INFORMS Journal on Computing
Vol. 19, No. 1, Winter 2007, pp. 101–111
ISSN 1091-9856 | EISSN 1526-5528 | 07 | 1901 | 0101
DOI 10.1287/ijoc.1050.0141

Simulation Allocation for Determining the Best Design in the Presence of Correlated Sampling

Michael C. Fu
The Robert H. Smith School of Business, Institute for Systems Research, University of Maryland, College Park, Maryland 20742-1815, USA, mfu@rhsmith.umd.edu

Jian-Qiang Hu
Department of Manufacturing Engineering, Boston University, Brookline, Massachusetts 02446, USA, hqiang@bu.edu

Chun-Hung Chen
Department of Systems Engineering & Operations Research, George Mason University, 4400 University Drive, MS 4A6, Fairfax, Virginia 22030, USA, cchen9@gmu.edu

Xiaoping Xiong
Freddie Mac, 1551 Park Run Drive, McLean, Virginia 22102, xiaoping_xiong@freddiemac.com

We consider the problem of efficiently allocating simulation replications in order to maximize the probability of selecting the best design under the scenario in which system performances are sampled in the presence of correlation. In the case of two designs, we are able to derive the optimal allocation exactly, and find that in the presence of positive correlation, unless the variance of one design is significantly larger than that of the other, the number of simulation replications should be identical. In extending to a general number of competing designs, an approximation for the asymptotically optimal allocation is obtained. The approximation coincides with the independent case derived previously in the limit as the correlation vanishes and also agrees with the two-design exact solution. Furthermore, the allocations prescribed by the results seem to match intuition, in terms of the relationship to correlations and relative variances between designs, again suggesting that equal allocation is optimal for sufficiently high positive correlation. An allocation algorithm based on the approximation is proposed and tested on several numerical examples.

Key words: simulation budget allocation; correlation; optimal sampling schemes; multiple-comparison procedures
History: Accepted by Susan Sanchez, Area Editor for Simulation; received October 2002; revised August 2003, May 2004, December 2004, March 2005; accepted March 2005.

1. Introduction

The problem we consider is selecting the best design among a finite number of choices, where the performance of each design must be estimated with some uncertainty through stochastic sampling. The primary context is simulation, where the goal is to determine the best allocation of simulation replications among the various designs. This problem setting falls under the well-established branch of statistics known as ranking and selection or multiple comparison procedures. In the context of simulation, Goldsman and Nelson (1998) provide a nice overview of this field. The ranking-and-selection algorithms determine the number of simulation replications required for each design in order to guarantee a pre-specified level of correct selection, whereas multiple-comparison procedures provide confidence intervals on estimated performance differences between systems. More recently, Chen et al. (1997, 2000), Chen and Kelton (2000), and Chick and Inoue (2001a, b) have approached the problem from the perspective of allocating a fixed number of simulation replications in order to maximize the probability of correct selection. Chen (cf. Chen et al. 1997, 2000) has called this *optimal computing budget allocation* (OCBA). Although these appear to be dual problems, a closer look at the specific formulations reveals some important differences, which lead to different solutions. In particular, most of the classical statistics procedures lead to allocations that depend only on variances, whereas the latter approaches lead to the more intuitively appealing conclusion that differences in means should also affect the allocation scheme. However, previous OCBA work (Chen et al. 1997, 2000, 2003) considers only the independent case, whereas we investigate the correlated case. This work thus parallels the work in ordinal optimization of Deng et al. (1992) and the two-stage Bayesian setting considered in Chick and Inoue (2001b).

Correlation is especially important in various scenarios such as simulation or experimental-design (clinical-trial) settings that allow the same subject to be used for multiple experiments (treatments). The correlation is induced purposely in order to reduce the number of simulation replications or trials required to come to a conclusion, which in our framework, is the selection of the best design. Glasserman and Yao (1992), Glasserman and Vakili (1994), and Dai and Chen (1997) also investigate the effects of induced correlation on estimating performance, but do not investigate the sampling or simulation allocation problem. Multiple-comparison and ranking-and-selection procedures that incorporate common random numbers include Yang and Nelson (1991), Nelson and Matejcik (1995), and Kim and Nelson (2001). Since the primary goal of these procedures is to guarantee a specified level of correct selection, they tend to be on the conservative side in their computational requirements. When computation is relatively expensive, or the simulation budget is relatively tight, an efficient resource allocation becomes more crucial. Our work focuses on the effects of correlation on the optimal allocation of simulation effort.

We derive optimal allocations for the setting of maximizing the probability of correct selection subject to a budget constraint on the total number of samples, when there is correlated sampling of the estimated design performances. Our results can be used for specifying efficient allocations based on estimates of the means, variances, and correlations. For the case of two designs, we find the exact optimal allocation. When the two variances are the same, equal allocation is optimal for *any* positively correlated sampling; otherwise, the greater the difference in the variances, the higher level the level of positive correlation required for equal allocation to be optimal, and if the variance difference is sufficiently large, equal allocation can never be optimal. For more than two competing designs, an approximation for the asymptotically (as the number of samples increases) optimal allocation is obtained, and an allocation algorithm based on the approximation is proposed. The solution obtained from the approximation coincides with the independent case derived previously in Chen et al. (2000) in the limit as the correlation vanishes. In both the exact and the approximate solutions, the derived optimal allocations depend on the following factors: the variances of each of the designs, the relative difference between the means of each design and the mean of the best design, and the correlation between each design and the best design. The first two factors are also present in the independent case. The second factor does not come into play in the two-design case, since there is just a single relative difference, nor in the special three-design case where the two non-optimal means are equal, which leads to the two relative differences being identical.

The rest of the paper is organized as follows. Section 2 provides the problem setting, which is followed by the exact solution for the case of two designs in §3. An approximate solution for more than two designs is derived in §4, and some bounds on the approximation are provided. In §5, the proposed algorithm based on the results is presented. In §6, the algorithm is applied to several examples, including a queueing system, to illustrate the effects of correlation on the probability of correct selection as a function of the sampling budget. Section 7 offers some concluding remarks.

2. Problem Setting

We introduce the following notation:

T = total number of simulation replications (budget),
k = number of designs,
N_i = number of simulation replications allocated to design i,
$\tilde{J}_{im}$ = the mth simulation replication for design i, $m = 1, \ldots, N_i$,
$\bar{J}_i = (1/N_i)\sum_{m=1}^{N_i} \tilde{J}_{im}$, the sample average for design i,
μ_i = mean for design i,
σ_i^2 = variance for design i,
C_{ij} = covariance between paired replications of design i and j ($=\sigma_i^2$ for $j = i$),
$\rho_{ij} = C_{ij}/(\sigma_i\sigma_j)$, correlation between paired replications of design i and j.

The objective is to find a simulation budget allocation that maximizes the probability of correct selection (PCS), where "correct selection" is defined as picking the best design, which we will take as the design having maximum mean. Without loss of generality, we assume $\mu_1 > \mu_i \ \forall\, i > 1$, i.e., design 1 is the best. More precisely, we want to maximize $P(\bar{J}_1 - \bar{J}_i > 0,\ i = 2, \ldots, k)$ by determining the values of $N_1, N_2, \ldots, N_k$ subject to $N_1 + N_2 + \cdots + N_k = T$. This formulation implicitly assumes that the computational cost of each replication is constant across designs; however, it can be generalized to other settings by changing the budget constraint to $c_1N_1 + c_2N_2 + \cdots + c_kN_k = T$, where c_i reflects the (relative) cost of a replication for design i, and T reflects a more general computing budget rather than simply the total number of simulation replications.

