreases as the differences $\hat{\delta}_{i,b} = \bar{X}_i - \bar{X}_b (> 0)$ increase. This makes sense because as $\hat{\delta}_{i,b}$ increases, it is more likely that $\bar{X}_i > \bar{X}_b$. In other words, as the observed difference of sample means across alternatives $\hat{\delta}_{i,b}$ increases, it is less likely that $\mu_i < \mu_b$.

The ratio of the average number of simulation replications allocated for design 10 and design 1 of Rinott's procedure is 1.0019 (527/526) when $P^* = 0.90$ and $n_0 = 20$, 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 0.0437 (20/458) under the ETSS procedure, see Table 2. This is where ETSS based procedures can significantly improve the efficiency of the Rinott procedure, i.e., the performance measure of inferior designs are far away from the best design. Note that 0.0437 is much larger than the theoretical value 0.01 ($(d^*/d_{i_{10}})^2 (S_{10}(n_0)/S_1(n_0))^2 = (0.9/9)^2$) because $n_0 > N_i$ for some design i. For example, if we use $n_0 = 5$ ($\geq 0.01 \times 458$), the ETSS procedure would have eliminated design 10. Moreover, the self-included intrinsic subset pre-selection of the ETSS procedure has better performance with larger n_0, i.e., inferior designs having $N_i = n_0$.

Table 2: Detailed Sample Sizes for $P^* = 0.90$ and $n_0 = 20$ of Experiment 1

Dn	Rinott	ETSS	$\mathrm{AT}_{.5}$	AT_S	AT_2	AT_C
1	526	458	485	506	516	528
2	527	304	354	426	466	528
3	526	175	230	315	387	497
4	526	96	127	206	283	194
5	522	48	67	117	174	68
6	524	30	37	62	94	35
7	522	22	25	34	51	24
8	526	20	21	25	30	20
9	526	20	20	21	22	20
10	527	20	20	20	20	20

Table 3: Detailed Sample Sizes for $P^* = 0.95$ and $n_0 = 20$ of Experiment 1

Dn	Rinott	ETSS	$\mathrm{AT}_{.5}$	AT_S	AT_2	AT_C
1	669	579	615	645	661	671
2	666	389	458	541	596	670
3	666	233	294	395	492	632
4	668	118	169	258	351	244
5	664	60	87	148	218	86
6	669	34	44	74	117	44
7	671	24	29	41	60	28
8	666	21	22	26	33	22
9	661	20	20	22	24	20
10	666	20	20	20	21	20

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, $X_{ij} \sim \mathcal{N}(i, (6 + (i-1)/2)^2)$, $i = 1, 2, \ldots, 10$.

The results are listed in Tables 4 through 6. Because most designs have larger variances than designs in Experiment 1, the total simulation replications are greater than Experiment 1. We are less confident of the best selection in the first stage of this setting. Therefore, more simulation replications are needed to obtain the desired confidence. All procedures allocate more additional simulation replications for designs because of larger variances. For Rinott's procedure, the simulation replications allocation is based entirely on the variances; thus, $N_i > N_j$ when $S_i(n_0) > S_j(n_0)$. The ETSS and AT$_a$ procedures take into consideration the difference of sample means; therefore, $N_i < N_j$ even though $S_i(n_0) > S_j(n_0)$. The ETSS procedure has 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. The results of using CRNs are similar with Experiment 1. Since inferior designs have larger variances, we are not confident to exclude those designs from further simulations. That is, based on n_0 samples from the first stage, the ETSS procedure cannot conclude that $\bar{X}_i > \bar{X}_b$, for $i \neq b$ with a desired confidence. Therefore, additional samples are required for those designs to increase the confidence.

