Proceedings of the 2015 Winter Simulation Conference
L. Yilmaz, W. K V. Chan, I. Moon, T. M. K. Roeder, C. Macal, and M. D. Rossetti, eds.

NON-MONOTONICITY OF PROBABILITY OF CORRECT SELECTION

Yijie Peng

Department of Management Science
Fudan University
220 Handan Road
Yangpu, SH, 200433, CHINA

Chun-Hung Chen

Department of SE & OR
George Mason University
4400 University Drive
Fairfax, VA, 22030, USA

Michael C. Fu

The Robert H. Smith School of Business
Institute for Systems Research
University of Maryland
College Park, MD, 20742, USA

Jian-Qiang Hu

Department of Management Science
Fudan University
220 Handan Road
Yangpu, SH, 200433, CHINA

ABSTRACT

In Peng et al. (2015b), we show that the probability of correct selection (PCS), a commonly used metric, is not necessarily monotonically increasing with respect to the number of simulation replications. A simple counterexample where the PCS may decrease with additional sampling is provided to motivate the problem. The reference identifies the induced correlations as the source of the non-monotonicity, and characterizes the general scenario under which the phenomenon occurs by a condition where coefficient of variations of the difference in sample means are large. Numerical examples further illustrate the non-monotonic behavior of the PCS for some well-known sampling schemes.

1 INTRODUCTION

Statistical ranking and selection procedures that have been developed for simulation optimization problems involving a given finite set of fixed alternatives include indifference zone (IZ) procedures (Rinott 1978, Bechhofer et al. 1995, Goldsman and Nelson 1998, Hong and Nelson 2005, Kim and Nelson 2006), optimal computing budget allocation (OCBA) (Chen et al. 2000, Chen and Lee 2011, Pasupathy et al. 2014, Peng et al. 2015a, Xu et al. 2015), and expected value of information (EVI) procedures (Chick and Inoue 2001, Branke et al. 2007). In these procedures, one of the most commonly used metrics is the probability of correct selection (PCS). Recently, Peng et al. (2015b) demonstrated that PCS is not necessarily monotone in the number of simulation replications. In particular, the PCS may actually decrease with additional sampling under certain scenarios if the additional replications are allocated to the best alternative. After providing a simplified counterexample to demonstrate the phenomenon, we review some of the theory behind the main results in that paper, which involves analyzing the joint distribution on the difference of pairwise sample means and the induced correlations when there are at least three alternatives. For the general case, a lower bound on the derivative of PCS with respect to the number of simulation replications is provided, which leads to a necessary condition under which the PCS is non-monotonic and a sufficient condition under which the PCS is increasing. In the special case of three designs, the conditions can be determined by one factor, i.e., the coefficient of variations of the difference in sample means. Non-monotonicity of PCS occurs under the condition in which the variances are large relative to the differences in means and the sample size, or in other words a setting where the coefficient of variations (CVs) of the difference in sample means are large (DISMAL), henceforth referred to as the CVs-DISMAL scenario. This corresponds, for example, to the least favorable configuration IZ setting with relatively large variances and small sample size. Even in such situations, the PCS can be shown to be increasing if additional replications are allocated to alternatives other than the best, but such an allocation may contradict, for example, the proportional-to-variance (PTV) allocations typically specified by IZ procedures, as well as OCBA ratio allocations, which are derived in the asymptotic regime where the CVs go to zero. Specifically, numerical experiments illustrate in the CVs-DISMAL scenario the PCS of PTV, OCBA and EVI procedures actually decrease.

Induced correlations were also considered in the optimal selection policy in a Bayesian framework, which has been studied in Gupta and Miescke (1986), Berger and Deely (1988), Gupta and Miescke (1989), Gupta and Miescke (1996), and Peng et al. (2015a). By proving that the value of information is non-negative, it follows that the integrated PCS (IPCS) under a Bayesian framework with an optimal selection policy must be increasing with respect to the number of simulation replications. However, calculation of the optimal selection policy in general is computationally expensive for ranking and selection, although in principle it can be numerically estimated by particle filtering algorithms. Incorporating the induced correlation structure into the allocation policy would be another potential way to ensure the monotonicity of the IPCS, and deserves further exploration.