We assume that $\{\tilde{J}_{im},\ m = 1, \ldots, N_i,\ i = 1, \ldots, k\}$ has a joint normal distribution, and

$$\operatorname{cov}(\tilde{J}_{im}, \tilde{J}_{jn}) = \begin{cases} 0, & \text{if } m \neq n, \\ C_{ij}, & \text{otherwise.} \end{cases}$$

Then, both $\{\bar{J}_i,\ i=1,\ldots,k\}$ and $\{\bar{J}_1-\bar{J}_i,\ i=2,\ldots,k\}$ also have joint normal distributions with respective covariance matrices

$$\Theta = \left[\frac{C_{ij}}{\max(N_i, N_j)}\right]_{k\times k},$$

and

$$\Lambda = [\Lambda_{ij}]_{(k-1)\times(k-1)} = \chi^T \Theta \chi,$$

where

$$\chi = \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ -1 & 0 & 0 & \cdots & 0 \\ 0 & -1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 \end{pmatrix}_{k\times(k-1)}.$$

To calculate $P(\bar{J}_1-\bar{J}_i>0,\ i=2,\ldots,k)$, we express $(\bar{J}_1-\bar{J}_2,\ldots,\bar{J}_1-\bar{J}_k)$ as

$$\begin{aligned}(\bar{J}_1-\bar{J}_2,\ldots,\bar{J}_1-\bar{J}_k) \\ &= (z_1,\ldots,z_{k-1})U + (\mu_1-\mu_2,\ldots,\mu_1-\mu_k), \quad (1)\end{aligned}$$

where $z_1,\ldots,z_{k-1}$ are i.i.d. standard normal random variables and $U=[u_{ij}]_{(k-1)\times(k-1)}$ is an upper triangular matrix (i.e., $u_{ij}=0$ if $i>j$) such that $U^TU=\Lambda$. For $k=2$, we have

$$u_{11}^2 = \Lambda_{11} = \frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2} - \frac{2C_{12}}{\max(N_1, N_2)}. \quad (2)$$

Based on (1), we have

$$\bar{J}_1 - \bar{J}_i = \sum_{m=1}^{i-1} u_{m,i-1} z_m + \mu_1 - \mu_i,$$

for $i=2,\ldots,k$. Therefore,

$$\begin{aligned} P(\bar{J}_1-\bar{J}_i>0,\ i=2,\ldots,k) \\ &= P\left(\sum_{m=1}^{i-1} u_{m,i-1}z_m > \mu_i-\mu_1,\ i=2,\ldots,k\right) \\ &= \frac{1}{(\sqrt{2\pi})^{k-1}} \\ &\quad \cdot \int_{u_{11}x_1\ge\mu_2-\mu_1} \cdots \int_{\sum_{m=1}^{k-1} u_{m,k-1}x_m \ge \mu_k-\mu_1} \prod_{m=1}^{k-1} \exp\left(-\frac{x_m^2}{2}\right) \\ &\quad \cdot dx_{k-1}\cdots dx_1. \quad (3)\end{aligned}$$

The objective herein is to maximize the PCS (3) subject to a limited computing budget, i.e., $N_1+N_2+\cdots+N_k=T$.

3. Solution for Two Designs

In the case of two designs, i.e., $k=2$, the PCS (3) simplifies to

$$P(\bar{J}_1-\bar{J}_2>0) = \frac{1}{\sqrt{2\pi}} \int_{u_{11}x_1\ge\mu_2-\mu_1} \exp\left(-\frac{x_1^2}{2}\right) dx_1,$$

so that maximizing the probability is simply a matter of taking the lower limit of the defining integral as small as possible, i.e., minimizing u_{11} as given by (2). Thus, the problem reduces to

$$\min_{N_1, N_2}\left\{\frac{\sigma_1/\sigma_2}{N_1} + \frac{\sigma_2/\sigma_1}{N_2} - \frac{2\rho_{12}}{\max(N_1, N_2)}\right\} \quad (4)$$

subject to $N_1+N_2=T$.

To simplify notation, we drop the subscript on ρ_{12} and write simply ρ throughout. We also introduce $c=\sigma_1/\sigma_2$. Without loss of generality, we assume $c\ge 1$. If $c<1$, then reverse the roles of N_1 and N_2, i.e., N_1 *corresponds to the system with greater variance*. Note that we can do this because the optimization problem for two systems given by (4) involves neither μ_1 nor μ_2. Using the introduced notation, we rewrite (4) as

$$\min_{N_1, N_2}\left\{\frac{c}{N_1} + \frac{1/c}{N_2} - \frac{2\rho}{\max(N_1, N_2)}\right\}. \quad (5)$$

Note that the objective function in (5) provides some intuition for the effect of correlation on the relative simulation allocation. It is known that when $\rho=0$, the optimal solution is

$$\frac{N_1}{N_2} = \frac{\sigma_1}{\sigma_2},$$

i.e., the simulation allocation depends only on the standard deviations when there is no correlation in the sampling.

We solve (5) by treating it as a continuous-variable optimization problem. In actual implementation, a rounding scheme is required to obtain integer solutions (see §6). This is also supported by simulation results in Chen et al. (2005), which indicate relative insensitivity to small perturbations in the allocation. First, we have the following:

Lemma 1. *There exists a solution to* (5) *such that* $N_1\ge N_2$.

Proof. Assume otherwise, i.e., that $N_1<N_2$, and let $\widetilde{N}_1$ and $\widetilde{N}_2$ denote a minimizing pair of (5) under this constraint. Then, since $c\ge 1$ and $\widetilde{N}_2>\widetilde{N}_1$, we have

$$c(\widetilde{N}_2-\widetilde{N}_1) \ge \frac{1}{c}(\widetilde{N}_2-\widetilde{N}_1).$$

Now, by reversing the allocation, a quick check gives

$$\frac{c}{\widetilde{N}_1} + \frac{1/c}{\widetilde{N}_2} = \frac{c\widetilde{N}_2+\widetilde{N}_1/c}{\widetilde{N}_1\widetilde{N}_2} \ge \frac{c\widetilde{N}_1+\widetilde{N}_2/c}{\widetilde{N}_1\widetilde{N}_2} = \frac{c}{\widetilde{N}_2} + \frac{1/c}{\widetilde{N}_1},$$

providing an $N_1 \geq N_2$ solution at least as good as the best $N_1 < N_2$ solution. □

Thus, we can assume $N_1 \geq N_2$, so the function to be minimized in (5) becomes:

$$h(N_1) = \frac{c-2\rho}{N_1} + \frac{1/c}{T-N_1}. \tag{6}$$

First note that if $c - 1/c \leq 2\rho$, then (6) is monotonically increasing in N_1 on the range $N_1 \geq N_2$, so the minimum occurs at the boundary $N_1 = N_2$. Otherwise, differentiating (6) and setting to zero, one obtains:

$$\frac{N_1}{N_2} = \sqrt{c(c-2\rho)}. \tag{7}$$

The ratio given by (7) can provide a further indication of the effect of correlation on the relative allocation. In the case of independent simulations (i.e., $\rho = 0$), $N_1/N_2 = c$, where c is the ratio of the standard deviations.

Substituting $N_2 = T - N_1$, we obtain the final solution:

$$N_1^* = \begin{cases} T\dfrac{\sqrt{c(c-2\rho)}}{1+\sqrt{c(c-2\rho)}} \geq N_2^* & 2\rho \leq c - 1/c; \\ T/2 = N_2^* & 2\rho \geq c - 1/c; \end{cases} \tag{8}$$

with $N_2^* = T - N_1^*$, the solution being continuous across the boundary $\rho = c - 1/c$. Not surprisingly, the optimal allocation depends on the magnitude of the correlation and the relative variances. However, for large-enough differences in the variances, the second condition in (8) is impossible, thereby leaving only a single region. In particular, if $c > 1+\sqrt{2}$, then $c - 1/c > 2 \geq 2\rho$, so the optimal allocation is provided by (7). At the other extreme, the special case of equal variances ($c = 1$ or $\sigma_1 = \sigma_2$), the second condition in (8) leads to the conclusion that for any $\rho \geq 0$, the number of simulation replications should be identical, just as was the case for independent systems. For strictly negative correlation with $c = 1$, there is an alternate optimum for the first condition in (8) that simply switches the allocation of N_1 and N_2. This should be clear from the proof of Lemma 1 when $c = 1$, where the last inequality in the next-to-last line becomes an equality.

4. Approximate Solution for More Than Two Designs

In general, the allocation problem is analytically intractable for more than two designs, so we propose in §4.1 an approximate problem. An exact solution scheme to the approximate problem is given in §4.2 and applied to solving a special case for three designs in §4.3. Throughout, we treat the non-negative variables $\{N_i\}$ to be optimized as continuous.