Table 4: P(CS) and Sample Sizes of Experiment 2

	$P^* = 0.90$		$P^* = 0.95$	
Procedure	P(CS)	$\overline{T}$	P(CS)	$\overline{T}$
Rinott(20)	99.30%	10224	99.78%	13020
ETSS(20)	93.73%	1518	95.22%	1933
AT$_{.5}$(20)	95.89%	1835	97.19%	2302
AT$_S$(20)	97.96%	2420	98.60%	3069
AT$_2$(20)	98.50%	2922	99.18%	3697
AT$_C$(20)	100.00%	2507	100.00%	3185
Rinott(30)	99.47%	9757	99.74%	12280
ETSS(30)	96.39%	1489	97.15%	1836
AT$_{.5}$(30)	97.60%	1763	98.06%	2200
AT$_S$(30)	98.87%	2172	99.22%	2701
AT$_2$(30)	99.22%	2821	99.57%	3524
AT$_C$(30)	100.00%	2178	100.00%	2730

Table 5: Detailed Sample Sizes for $P^* = 0.90$ and $n_0 = 20$ of Experiment 2

Dn	Rinott	ETSS	AT$_{.5}$	AT$_S$	AT$_2$	AT$_C$
1	526	447	474	505	515	527
2	619	345	407	488	534	620
3	717	241	312	423	509	659
4	822	161	221	326	428	333
5	932	106	143	237	322	140
6	1053	69	94	163	228	77
7	1178	50	67	111	153	51
8	1315	37	46	71	106	38
9	1457	31	37	52	72	31
10	1601	26	30	40	51	27

Table 6: Detailed Sample Sizes for $P^* = 0.95$ and $n_0 = 20$ of Experiment 2

Dn	Rinott	ETSS	AT$_{.5}$	AT$_S$	AT$_2$	AT$_C$
1	664	565	599	641	658	670
2	785	444	510	624	683	787
3	911	313	390	540	643	837
4	1042	209	274	417	542	426
5	1181	136	187	301	401	180
6	1344	90	121	203	291	98
7	1500	63	82	133	195	65
8	1679	44	58	94	125	48
9	1861	35	42	62	90	38
10	2049	30	35	48	64	32

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, $X_{ij} \sim \mathcal{N}(i, (6-(i-1)/2)^2)$, $i = 1, 2, \ldots, 10$.

The results are listed in Tables 7 through 9. Because most designs have smaller variances than Experiment 1, the total simulation replications are smaller than Experiment 1. We have more confidence of the best selection in the first stage of this setting. Therefore, fewer simulation replications are needed to obtain the desired confidence. All procedures allocate smaller additional simulation replications for designs with inferior designs in this setting, i.e., the variances decrease as the sample means increase. Once again, AT$_a$ has better coverage than ETSS with little additional replications or batches. Moreover, with CRNs the procedure effectively allocates extra replications or batches to more promising alternatives and improves P(CS). Since inferior designs have smaller variances, we are confident to exclude those designs from further simulations, i.e., based on the first-stage information the ETSS procedure has concluded with certain confidence that $\bar{X}_i > \bar{X}_b$ for design i with $N_i = n_0$.

Table 7: P(CS) and Sample Sizes of Experiment 3

	$P^* = 0.90$		$P^* = 0.95$	
Procedure	P(CS)	$\overline{T}$	P(CS)	$\overline{T}$
Rinott(20)	99.48%	2361	99.66%	2989
ETSS(20)	95.78%	1038	96.69%	1279
AT$_{.5}$(20)	97.18%	1164	97.82%	1431
AT$_S$(20)	98.56%	1362	98.84%	1695
AT$_2$(20)	99.02%	1549	99.54%	1948
AT$_C$(20)	99.99%	1563	99.97%	1957
Rinott(30)	99.33%	2247	99.66%	2830
ETSS(30)	97.49%	1078	97.92%	1302
AT$_{.5}$(30)	98.32%	1206	98.86%	1473
AT$_S$(30)	98.77%	1348	99.37%	1653
AT$_2$(30)	99.27%	1579	99.64%	1954
AT$_C$(30)	100.00%	1476	100.00%	1816