The rest of the paper is organized as follows. In section 2, we define the problem setting and present a simple counterexample showing that the PCS does not necessarily increase with the simulation budget. Section 3 provides the theoretical explanation for the example by studying the sensitivity of PCS with respect to the number of replications allocated to each alternative. In Section 4, the performances of different allocation policies are tested. In Section 5, we conclude.

2 SETTING AND COUNTEREXAMPLE

We define the problem setting as choosing the maximum mean among k alternatives, and introduce the following notation:

μ_i: mean of ith alternative, $i=1,..,k$;
N_i: number of replications allocated to ith alternative, $i=1,..,k$;
X_{ij}: jth replication of ith alternative, $i=1,..,k$;
$\bar{X}_i \doteq \frac{\sum_{j=1}^{N_i} X_{ij}}{N_i}$, $i=1,..,k$.

Without loss of generality, assume the 1st alternative is the best, i.e., $\mu_1 > \mu_j$, $j=2,..,k$. Then the PCS is given by

$$PCS \doteq P\left(\bar{X}_1 > \bar{X}_2, \cdots, \bar{X}_1 > \bar{X}_k\right) .$$

More replications decrease the standard error of the sample mean, which intuition might suggest would always lead to higher PCS. However, the following counterexample demonstrates otherwise: for $k=3$, suppose $\mu_1 = 0.2$, $\mu_2 = \mu_3 = 0$, with the three alternatives independent and normally distributed, with common variance $\sigma^2 = 1$, $i=1,2,3$. With simulation replications increasing from $N_i = 1$, $i=1,2,3$, to $N_1 = 2$, $N_2 = N_3 = 1$, the PCS is changed from

$$PCS_1 = \int\int_{\{x_1>0,x_2>0\}} f\left(x_1,x_2;\ 0.2/\sqrt{2}, 0.2/\sqrt{2}, 1/2\right)\ dx_1 dx_2$$

to

$$PCS_2 = \int\int_{\{x_1>0,x_2>0\}} f\left(x_1,x_2;\ 0.2/\sqrt{3/2}, 0.2/\sqrt{3/2}, 1/3\right)\ dx_1 dx_2,$$

where f is the density of the two-dimensional normal distribution is given by

$$f(x_1,x_2;m_1,m_2,c_{23}) = \frac{1}{2\pi\sqrt{1-c_{23}^2}} \exp\left\{-\frac{(x_1-m_1)^2 - 2c_{23}(x_1-m_1)(x_2-m_2)+(x_2-m_2)^2}{2(1-c_{23}^2)}\right\} . \quad (1)$$

Numerical calculation shows PCS decreases from 0.39 to 0.37 with one additional replication allocated to the best alternative. The theoretical and graphical explanation for this simple counterexample will be discussed in detail in Section 3, but first we generalize it to highlight the main issues.

3 A SIMPLE COUNTEREXAMPLE FOR THREE ALTERNATIVES

For $k=3$, suppose $\mu_1 = 2\Delta$, $\mu_2 = \Delta$, $\mu_3 = 0$, $\Delta > 0$, with the three alternatives independent and normally distributed, with common variance $\sigma^2 = 1$, $i = 1,2,3$. Then we have

$$E\left[\frac{\bar{X}_1 - \bar{X}_2}{\sqrt{1/N_1+1/N_2}}\right] = \Delta\sqrt{\frac{N_1N_2}{N_1+N_2}}, \quad E\left[\frac{\bar{X}_1 - \bar{X}_3}{\sigma\sqrt{1/N_1+1/N_3}}\right] = 2\Delta\sqrt{\frac{N_1N_3}{N_1+N_3}},$$

$$\mathrm{var}\left(\frac{\bar{X}_1 - \bar{X}_2}{\sqrt{1/N_1+1/N_2}}\right) = \mathrm{var}\left(\frac{\bar{X}_1 - \bar{X}_3}{\sqrt{1/N_1+1/N_3}}\right) = 1,$$

and

$$\mathrm{cov}\left(\frac{\bar{X}_1 - \bar{X}_2}{\sqrt{1/N_1+1/N_2}}, \frac{\bar{X}_1 - \bar{X}_3}{\sqrt{1/N_1+1/N_3}}\right) = \sqrt{\frac{N_2N_3}{(N_1+N_2)(N_1+N_3)}} .$$

The positive correlation exists because the random variable $\bar{X}_1$ appears in both pairs of differences.