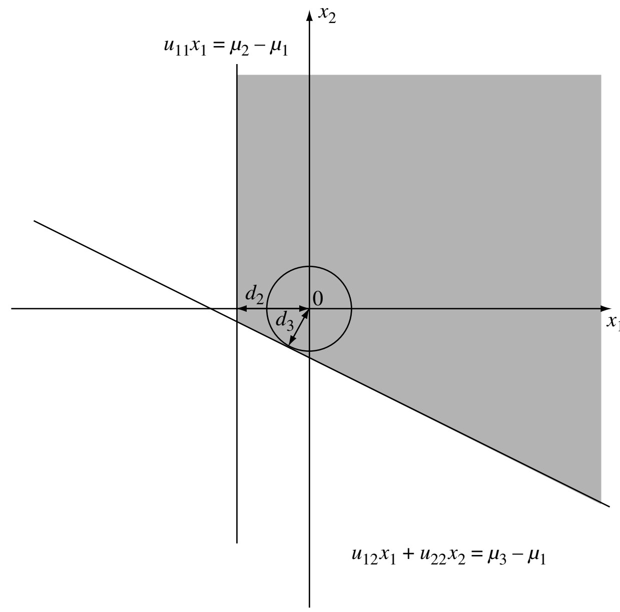

Figure 1 **Area of Integration for Approximation Is the Circle, Where Dominant Values of Integrand** $\exp(-(x_1^2+x_2^2)/2)$ **Are Captured**

4.1. An Approximate Problem

Integrating the PCS expression given by (3) over the desired region of integration is intractable. However, since the integrand in (3) involves terms that decrease exponentially, the integral will be dominated by the region closest to the origin. Thus, as a surrogate for maximizing (3), we maximize the size of the hypersphere centered at the origin that is contained in this region, which is equivalent to maximizing the hypersphere's radius given by $d \equiv \min(d_2, d_3, \ldots, d_k)$, where d_i is the distance from the origin to the hyperplane $\sum_{m=1}^{i} u_{m,i-1}x_m = \mu_i - \mu_1$ $(i = 2, \ldots, k)$. Figure 1 illustrates the problem for $k = 3$, showing the approximating circle relative to the original (shaded) region of integration. Simple algebra yields

$$d_i = \sqrt{\alpha_i/\Lambda_{i-1,i-1}} = \sqrt{\alpha_i/\lambda_i},$$

where

$$\alpha_i \equiv (\mu_1 - \mu_i)^2,$$

$$\lambda_i \equiv \Lambda_{i-1,i-1} = \frac{\sigma_1^2}{N_1} + \frac{\sigma_i^2}{N_i} - \frac{2C_{1i}}{\max(N_1, N_i)}.$$

The introduction of α_i and λ_i are to simplify notation in the analysis that follows. Let $\Omega = \{2, \ldots, k\}$ denote the set containing all indices except the best. Then the approximate problem can be rewritten as

$$\max_{\sum_{i=1}^{k} N_i = T} \min_{i \in \Omega} \sqrt{\alpha_i/\lambda_i},$$

or equivalently,

$$\min_{\sum_{i=1}^{k} N_i = T} \max_{i \in \Omega} (\lambda_i/\alpha_i). \tag{9}$$

We now derive a bound for the error between the original problem and the approximate problem. Let P_k denote the maximal value of the PCS (3) for a given budget T, and P_k' denote the PCS associated with the optimal solution to (9), with corresponding minimum radius d'. Obviously we have $P_k \geq P_k'$. Using (3), we have

$$\begin{aligned}0 &\leq P_k - P_k' \\ &\leq \frac{1}{(\sqrt{2\pi})^{k-1}} \int_{\sqrt{\sum_{m=1}^k x_m^2} \leq d'} \exp\left(-\sum_{m=1}^{k-1} \frac{x_m^2}{2}\right) dx_1 \cdots dx_{k-1} \\ &\quad + \frac{1}{(\sqrt{2\pi})^{k-1}} \int_{\sqrt{\sum_{m=1}^k x_m^2} > d'} \exp\left(-\sum_{m=1}^{k-1} \frac{x_m^2}{2}\right) dx_1 \cdots dx_{k-1} - P_k' \\ &\leq \frac{1}{(\sqrt{2\pi})^{k-1}} \int_{\sqrt{\sum_{m=1}^k x_m^2} > d'} \exp\left(-\sum_{m=1}^{k-1} \frac{x_m^2}{2}\right) dx_1 \cdots dx_{k-1} \\ &= C \int_{r > d'} r^{k-2} \exp(-r^2/2)\, dr \equiv B(r),\end{aligned}$$

where C is a constant depending on k and π. When all N_i are equal, we have $d' \sim O(\sqrt{T})$. Using L'Hôpital's rule, we can show $B(r) \sim O(T^{(k-3)/2} e^{-T})$, or more precisely, $\ln(B(r)) \sim O(T)$. This implies that the upper bound of the error decreases exponentially with a rate proportional to T.

4.2. Solution to the Approximate Problem

We first show that, without loss of generality, we can assume that for $i \in \Omega$, $\sigma_i > 0$ and $\mathrm{Var}(X_i - X_1) = \sigma_1^2 + \sigma_i^2 - 2C_{1i} > 0$. Suppose $\sigma_j = 0$ for some $j \in \Omega$. Then it should be clear that $N_j^* = 1$, i.e., we can then eliminate that design j from the problem. Similarly, if $\sigma_1^2 + \sigma_j^2 - 2C_{1j} = 0$ for some $j \in \Omega$, we know the difference between the best design and the jth design is constant almost surely, so again $N_j^* = 1$.

First we consider a degenerate case where $\sigma_1 = 0$. Again, it is optimal to set $N_1^* = 1$, reducing (9) to

$$\max_{\sum_{i=2}^k N_i = T-1} \min_{i \in \Omega} (N_i \alpha_i / \sigma_i^2).$$

Since $\min_{i\in\Omega}(N_i\alpha_i/\sigma_i^2) \leq N_j \alpha_j / \sigma_j^2$ for each $j \in \Omega$,

$$\min_{i\in\Omega}(N_i\alpha_i/\sigma_i^2) \leq \sum_{i=2}^k \beta_i (N_i \alpha_i/\sigma_i^2),$$

where $\{\beta_i\}$ can be any sequence of positive numbers summing to 1. Letting $\beta_i = (\sigma_i^2/\alpha_i)/(\sum_{j=2}^k \sigma_j^2/\alpha_j)$, we have $\min_{i\in\Omega}(N_i\alpha_i/\sigma_i^2) \leq (T-1)/(\sum_{j=2}^k \sigma_j^2/\alpha_j)$, an upper bound that can be achieved only if all $N_i\alpha_i/\sigma_i^2$ are equal. Hence, the optimal allocation will be given by $N_i = (T-1)\beta_i$.

Henceforth, we assume $\sigma_1 > 0$. Denote $\{N_i^*\}$ as the optimal solution to (9), and λ_i^* as the associated value of λ_i. The following theorem is an important characterization of $\{N_i^*\}$.

Theorem 1. *Any optimal solution $\{N_i^*\}$ to (9) has all λ_i^*/α_i (hence d_i) equal ($i \geq 2$).*

Proof. We establish the claim by contradiction by showing that if λ_i^*/α_i are not all equal, we can construct a new allocation $\{N_i'\}$ that is better than $\{N_i^*\}$. Let $Y_1^* = \max_{i\in\Omega}(\lambda_i/\alpha_i)$ and $\Omega_1 = \{i \in \Omega \mid \lambda_i^* = \alpha_i Y_1^*\}$, which correspond, respectively, to the (optimal) value of (9) and to the designs that achieve the optimum. Assuming the claim is not true, we have that $\Omega_1 \neq \Omega$, and $Y_2^* \equiv \max_{i\in\Omega\setminus\Omega_1}(\lambda_i/\alpha_i) < Y_1^* = \max_{i\in\Omega_1}(\lambda_i/\alpha_i)$. Defining $\Omega_2 = \{i \in \Omega_1 \mid N_i^* > N_1^*\}$ and $\Omega_3 = \{i \in \Omega_1 \mid N_i^* \leq N_1^*\}$, we construct an allocation $\{N_i'(\varepsilon),\ i \in \Omega\}$ as follows:

$$N_i'(\varepsilon) = \begin{cases} N_1^* + \varepsilon & i = 1, \\ N_i^* & i \in \Omega_2, \\ N_i^* + \varepsilon & i \in \Omega_3, \\ N_i^* - C_0\varepsilon & i \in \Omega\setminus\Omega_1, \end{cases}$$

where $C_0 = (|\Omega_3| + 1)/(|\Omega| - |\Omega_1|)$ and $\varepsilon > 0$ is to be specified. The constructed allocation $\{N_i'(\varepsilon)\}$ retains feasibility such that $\sum_{i=1}^k N_i'(\varepsilon) = T$, $N_i'(\varepsilon) \geq 0\ \forall i$, as long as $\varepsilon < C_1 \equiv \min_{i\in\Omega\setminus\Omega_1}(N_i^*/C_0)$. We will show that for sufficiently small ε, this allocation gives $\{\lambda_i\}$ leading to the contradiction $\max_{i\in\Omega}(\lambda_i/\alpha_i) < Y_1^*$, i.e., a lower value of (9).