Table 8: Detailed Sample Sizes for $P^* = .90$ and $n_0 = 20$ of Experiment 3

Dn	Rinott	ETSS	$AT_{.5}$	AT_S	AT_2	AT_C
1	525	470	493	512	518	527
2	443	267	308	362	400	444
3	367	125	162	223	282	338
4	297	49	69	111	162	116
5	233	24	29	47	73	35
6	179	20	20	23	31	21
7	131	20	20	20	20	20
8	91	20	20	20	20	20
9	59	20	20	20	20	20
10	33	20	20	20	20	20

Table 9: Detailed Sample Sizes for $P^* = .95$ and $n_0 = 20$ of Experiment 3

Dn	Rinott	ETSS	$AT_{.5}$	AT_S	AT_2	AT_C
1	665	600	619	648	663	671
2	558	334	387	457	511	563
3	465	155	204	284	355	429
4	374	61	84	142	206	147
5	295	27	33	56	93	42
6	228	20	21	25	36	22
7	167	20	20	20	21	20
8	116	20	20	20	20	20
9	74	20	20	20	20	20
10	42	20	20	20	20	20

5 CONCLUSIONS

Many two-stage indifference-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. Moreover, the performance of the intrinsic subset pre-selection of the ETSS procedure is as good as the procedure described in Goldsman and Nelson (2001). 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 in the first stage. We propose adding a conservative adjustment to ETSS to increase P(CS). We recommend using the adjustment $a = S_b(\bar{X}_b)/d^*$, which is computed dynamically according to the variance of the best alternative at the first stage. 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. However, conservative users can increase the value of the adjustment to meet their requirements. For example, let $U(\bar{X}_b)$ be the upper one-tailed P^* confidence limit of μ_b.

Moreover, since the quality of the first-stage sample means have great influence in the performance of ETSS procedure, we recommend using a larger first-stage sample size for the ETSS procedure. Note that with the same minimal required possibility of correct selection P^*, the total sample sizes using $n_0 = 30$ are less than using $n_0 = 20$ and achieve higher P(CS) at the same time in all our experiments with AT_S.

Our experimental results show that the AT_a (adjusted ETSS) procedure is a powerful tool for selecting the best design out of k alternatives. The main advantage of the AT_a 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. Furthermore, the added adjustment efficiently improves P(CS) of the ETSS procedure with only slightly larger simulation replications. The simplicity of this method should make it attractive to simulation practitioners or software developers.

One drawback of two-stage selection procedures is that they rely heavily on the information from only one stage. To eliminate this drawback, we are working on sequentializing the ETSS procedure.

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.

Berger, J. O., and J. Deely. 1994. A Bayesian Approach to Ranking and Selection of Related Means with Alternative to AOV Methodology. *Journal of American Statistics Association* 83: 364–373.

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. 2001. Using Common Random Numbers with Two-Stage Selection Procedures. In *Proceedings of the 2001 Winter Simulation Conference*, ed. B.A. Peters, J.S. Smith, D.J. Medeiros, and M.W. Rohrer, 408–416. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.

Chen, E. J., and W. D. Kelton. 2000a. An Enhanced Two-Stage Selection Procedure. In *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. 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, 617–626. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.

Chick, S. E. 1997. Selecting The Best System: A Decision-Theoretic Approach. In *Proceedings of the 1997 Winter Simulation Conference*, ed. S. Andradóttir, K.J. Healy, D.H. Withers, and B.L. Nelson, 326-333. 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., and B. L. Nelson. 2001. Statistical Selection of The Best System. In *Proceedings of the 2001 Winter Simulation Conference*, ed. B.A. Peters, J.S. Smith, D.J. Medeiros, and M.W. Rohrer, 139–146. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.

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

Rinott, Y. 1978. On Two-stage Selection Procedures and Related Probability Inequalities. *Communications in Statistics* A7: 799-811.

Tong, Y. L. 1980. *Probability Inequalities in Multivariate Distributions.* New York: Academic Press.

Wilcox, R. R., 1984. A Table for Rinott's Selection Procedure. *Journal of Quality Technology* Vol. 16, No. 2: 97-100.

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 `<chenej@basf.com>`.