If $N_i = N$, $i = 1,2,3$, then $(\bar{X}_1 - \bar{X}_2)/\sqrt{2/N}$ and $(\bar{X}_1 - \bar{X}_3)/\sqrt{2/N}$ jointly follow a two-dimensional normal distribution with means

$$E\left[\frac{\bar{X}_1 - \bar{X}_2}{\sqrt{2/N}}\right] = \Delta\sqrt{N/2}, \quad E\left[\frac{\bar{X}_1 - \bar{X}_3}{\sqrt{2/N}}\right] = \Delta\sqrt{2N},$$

and variances and covariance given by

$$\mathrm{var}\left(\frac{\bar{X}_1 - \bar{X}_2}{\sqrt{2/N}}\right) = \mathrm{var}\left(\frac{\bar{X}_1 - \bar{X}_3}{\sqrt{2/N}}\right) = 1, \quad \mathrm{cov}\left(\frac{\bar{X}_1 - \bar{X}_2}{\sqrt{2/N}}, \frac{\bar{X}_1 - \bar{X}_3}{\sqrt{2/N}}\right) = 1/2.$$

Then as the differences between the means of the three designs go to zero, we know $\lim_{\Delta\to\infty} PCS = 1/3$ by symmetry, where

$$PCS = P\left(\frac{\bar{X}_1 - \bar{X}_2}{\sqrt{2/N}} > 0, \frac{\bar{X}_1 - \bar{X}_3}{\sqrt{2/N}} > 0\right) .$$

Therefore, if Δ is very small and the number of replications for each alternative is relatively small, the PCS should be close to $1/3$.

If $N_2 = N_3 = N$ is kept fixed, then as $N_1 \to \infty$, $\bar{X}_1 \to 2\Delta$ a.s.,

$$E\left[\frac{\bar{X}_1 - \bar{X}_2}{\sqrt{1/N}}\right] = \Delta\sqrt{N}, \quad E\left[\frac{\bar{X}_1 - \bar{X}_3}{\sqrt{1/N}}\right] = 2\Delta\sqrt{N},$$

and

$$\mathrm{var}\left(\frac{\bar{X}_1 - \bar{X}_2}{\sqrt{1/N}}\right) = \mathrm{var}\left(\frac{\bar{X}_1 - \bar{X}_3}{\sqrt{1/N}}\right) = 1, \quad \mathrm{cov}\left(\frac{\bar{X}_1 - \bar{X}_2}{\sqrt{1/N}}, \frac{\bar{X}_1 - \bar{X}_3}{\sqrt{1/N}}\right) = 0 .$$

With N_1 going to ∞, the randomness of $\bar{X}_1$ disappears, so that the correlation of the pairs of the differences goes to zero. In other words, if the observation for the best alternative is deterministic, the differences $(\bar{X}_1 - \bar{X}_i)$, $i = 1,..,k$, are independent. Then,

$$\lim_{\Delta\to 0} PCS = \lim_{\Delta\to 0} P\left(\frac{\bar{X}_1-\bar{X}_2}{\sqrt{1/N}} > 0, \frac{\bar{X}_1-\bar{X}_3}{\sqrt{1/N}} > 0\right)$$
$$= \lim_{\Delta\to 0} P\left(\frac{\bar{X}_2-\Delta}{\sqrt{1/N}} < \Delta\sqrt{N}\right) P\left(\frac{\bar{X}_3}{\sqrt{1/N}} < 2\Delta\sqrt{N}\right) = 1/4 \ .$$

Therefore, if Δ is very small and the number of replications N for alternative 2 and 3 is relatively small, the PCS should be close to $1/4$. Thus, when N_1 increases to ∞, the PCS decreases from approximately 1/3 to approximately 1/4.