Let $\lambda_i'(\varepsilon)$ be the value of λ_i associated with $N_i'(\varepsilon)$, and define

$$Y_1'(\varepsilon) = \max_{i\in\Omega_1}(\lambda_i'(\varepsilon)/\alpha_i), \quad Y_2'(\varepsilon) = \max_{i\in\Omega\setminus\Omega_1}(\lambda_i'(\varepsilon)/\alpha_i).$$

For all $i \in \Omega_1$, $N_i'(\varepsilon) - N_1'(\varepsilon)$ and $N_i^* - N_1^*$ have the same sign, given $\varepsilon < C_2 \equiv \min(C_1, \min_{i\in\Omega_2} N_i^* - N_1^*)$, where $\min_{i\in\Omega_2} N_i^* = \infty$ if $\Omega_2 = \phi$. Taking $\varepsilon < C_2$, we check the sign of $\lambda_i'(\varepsilon) - \lambda_i^*$.

- $i \in \Omega_2$

$$\begin{aligned}\lambda_i'(\varepsilon) - \lambda_i^* &= \frac{\sigma_1^2}{N_1^* + \varepsilon} + \frac{\sigma_i^2 - 2C_{1i}}{N_i^*} - \frac{\sigma_1^2}{N_1^*} - \frac{\sigma_i^2 - 2C_{1i}}{N_i^*} \\ &= -\frac{\varepsilon\sigma_1^2}{N_1^*(N_1^* + \varepsilon)} < 0,\end{aligned}$$

since $\varepsilon > 0$ and $\sigma_1 > 0$.

- $i \in \Omega_3$

$$\begin{aligned}\lambda_i'(\varepsilon) - \lambda_i^* &= \frac{\sigma_1^2 - 2C_{1i}}{N_1^* + \varepsilon} + \frac{\sigma_i^2}{N_i^* + \varepsilon} - \frac{\sigma_1^2 - 2C_{1i}}{N_1^*} - \frac{\sigma_i^2}{N_i^*} \\ &= -\varepsilon\left[\frac{\sigma_1^2 - 2C_{1i}}{N_1^*(N_1^* + \varepsilon)} + \frac{\sigma_i^2}{N_i^*(N_i^* + \varepsilon)}\right] \\ &\leq -\frac{\varepsilon(\sigma_1^2 + \sigma_i^2 - 2C_{1i})}{N_1^*(N_1^* + \varepsilon)} < 0,\end{aligned}$$

since $N_i^* \leq N_1^*$, $\varepsilon > 0$ and $\sigma_1^2 + \sigma_i^2 - 2C_{1i} > 0$.

We have shown $\lambda_i'(\varepsilon) < \lambda_i^*\ \forall i \in \Omega_1$, $\varepsilon \in (0, C_2)$; hence $Y_1'(\varepsilon) < Y_1^*$. Since $Y_2'(\varepsilon)$ is a continuous function of ε and $Y_2'(0) = Y_2^* < Y_1^*$, we have $Y_2'(\varepsilon) < Y_1^*$ for ε sufficiently small, hence $\max_{i\in\Omega}(\lambda'_i(\varepsilon)/\alpha_i) = \max(Y'_1(\varepsilon), Y'_2(\varepsilon)) < Y_1^*$, establishing that $\{N'_i(\varepsilon)\}$ is a better solution than $\{N_i^*\}$. □

With Theorem 1, we can rewrite the problem given by (9) as follows:

$$\min \; Y \tag{10}$$

$$\text{subject to} \;\; \frac{\sigma_1^2}{N_1} + \frac{\sigma_i^2}{N_i} - \frac{2C_{1i}}{\max(N_1, N_i)} = Y\alpha_i, \tag{11}$$

$$\sum_{i=1}^{k} N_i = T \tag{12}$$

$$Y \geq 0, \quad N_i \geq 0.$$

Let $\{N_i^*\}$ and Y^* correspond to the respective optimal values of $\{N_i\}$ and Y for the problem. The following is the main result.

Theorem 2. *Defining* $M = Y^* N_1^*$ *and*

$$A_i \equiv \frac{\sigma_1^2 + \sigma_i^2 - 2C_{1i}}{\alpha_i}, \quad i = 2, \ldots, k, \tag{13}$$

an optimal solution to the approximation problem (9) *is given by*

$$N_i^* = \begin{cases} \dfrac{\sigma_i^2}{M\alpha_i - \sigma_1^2 + 2C_{1i}} N_1^* \leq N_1^* \\ \qquad \text{if } 2\rho_{1i} \geq \sigma_i/\sigma_1 \text{ or } A_i < M; \\ \dfrac{\sigma_i^2 - 2C_{1i}}{M\alpha_i - \sigma_1^2} N_1^* \geq N_1^* \quad \text{otherwise.} \end{cases} \tag{14}$$

Remarks.

1. The "solution" provided by (14) depends on M, which itself depends on unknown quantities; hence, it is an implicit solution. The next theorem will provide a way to determine M, upon which the normalization constraint (12) will then yield the allocation solution.
2. As in the two-design case, correlations ($2\rho_{1i}$) and ratio of standard deviations (σ_i/σ_1) characterize the solution.
3. From the theorem, it is clear that $N_i^* = N_1^*$ iff $A_i = M$. Thus, equal allocation implies A_i is constant for all $i > 1$. We showed that equal allocation can be optimal for the exact two-design problem in §3, and we will show that it may be optimal for the approximation in a special three-design case considered in §4.3 where $A_2 = A_3$. However, in general, an equal allocation cannot be optimal for the approximation for more than two designs, because if A_i is not constant across Ω, then Theorem 2 implies that at least one $N_i^* \neq N_1^*$.
4. Note that Theorem 2 implies that $2\rho_{1i} \geq \sigma_i/\sigma_1 \Rightarrow A_i \leq M$, which also follows by applying $N_i \leq N_1$ to (11).

For the proof and subsequent discussion, we introduce some additional notation. Let $I = \{i \in \Omega \mid 2\rho_{1i} < \sigma_i/\sigma_1\}$ and $J = \Omega \backslash I = \{i \in \Omega \mid 2\rho_{1i} \geq \sigma_i/\sigma_1\}$.

Proof of Theorem 2. Solving (11) for the two cases $N_i \leq N_1$ and $N_i \geq N_1$ allows the removal of the "max" operator in (11), leading directly to the two cases in (14). Conversely, if $A_i < M$, then

$$\frac{\sigma_i^2}{M\alpha_i - \sigma_1^2 + 2C_{1i}} < 1, \qquad \frac{\sigma_i^2 - 2C_{1i}}{M\alpha_i - \sigma_1^2} < 1.$$

Similarly, the reversed sign and equality can also be established for $A_i > M$ and $A_i = M$, respectively, so

(i) $A_i < M$ iff $N_i^* < N_1^*$, and the first case of (14) holds;

(ii) $A_i > M$ iff $N_i^* > N_1^*$, and the second case of (14) holds;

(iii) $A_i = M$ iff $N_i^* = N_1^*$, and both cases of (14) hold (coincide).

Thus, (14) will follow directly from this and the result $2\rho_{1i} \geq \sigma_i/\sigma_1 \Rightarrow N_i^* \leq N_1^*$, so it only remains to establish this last implication.

The proof again proceeds by contradiction by constructing a new allocation $\{N'_i\}$ that is better than $\{N_i^*\}$, assuming that there exists an i such that $N_i^* > N_1^*$ with $2\rho_{1i} \geq \sigma_i/\sigma_1$ ($i \in J$). We first define $I' = \{i \in I \mid N_i^* > N_1^*\}$, $\Omega' = \{i \in \Omega \mid N_i^* \leq N_1^*\}$, and $J' = \{i \in J \mid N_i^* > N_1^*\} \neq \phi$.

Similar to the previous proof, we define $\{N'_i(\varepsilon)\}$ as follows:

$$N'_i(\varepsilon) = \begin{cases} N_1^* + \varepsilon & i = 1, \\ N_i^* & i \in I', \\ N_i^* + \varepsilon & i \in \Omega', \\ N_i^* - C_3\varepsilon & i \in J', \end{cases}$$

where $C_3 = (1 + |\Omega'|)/|J'|$ and $\varepsilon > 0$ is sufficiently small. Taking $\varepsilon < C_4 \equiv \min(\min_{i\in J'}((N_i^* - N_1^*)/(C_3+1)), \min_{i\in I'} N_i^* - N_1^*)$, where $\min_{i\in I'} N_i^* = \infty$ if $I' = \phi$, $N'_i(\varepsilon) - N'_1(\varepsilon)$ and $N_i^* - N_1^*$ have the same sign for $i \in \Omega\backslash J'$. Again, the constructed allocation $\{N'_i(\varepsilon)\}$ retains feasibility such that $\sum_{i=1}^{k} N'_i(\varepsilon) = T$, $N'_i(\varepsilon) \geq 0 \; \forall\, i$, since $\varepsilon < \min_{i\in J'}((N_i^* - N_1^*)/(C_3+1)) < \min_{i\in J'}(N_i^*/C_3)$. Now we check the sign of $\lambda'_i(\varepsilon) - \lambda_i^*$.

- $i \in I'$

$$\lambda'_i(\varepsilon) - \lambda_i^* = \frac{1}{\alpha_i}\left[\frac{\sigma_1^2}{N_1^* + \varepsilon} + \frac{\sigma_i^2 - 2C_{1i}}{N_i^*} - \frac{\sigma_1^2}{N_1^*} - \frac{\sigma_i^2 - 2C_{1i}}{N_i^*}\right]$$
$$= -\frac{\varepsilon}{\alpha_i}\frac{\sigma_1^2}{N_1^*(N_1^* + \varepsilon)} < 0,$$

since $\varepsilon > 0$ and $\sigma_1 > 0$.

- $i \in \Omega'$

$$\lambda'_i(\varepsilon) - \lambda_i^* = \frac{1}{\alpha_i}\left[\frac{\sigma_1^2 - 2C_{1i}}{N_1^* + \varepsilon} + \frac{\sigma_i^2}{N_i^* + \varepsilon} - \frac{\sigma_1^2 - 2C_{1i}}{N_1^*} - \frac{\sigma_i^2}{N_i^*}\right]$$
$$= -\frac{\varepsilon}{\alpha_i}\left[\frac{\sigma_1^2 - 2C_{1i}}{N_1^*(N_1^* + \varepsilon)} + \frac{\sigma_i^2}{N_i^*(N_i^* + \varepsilon)}\right]$$
$$\leq -\frac{\varepsilon}{\alpha_i}\frac{\sigma_1^2 + \sigma_i^2 - 2C_{1i}}{N_1^*(N_1^* + \varepsilon)} < 0,$$

since $N_i^* \leq N_1^*$, $\varepsilon > 0$ and $\sigma_1^2 + \sigma_i^2 - 2C_{1i} > 0$.

- $i \in J'$

$$\lambda_i'(\varepsilon) - \lambda_i^* = \frac{1}{\alpha_i}\left[\frac{\sigma_1^2}{N_1^* + \varepsilon} + \frac{\sigma_i^2 - 2C_{1i}}{N_i^* - C_3\varepsilon} - \frac{\sigma_1^2}{N_1^*} - \frac{\sigma_i^2 - 2C_{1i}}{N_i^*}\right]$$
$$= -\frac{\varepsilon}{\alpha_i}\left[\frac{\sigma_1^2}{N_1^*(N_1^* + \varepsilon)} + \frac{2C_{1i} - \sigma_i^2}{N_1^*(N_1^* - C_3\varepsilon)}\right] < 0,$$

where the last step uses $2C_{1i} - \sigma_i^2 \geq 0$ because $i \in J' \subset J$.

Since the sign of $\lambda_i'(\varepsilon) - \lambda_i^* < 0\ \forall i$, following arguments similar to those in the proof of Theorem 1 establishes that $\{N_i'(\varepsilon)\}$ is a better solution than $\{N_i^*\}$. □

Theorem 3.

$$M = \arg\min_{x \geq A_0} \frac{h(x)}{T},$$

where

$$h(x) \equiv x + \sum_{i \in J \cup I \setminus I(x)} \frac{\sigma_i^2 x}{\alpha_i x - \sigma_1^2 + 2C_{1i}} + \sum_{i \in I(x)} \frac{(\sigma_i^2 - 2C_{1i})x}{\alpha_i x - \sigma_1^2},$$
$$I(x) = \{i \in I \mid A_i > x\},$$
$$A_0 \equiv \left(\max_{i \in J} A_i\right) \vee \left(\max_{i \in I} \frac{\sigma_1^2}{\alpha_i}\right), \tag{15}$$

where $a \vee b \equiv \max(a, b)$.

Proof. From Theorem 2, for $i \in I$,

$$Y^*\alpha_i = \frac{\sigma_1^2}{N_1^*} + \frac{\sigma_i^2}{N_i^*} - \frac{2C_{1i}}{\max(N_1^*, N_i^*)}$$
$$\geq \frac{\sigma_1^2}{N_1^*} + \frac{\sigma_i^2 - 2C_{1i}}{\max(N_1^*, N_i^*)} > \frac{\sigma_1^2}{N_1^*} \implies M > \frac{\sigma_1^2}{\alpha_i},$$

so combining with Remark 4, we have $M \geq A_0$.

Substituting (14) into (12) yields

$$\left[1 + \sum_{i \in J \cup I \setminus I(M)} \frac{\sigma_i^2}{\alpha_i M - \sigma_1^2 + 2C_{1i}} + \sum_{i \in I(M)} \frac{\sigma_i^2 - 2C_{1i}}{\alpha_i M - \sigma_1^2}\right] N_1 = T. \tag{16}$$

Since $Y^* = M/N_1^*$, the approximation problem (10)/(11)/(12) leads to the conclusion of Theorem 3. □

We now develop a procedure to find M by further characterizing the function $h(x)$, and by breaking the optimization problem down into intervals characterized by values of A_i above the minimum value A_0. First, define the set containing those elements:

$$I_0 = \{i \in I \mid A_i > A_0\},$$

and then define the "order statistics" of elements in I_0 by $\{A_{(i)}, i = 1, \ldots, |I_0|\}$, where $A_{(i)}$ is the ith smallest element in $\{A_i, i \in I_0\}$, i.e., $A_{(i)} \leq A_{(j)}$ for $i < j$. Also, let $A_{(0)} = A_0$ and $A_{(|I_0|+1)} = \infty$. Figure 2 depicts these quantities graphically. Now we define the (decreasing) sets

$$I_n = \{i \in I \mid A_i > A_{(n)}\},$$

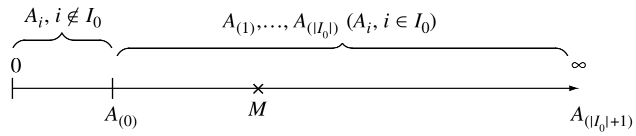

Figure 2 **Locating M in Intervals Defined by Order Statistics $A_{(i)}$**

which will replace the set $I(x)$ in Theorem 3. With these definitions, we can rewrite the optimization problem in Theorem 3.

Corollary 1.

$$M = \arg_{x \geq A_0}\left\{\min_{n \in \{0, \ldots, |I_0|\}} \min_{x \in [A_{(n)}, A_{(n+1)}]} \frac{h_n(x)}{T}\right\},$$

where

$$h_n(x) \equiv x + \sum_{i \in \Omega \setminus I_n} \frac{\sigma_i^2 x}{\alpha_i x - \sigma_1^2 + 2C_{1i}} + \sum_{i \in I_n} \frac{(\sigma_i^2 - 2C_{1i})x}{\alpha_i x - \sigma_1^2}. \tag{17}$$

Since h_n is smooth on $[A_{(n)}, A_{(n+1)}]$ for all $n \in \{0, \ldots, |I_0|\}$, in order to calculate the minimum, we need to locate all the zeros, if any, of the derivative of h_n in $[A_{(n)}, A_{(n+1)}]$. Straightforward differentiation of (17) yields

$$h_n'(x) \equiv 1 - \sum_{i \in \Omega \setminus I_n} \frac{\sigma_i^2(\sigma_1^2 - 2C_{1i})}{(\alpha_i x - \sigma_1^2 + 2C_{1i})^2} - \sum_{i \in I_n} \frac{\sigma_1^2(\sigma_i^2 - 2C_{1i})}{(\alpha_i x - \sigma_1^2)^2}. \tag{18}$$

Since x appears as a quadratic term in the denominator of the summands, and there are $(k-1)$ total terms in the summations, the equation $h_n'(x) = 0$ can be rewritten as a polynomial equation of degree $2k-2$, and the number of zeros might be as many as $2k-2$. Theorem 4 removes that possibility, reducing the total number of possible candidates for the global optimizer to $|I_0| + 1$.