4 THEORETICAL ANALYSIS

We introduce the following additional notation:

σ_i^2: variance of ith alternative, $i = 1,..,k$;
ϕ: density of standard normal distribution;
Φ: cumulative distribution function (CDF) of standard normal distribution;

$$v_i \doteq \frac{\sigma_i^2}{N_i}, \quad i = 1,..,k;$$
$$m_i \doteq \frac{\mu_1-\mu_i}{\sqrt{v_1+v_i}}, \quad i = 2,..,k;$$
$$c_{ij} \doteq \frac{v_1}{\sqrt{v_1+v_i}\sqrt{v_1+v_j}}, \quad i,j = 2,..,k,\ i \neq j;$$
$$\tilde{m}_j^{(i)} \doteq (m_j - m_i c_{ij})/\sqrt{1-c_{ij}^2}, \quad j \neq i,\ i = 2,..,k;$$
$$Y_i \doteq \frac{\bar{X}_1-\bar{X}_i}{\sqrt{v_1+v_i}} - m_i, \quad i = 2,..,k \ .$$

Assume $X_{ij} \sim N(\mu_i, \sigma_i^2)$, i.i.d. for $j = 1,...,N_i$, and mutually independent for $i = 1,..,k$, so $\bar{X}_i \sim N(\mu_i, v_i)$ and mutually independent for $i = 1,..,k$, and $Y \doteq (Y_2,\cdots,Y_k) \sim N(0,\Sigma)$, where

$$\Sigma = \begin{pmatrix} 1 & c_{23} & \cdots & c_{2k} \\ c_{23} & 1 & \cdots & c_{3k} \\ \vdots & \vdots & \ddots & \vdots \\ c_{2k} & c_{3k} & \cdots & 1 \end{pmatrix} .$$

The independent normal distribution assumption is common, often justified by the central limit theorem, e.g., by using batching. The PCS can be rewritten as

$$PCS = P\left(Y_2 > -m_2, \cdots, Y_k > -m_k\right) \ .$$

The apparent paradox in the counterexample can be explained by considering the sensitivity of the PCS with respect to the number of replications.

Theorem 1 (Peng et al. 2015b)

(i) For $k=2$, $\frac{d}{dN_1}PCS>0$ and $\frac{d}{dN_2}PCS>0$;

(ii) For $k\geq 3$, $\frac{d}{dN_i}PCS>0$, $i=2,..,k$, and

$$\frac{d}{dN_1}PCS \geq \sum_{i=2,..,k}\frac{dm_i}{dN_1}\phi(-m_i)\prod_{j\neq i}\Phi(\tilde{m}_j^{(i)})$$
$$+2\sum_{2\leq i<j\leq k}\frac{dc_{ij}}{dN_1}\frac{1}{2\pi\sqrt{1-c_{ij}^2}}\exp\left(-\frac{m_i^2+m_j^2}{2(1+\arcsin(c_{ij}))}\right).$$

Remark. The proof of the theorem can be found in Peng et al. (2015b). The lower bound leads to a sufficient condition for $\frac{d}{dN_1}PCS\geq 0$ or necessary condition for $\frac{d}{dN_1}PCS\leq 0$. An additional allocated replication affects two parts of the PCS: the mean vector of the joint distribution of $(Y_2,..,Y_k)$ and the correlation of the joint normally distributed random variable. For the second through k-th alternatives, both contributions to the PCS are positive, whereas for the first alternative, the first contribution is positive but the second is negative. In the following sections, we will see that in the CVs-DISMAL scenario, the second contribution dominates.