Theorem 4. *$h_n'(\cdot)$ has at most one zero on the interval $[A_{(n)}, A_{(n+1)}]$.*

Proof. We can rewrite $h_n'(x)$ in (18) as $1 - \sum_{i=2}^k b_i'/(x - b_i)^2$, where

$$b_i' \equiv \begin{cases} \sigma_i^2(\sigma_1^2 - 2C_{1i})/\alpha_i^2, & i \in \Omega \setminus I_n, \\ \sigma_1^2(\sigma_i^2 - 2C_{1i})/\alpha_i^2, & i \in I_n, \end{cases} \quad \text{and}$$

$$b_i \equiv \begin{cases} (\sigma_1^2 - 2C_{1i})/\alpha_i, & i \in \Omega \setminus I_n, \\ \sigma_1^2/\alpha_i, & i \in I_n. \end{cases}$$

So we have

$$h_n'(x)x^2 = x^2 - \sum_{i=2}^{k} b_i' + \sum_{i=2}^{k} \frac{-2b_i'b_i(x-b_i/2)}{(x-b_i)^2}.$$

It is easy to show $b_i'b_i \geq 0$. We can also show $A_{(n)} > b_i$ for each $I \in \Omega$: for $i \in I_n$, $A_{(n)} > A_0 \geq b_i$ by the definition of A_0; for $i \in \Omega\backslash I_n$, $A_{(n)} \geq A_i > b_i$ by the definition of A_i and $\sigma_i > 0$. Hence, $x \in [A_{(n)}, A_{(n+1)}]$ implies $x > b_i$. Then we can show $h_n'(x)x^2$ is strictly increasing by checking the sign of its derivative. This further implies it has at most one zero when $x \geq A_{(n)} > 0$, and the same holds for $h_n'(x)$. □

If $h_n'(x)$ has no zero within the interval $x \in [A_{(n)}, A_{(n+1)}]$, then the minimum of $h_n(x)$ over that interval occurs at one of the interval endpoints $x = A_{(n)}$ or $x = A_{(n+1)}$. To see which is the case, we need only check the signs of $h_n'(x)$ at the endpoints. If the LHS is negative and the RHS positive, then there is a unique local minimum in the interval at the zero of $h_n'(x)$. If both are positive, then the minimum occurs at the left endpoint. If both are negative, then the minimum occurs at the right endpoint and can be ignored, because if intervals are checked from left to right, this value will be included in the next interval considered. If the LHS is positive and the RHS is negative, then the minimum could be at either endpoint, but again the RHS is included in the next interval, so only the LHS needs to be checked. This leads to the scheme shown as Figure 3 for determining M in Theorem 3, and then the optimal allocation for the problem given by (10)/(11)/(12).

Remarks.

1. If $N_1 > N_i$ $\forall i \neq 1$ (iff $A_i < M$ $\forall i \neq 1$), then the search reduces to considering the single (rightmost, cf. Figure 2) interval $[A_{(|I_0|)}, \infty)$, and M occurs at the single zero of $h'_{(|I_0|)}$, i.e.,

$$1 = \sum_{i=2}^{k} \frac{\sigma_i^2(\sigma_1^2 - 2C_{1i})}{(\alpha_i x - \sigma_1^2 + 2C_{1i})^2}.$$

Using this and the first case in (14) for $(N_i^*/N_1^*)^2$ gives

$$N_1^* = \sqrt{\sum_{i=2}^{k} \frac{\sigma_1^2 - 2C_{1i}}{\sigma_i^2}(N_i^*)^2}.$$

For $N_1 \gg N_i$, (11) simplifies to $\sigma_i^2/N_i^* \approx Y^*\alpha_i$ $\forall i \neq 1$, so we have

$$\frac{N_i^*}{N_j^*} \approx \frac{\sigma_i^2/\sigma_j^2}{\alpha_i/\alpha_j}, \quad i, j \neq 1.$$

These two equations lead to a solution that is very simple to compute. Furthermore, the solution corresponds to the allocation in Chen et al. (2000) for the independent case ($C_{1i} = 0$ $\forall i \neq 1$).

2. If A_0 is sufficiently large such that I_0 is empty, the search simplifies to a single interval $[A_0, \infty)$ (cf. Figure 2) with $N_i \leq N_1$ $\forall i \neq 1$.

3. Step 3 can be simplified for some special cases. For example, if $C_{1i} \geq 0$ for all $i \in I$, we can show that $a_i^- \leq a_{i+1}^+$ for all $i \in \{0, \ldots, |I_0|-1\}$. Then we can apply a stopping rule in the scheme: $a_i^+ a_i^- < 0$ for some i; both $a_i^- \leq 0$ and $a_{i+1}^+ \geq 0$ for some i. The global optimizer will be zero in interval $[A_{(i)}, A_{(i+1)}]$ for the former case and $A_{(i)}$ for the latter. In particular, if $C_{1i} = 0$ $\forall i \in I$, $h_i'(x)$ are identical, and we need only consider the single interval $[A_0, \infty)$ to locate the unique optimizer.

- **Inputs:** $\{\mu_i\}$ and $\{C_{ij}\}$.
- *Step* 1. Calculate $\{A_i,\ i = 0, 1, \ldots, k\}$ via (15) and (13).
- *Step* 2. Determine the set $I_0 = \{i \in I \mid A_i > A_0\}$ (possibly empty), and the "order statistics" of elements in I_0: $\{A_{(i)},\ i = 1, \ldots, |I_0|\}$, where $A_{(i)}$ is the ith smallest element in $\{A_i,\ i \in I_0\}$.
- *Step* 3. Calculate $a_n^+ = \lim_{x \downarrow A_{(n)}^+} h_n'(x)$, $a_n^- = \lim_{x \uparrow A_{(n+1)}^-} h_n'(x)$, $n = 0, \ldots, |I_0|$, where h_n' is given by (18).
- *Step* 4. Set $M := A_0$, $h^* := \lim_{x \downarrow A_{(0)}} h_0(x)$, where h_n is given by (17).
- *Step* 5. For $n = 0$ to $|I_0|$,
  locate local potential minimum on interval $[A_{(n)}, A_{(n+1)}]$:

$$x^* := \begin{cases} \underset{x \in [A_{(n)}, A_{(n+1)}]}{\arg} \{h_n'(x) = 0\} \text{ (unique)} & a_n^+ < 0 \text{ and } a_n^- > 0, \\ A_{(n)} & \text{otherwise.} \end{cases}$$

  If $h_n(x^*) < h^*$, then $M := x^*$, $h^* := h_n(x^*)$ (else no change).
- *Step* 6. Using M, find N_1^* via (16), and then N_i^*, $i = 2, \ldots, k$, via (14).
- **Return** optimal allocation $\{N_i^*\}$ for the approximation problem.

Figure 3 **Solution Scheme for Approximation Problem**

4.3. A Special Case for Three Designs

We consider a special case where designs 2 and 3 are symmetric, i.e., $\mu_2 = \mu_3$, $\sigma_2 = \sigma_3$, and $\sigma_{12} = \sigma_{13}$, so that $N_2 = N_3$, $\lambda_{22} = \lambda_{33}$, and $d_2 = d_3$. Then the original approximation problem of maximizing $\min(d_2, d_3)$ is reduced to minimizing $\lambda_{22} = \lambda_{33}$, i.e.,

$$\min_{N_1, N_2} \left\{ \frac{\sigma_1/\sigma_2}{N_1} + \frac{\sigma_2/\sigma_1}{N_2} - \frac{2\rho_{12}}{\max(N_1, N_2)} \right\} \tag{19}$$

subject to $N_1 + 2N_2 = T$. The simplification leads to two cases to consider. Let $c = \sigma_1/\sigma_2 = \sigma_1/\sigma_3$.

Case 1. $2\rho_{12} = 2\rho_{13} < \sigma_2/\sigma_1 = \sigma_3/\sigma_1 = 1/c$, i.e., $I = \Omega$.

Simple calculation gives us $A_0 = \sigma_1^2/\alpha_2$ and $A_2 = A_3 = (\sigma_1^2 + \sigma_2^2 - 2C_{12})/\alpha_2 = A_{(1)} = A_{(2)}$. Also we have $h_0'(x) = 1 - 2\sigma_1^2(\sigma_2^2 - 2C_{12})/(x\alpha_2 - \sigma_1^2)^2$ on $(A_{(0)}, A_{(1)}]$ and $h_2'(x) = 1 - 2\sigma_2^2(\sigma_1^2 - 2C_{12})/(x\alpha_2 - \sigma_1^2 + 2C_{12})^2$ on $[A_{(1)}, \infty)$. These functions yield $a_0^+ = -\infty$, $a_0^- = (1 - 2c^2 - 2c\rho_{12})/(1 - 2c\rho_{12})$, $a_2^+ = 1 - 2c^2 + 4c\rho_{12}$, and $a_2^- = 1$. Since $1 - 2c\rho_{12} > 0$ by assumption, we only have to look at the signs of $1 - 2c^2 - 2c\rho_{12}$ and $1 - 2c^2 + 4c\rho_{12}$.

If $1-2c^2-2c\rho_{12}<0$, $A_{(1)}$ is a minimizer in $(A_{(0)}, A_{(1)}]$, and we only have to look at the interval $[A_{(1)}, \infty)$. If $1-2c^2+4c\rho_{12}\geq 0$, $A_{(1)}$ is the minimizer in this interval; thus the optimizer and the first case in (14) gives us $N_i^* = N_1^*$, i.e., $N_1^* = N_2^* = N_3^* = T/3$. If $1-2c^2+4c\rho_{12}<0$, the optimizer is achieved by solving $h_2'(x)=0$, i.e., $x^* = (\sigma_1^2 - 2C_{12} + \sqrt{2\sigma_2^2(\sigma_1^2-2C_{12})})/\alpha_2$. Then (16) gives us $N_1^* = (\sqrt{c^2/2-c\rho_{12}}/(1+\sqrt{c^2/2-c\rho_{12}}))T$.

If $1-2c^2-2c\rho_{12}\geq 0$, there is a minimizer in $(A_{(0)}, A_{(1)}]$. If $1-2c^2+4c\rho_{12}\geq 0$, $A_{(1)}$ is a minimizer in $[A_{(1)}, \infty)$; thus we only need to solve $h_0'(x)=0$, which gives $x_1^* = (\sigma_1^2+\sqrt{2\sigma_1^2(\sigma_2^2-2C_{12})})/\alpha_2$. Then we can get $N_1^* = (c/(c+\sqrt{2-4c\rho_{12}}))T$. However, there is another minimizer in $[A_{(1)}, \infty)$, given $1-2c^2+4c\rho_{12}<0$, which is $x_2^* = (\sigma_1^2-2C_{12}+\sqrt{2\sigma_2^2(\sigma_1^2-2C_{12})})/\alpha_2$. Therefore, we have to compare $h_0(x_1^*)$ and $h_2(x_2^*)$ to determine which is the optimal N_1^*. In other words,

$$N_1^* = \begin{cases} \dfrac{c}{c+\sqrt{2-4c\rho_{12}}}T & \text{if } h_0(x_1^*) < h_2(x_2^*), \\[2ex] \dfrac{\sqrt{c^2/2-c\rho_{12}}}{1+\sqrt{c^2/2-c\rho_{12}}}T & \text{if } h_0(x_1^*) \geq h_2(x_2^*), \end{cases} \tag{20}$$

and thus we obtain the overall allocation:

$$N_1^* = \begin{cases} \dfrac{T}{3} & 1-2c^2-2c\rho_{12}<0 \text{ and} \\ & 1-2c^2+4c\rho_{12}\geq 0, \\[1ex] \dfrac{c}{c+\sqrt{2-4c\rho_{12}}}T & 1-2c^2-2c\rho_{12}\geq 0 \text{ and} \\ & 1-2c^2+4c\rho_{12}\geq 0, \\[1ex] \dfrac{\sqrt{c^2/2-c\rho_{12}}}{1+\sqrt{c^2/2-c\rho_{12}}}T & 1-2c^2-2c\rho_{12}<0 \quad \text{and} \\ & 1-2c^2+4c\rho_{12}<0, \\ \text{RHS of Eq. (20)} & 1-2c^2-2c\rho_{12}\geq 0 \text{ and} \\ & 1-2c^2+4c\rho_{12}<0. \end{cases}$$

Case 2. $2\rho_{12} = 2\rho_{13} \geq \sigma_2/\sigma_1 = \sigma_3/\sigma_1 = 1/c$, i.e., $J=\Omega$. The solution process is even simpler than for case 1, leading to

$$N_1^* = \begin{cases} \dfrac{T}{3} & 1-2c^2+4c\rho_{12}\geq 0, \\[1ex] \dfrac{\sqrt{c^2/2-c\rho_{12}}}{1+\sqrt{c^2/2-c\rho_{12}}}T & 1-2c^2+4c\rho_{12}<0. \end{cases}$$

Since $1-2c^2-2c\rho_{12}<0$ in case 2, this allocation turns out to be already included in the previous solution.

5. Allocation Procedure

Based on the solution procedure to the approximate problem, we propose a two-stage allocation algorithm. The initial stage is used to estimate the means and variances/covariances, and the second stage calculates $\{N_i\}$ based on these estimates to allocate the bulk of the computational budget (assuming $kn_0 \ll T$):

Correlated Budget Allocation (CBA) Algorithm

- **Inputs:** k (# designs), T (total simulation budget), and n_0 (initial sample size for each design).
- *Step* 1. Estimate $\{\mu_i\}$ and $\{C_{ij}\}$, based on n_0 replications of each design.
- *Step* 2. Use estimates of $\{\mu_i\}$ and $\{C_{ij}\}$ as input, and then determine new $\{N_i\}$ by applying the solution scheme described in Figure 3.
- *Step* 3. Perform $(N_i - n_0)^+$ additional replications of design i, $i=1,\ldots,k$.
- **Return** design with largest (overall) sample mean.

Practical implementation involves a choice of n_0. A smaller n_0 provides added flexibility for better allocation of the remaining computing budget, but if n_0 is too small, the estimates of the mean and the variance may be very poor, resulting in a suboptimal allocation. A large value of n_0 may waste computation time in simulating non-critical designs. Chen et al. (2005) recommend a value between 10 and 20, but clearly this can be problem-dependent.

6. Numerical Examples

In this section, we consider several numerical examples in order to investigate the performance of the proposed CBA algorithm. We compare the probability of correct selection (PCS) for three allocations: "CBA" will denote the allocation obtained using the CBA algorithm; "IBA" will denote the *independent* budget allocation (IBA) calculated using the CBA algorithm under zero correlation ($C_{ij}=0$ for all i, j); "EBA" will denote *equal* budget allocation, i.e., $N_1 = N_2 = \cdots = N_k = T/k$. For each example, the PCS is estimated using 100,000 independent "macro" replications of each budget allocation. The estimated standard errors are under 10^{-3}, so the estimates are approximately accurate to three decimal places. Knowledge of which is the true best design is used only to estimate the PCS *after* carrying out a particular allocation prescribed by an algorithm, *not* in the allocation procedures (IBA or CBA) themselves. In all cases, $k=10$, $T=500$, and $n_0=10$ for each design (100 initial samples total), leaving 400 samples to be allocated in the second stage of the IBA and CBA algorithms. To test the sensitivity with respect to the initial allocation, the experiments were also run with $n_0=20$, and this yielded slightly higher PCS results for both IBA and CBA (results are not reported here). We also increased T iteratively for EBA in order to estimate the total sampling budget required to achieve the same level of PCS as CBA.

Table 1 Estimated PCS for Example 1

Correlation (%)	EBA	CBA	IBA
0	0.778	0.830 (44%)	0.830
20	0.808	0.867 (52%)	0.859
50	0.857	0.932 (64%)	0.903
90	0.996	0.997 (8%)	0.926

Note. Efficiency gain over EBA is in parentheses.

Example 1. Equal Variance

The first example we test is adopted from Example 1 in Chen et al. (2000): $\tilde{J}_{im} \sim N(10-i, 6^2)$, $i = 1, 2, \ldots, 10$, so the best design is the first. We consider correlations $\rho = 0.0$, 0.2, 0.5, and 0.9. The estimates of PCS for each procedure are given in Table 1. The "efficiency gain" in parentheses represents the savings in total computing budget by using CBA over using EBA to achieve the same level of PCS. For example, the entry 44% in the first row of Table 1 is calculated as follows: for EBA, it would take $T = 720$ to obtain the CBA-achieved PCS level of 0.830, representing a percentage difference of $220/500 \times 100\%$ in efficiency.

In all three procedures, we see that positive correlation improves simulation efficiency, i.e., the higher the correlation the higher level of PCS achieved, using the same total computing budget. In all cases, CBA performs the best, providing a savings of about 50% in the total computational budget over EBA to achieve the same level of PCS at all but the highest positive correlation, where the allocations nearly coincide.

Example 2. Unequal Variances

We next compare normally distributed designs with the following means and variances randomly (and independently) generated from $U(0, 10)$ and $U(24, 48)$ distributions, respectively:

i	1	2	3	4	5	6	7	8	9	10
μ_i	8.34	0.98	6.49	0.10	8.03	6.21	9.78	9.10	1.32	3.27
σ_i^2	34.35	44.34	28.12	35.93	44.49	43.39	39.72	24.31	42.10	24.42

In this example, design 7 is the best. Similar to Example 1, Table 2 indicates that positive correlation improves simulation efficiency and that CBA outperforms both EBA and IBA, though to a lesser degree.

Table 2 Estimated PCS for Example 2

Correlation (%)	EBA	CBA (% eff gain)	IBA
0	0.667	0.697 (22%)	0.697
20	0.691	0.733 (28%)	0.714
50	0.762	0.805 (28%)	0.748
90	0.957	0.959 (4%)	0.822

Note. Efficiency gain over EBA is in parentheses.

Example 3. Single-Server Queueing System

The last example is a simple queueing system: a first-come, first-served, single-server queue with renewal arrival process and i.i.d. service times. The ten designs correspond to $U/U/1$ queues with interarrival times $U(4, 14)$ and service times $U(1, 15.5 + 0.5i)$, $i = 1, \ldots, 10$. The proxy for design performance (μ_i) is the expected negative (to make it a maximization problem) average system time over the first 15 served customers. Design 1 is the best, since it has smallest mean service time, hence the largest expected negative average system time.

We consider two cases: one with independent sampling and one with correlated sampling. In the correlated-sampling case, each design shares the same arrival process, but the service times across designs are independent. This approximates correlation levels of larger problems such as the design and control of complex queueing networks, where it is usually possible to synchronize external arrival processes in the simulation. Based on some rough estimation analysis, this leads to a correlation of around 12% to 20% for the metric of interest. In the case of independent sampling, both interarrival times and service times across designs are independent. The mean average system time varies from 14 to 28 over the ten designs, while the variances vary from 33 to 120.

The estimated PCS values are provided in Table 3. Again, correlated sampling leads to improvement in simulation efficiency, and the CBA allocation is most efficient. The relative performance is roughly along the lines to that of the $\rho = 0$ and $\rho = 0.2$ cases in the previous two examples, although the reduction in computational burden is even more significant.

The last column in Table 3 considers the sequential version of CBA, denoted by "seqCBA" and described below. The seqCBA algorithm starts with $n_0 = 10$ samples for each design as before ($n_0 = 20$ was also tested, and in contrast to the two-stage algorithm, the large initial sample size actually leads to a slight degradation in the estimated PCS), but then instead of allocating all of the 400 remaining samples at once, it *sequentially* adds an *incremental* computing budget ($\Delta = 20$ is used here) in each iteration until the total computing budget T is exhausted. Clearly, the two-stage procedure is a special case with $\Delta = T - n_0 k$. The numerical results indicate that the sequential version

Table 3 Estimated PCS for Example 3

	EBA	CBA	IBA	seqCBA
Independent sampling	0.828	0.896 (64%)	0.896	0.970 (252%)
Correlated sampling	0.857	0.924 (72%)	0.917	0.981 (256%)

Note. Efficiency gain over EBA is in parentheses.

further improves the efficiency substantially, as now EBA requires more than triple the budget as seqCBA for the same PCS level. Analysis of this sequential dynamic sampling is beyond the scope of this paper, but more discussion can be found in Chen et al. (2004).

Sequential Correlated Budget Allocation (seqCBA) Algorithm

- **Inputs:** k (# designs), T (total simulation budget), n_0 (initial sample size for each design), and Δ (total incremental allocation at each iteration).
- *Step* 0 (Initialization). Let $N_i = n_0\ \forall i$; $\widehat{T} = n_0 k$ (total budget used so far). Perform N_i replications for each design i, $i = 1, \ldots, k$.
- *Step* 1. Estimate $\{\mu_i\}$ and $\{C_{ij}\}$, based on $\{N_i\}$ replications.
- *Step* 2. $\widehat{T} \leftarrow \widehat{T} + \min(\Delta, T - \widehat{T})$. Use estimates of $\{\mu_i\}$ and $\{C_{ij}\}$ as input, and then determine new $\{N_i\}$ by applying the solution scheme described in Figure 3.
- *Step* 3. Perform $(N_i - N_i^{old})^+$ additional replications of design i, $i = 1, \ldots, k$.
- **Repeat** Steps 1, 2, and 3 until $\widehat{T} = T$ (computing budget exhausted).
- **Return** design with largest (overall) sample mean.

7. Concluding Remarks

Clearly, correlation in the context of sampling when trying to compare performance between competing designs can lead to substantially reduced computational costs. Our work attempts to realize the benefits of such induced correlation by considering the computational budget allocation problem for this important case. We are able to derive the exact optimal allocation for the two-design case and characterize when equal allocation is optimal in the presence of positive correlation. Although an exact solution appears to be intractable for the general case, we are able to propose an approximate solution that seems to capture the important features, and we propose an allocation algorithm based on the results. Preliminary numerical results indicate that the algorithm is quite promising, as the most complicated of the three numerical examples, the single-server queueing example, yielded the best computational gains. Furthermore, sequential dynamic sampling yielded even more substantial savings. Simulations on larger systems are of course needed to verify these indications. In addition, one disadvantage of our approach is that a single application (“macro” replication) of the CBA algorithm does not provide an actual estimate of the achieved PCS. So another possible avenue of future research is to pursue the Bayesian framework of Chen et al. (2000) and Chick and Inoue (2001a, b), which produces Bayesian estimates of the PCS.

Acknowledgments

M. C. Fu was supported in part by the National Science Foundation under Grants DMI-9988867 and DMI-0323220, and by the Air Force Office of Scientific Research under Grants F496200110161 and FA95500410210. Much of this work was performed during his sabbatical leave at MIT in the spring of 2002. J. Q. Hu was supported in part by the National Science Foundation under Grant EEC-0088073. C. H. Chen was supported in part by NSF under Grants DMI-0002900, DMI-0049062, and IIS-0325074; by the NASA Ames Research Center under Grants NAG-2-1565 and NAG-2-1643; by FAA under Grant 00-G-016; and by the George Mason University Research Foundation. X. Xiong was supported in part by the National Science Foundation under Grant DMI-0323220. The authors thank the Simulation Area Editor and the anonymous Associate Editor and referees for many useful comments and suggestions that have led to a much improved paper.

References

Chen, C. H., D. He, M. C. Fu. 2006. Efficient dynamic simulation allocation in ordinal optimization. *IEEE Transactions Automatic Control* **51** 2005–2009.

Chen, C. H., D. He, E. Yücesan. 2003. Better-than-optimal simulation runs allocation. *Proc. Winter Simulation Conf.* 490–495.

Chen, C. H., J. Lin, E. Yücesan, S. E. Chick. 2000. Simulation budget allocation for further enhancing the efficiency of ordinal optimization. *Discrete Event Dynam. Systems: Theory Appl.* **10** 251–270.

Chen, C. H., E. Yücesan, L. Dai, H. C. Chen. 2005. Efficient computation of optimal budget allocation for discrete event simulation experiment. *IIE Trans.* Forthcoming.

Chen, E. J., W. D. Kelton. 2000. An enhanced two-stage selection procedure. *Proc. Winter Simulation Conf.* 727–735.

Chen, H. C., C. H. Chen, L. Dai, E. Yücesan. 1997. New development of optimal computing budget allocation for discrete event simulation. *Proc. Winter Simulation Conf.* 334–341.

Chick, S. E., K. Inoue. 2001a. New two-stage and sequential procedures for selecting the best simulated system. *Oper. Res.* **49** 732–743.

Chick, S. E., K. Inoue. 2001b. New procedures to select the best simulated system using common random numbers. *Management Sci.* **47** 1133–1149.

Dai, L., C. Chen. 1997. Rate of convergence for ordinal comparison of dependent simulations in discrete event dynamic systems. *J. Optim. Theory Appl.* **94** 29–54.

Deng, M., Y. C. Ho, J. Q. Hu. 1992. Effect of correlated estimation errors in ordinal optimization. *Proc. Winter Simulation Conf.*, IEEE, Piscataway, NJ, 231–240.

Glasserman, P., P. Vakili. 1994. Comparing Markov chains simulated in parallel. *Probab. Engrg. Inform. Sci.* **8** 309–326.

Glasserman, P., D. D. Yao. 1992. Some guidelines and guarantees for common random numbers. *Management Sci.* **38** 884–908.

Goldsman, D., B. L. Nelson. 1998. Comparing systems via simulation. J. Banks, ed. *Handbook of Simulation: Principles, Methodology, Advances, Applications, and Practice*, Chapter 8. John Wiley & Sons, New York, 273–306.

Kim, S.-H., B. L. Nelson. 2001. A fully sequential procedure for indifference-zone selection in simulation. *ACM Trans. Modeling Comput. Simulation* **11** 251–273.

Nelson, B. L., F. J. Matejcik. 1995. Using common random numbers for indifference-zone selection and multiple comparisons in simulation. *Management Sci.* **41** 1935–1945.

Yang, W. N., B. L. Nelson. 1991. Using common random numbers and control variates in multiple-comparison procedures. *Oper. Res.* **39** 583–591.