4.1 Analysis for Three Alternatives

For $k=3$, denote $\sigma_3^2=\sigma^2$, $N_3=N$, and

$$\begin{aligned}
&p_1 \doteq \sigma_1/\sigma, \quad p_2 \doteq \sigma_2/\sigma; \quad w_1 \doteq N_1/N, \quad w_2 \doteq N_2/N;\\
&\zeta_1 \doteq p_1^2/w_1, \quad \zeta_2 \doteq p_2^2/w_2;\\
&r \doteq \sigma/(\mu_1-\mu_2), \quad r^{'} \doteq \sigma/(\mu_1-\mu_3), \quad \tau \doteq r/r^{'};\\
&\alpha \doteq r/\sqrt{N}, \quad \alpha^{'} \doteq r^{'}/\sqrt{N}\ .
\end{aligned}$$

The parameter α (or $\alpha^{'}$) is the CV of the difference between $\bar{X}_1$ and $\bar{X}_2$ ($\bar{X}_1$ and $\bar{X}_3$). For $k=3$, whether or not allocating the replication to the best alternative leads to a decrease in the PCS is determined by a one-dimensional parameter.

For calculation of the lower bound given in (1), we need to specify

$$m_2=\frac{\mu_1-\mu_2}{\sqrt{v_1+v_2}}=\frac{1}{\alpha\sqrt{\zeta_1+\zeta_2}}, \quad m_3=\frac{\mu_1-\mu_3}{\sqrt{v_1+v_3}}=\frac{\tau}{\alpha\sqrt{\zeta_1+1}},$$

and

$$c_{23}=\frac{v_1}{\sqrt{v_1+v_2}\sqrt{v_1+v_3}}=\frac{\zeta_1}{\sqrt{\zeta_1+\zeta_2}\sqrt{\zeta_1+1}},$$

$$\begin{aligned}
\tilde{m}_3^{(2)} &= \left(m_2-\frac{m_1v_1}{\sqrt{v_1+v_2}\sqrt{v_1+v_3}}\right)\Big/\sqrt{1-\frac{v_1^2}{(v_1+v_2)(v_1+v_3)}}\\
&= \left(\frac{\tau}{\alpha\sqrt{\zeta_1+1}}-\frac{\zeta_1}{\alpha(\zeta_1+\zeta_2)\sqrt{\zeta_1+1}}\right)\Big/\sqrt{1-\frac{\zeta_1^2}{(\zeta_1+\zeta_2)(\zeta_1+1)}},
\end{aligned}$$

$$\begin{aligned}
\tilde{m}_2^{(3)} &= \left(m_1-\frac{m_2v_1}{\sqrt{v_1+v_2}\sqrt{v_1+v_3}}\right)\Big/\sqrt{1-\frac{v_1^2}{(v_1+v_2)(v_1+v_3)}}\\
&= \left(\frac{1}{\alpha\sqrt{\zeta_1+\zeta_2}}-\frac{\zeta_1}{\alpha(\zeta_1+1)\sqrt{\zeta_1+\zeta_2}}\right)\Big/\sqrt{1-\frac{\zeta_1^2}{(\zeta_1+\zeta_2)(\zeta_1+1)}},
\end{aligned}$$

$$\frac{dm_2}{dN_1} = \frac{\zeta_1}{2\alpha N w_1(\zeta_1+\zeta_2)^{3/2}}, \quad \frac{dm_3}{dN_1} = \frac{\tau\zeta_2}{2\alpha N w_1(\zeta_1+1)^{3/2}}$$
$$\frac{dc_{23}}{dN_1} = -\frac{\zeta_1}{2Nw_1\sqrt{\zeta_1+\zeta_2}\sqrt{\zeta_1+1}}\left(\frac{\zeta_2}{\zeta_2+\zeta_1}+\frac{1}{1+\zeta_1}\right).$$

The lower bound given in (1) can be given explicitly by $\Psi_1(\alpha)/N$, where

$$\Psi_1(\alpha) \doteq \alpha\left[a_1\phi\left(\frac{a_2}{\alpha}\right)\Phi\left(\frac{a_3}{\alpha}\right)+b_1\phi\left(\frac{b_2}{\alpha}\right)\Phi\left(\frac{b_3}{\alpha}\right)\right]-e_1\phi\left(\frac{e_2}{\alpha}\right), \tag{2}$$

and

$$\begin{aligned}
a_1 &\doteq \frac{\zeta_1}{2w_1(\zeta_1+\zeta_2)^{3/2}}, \quad b_1 \doteq \frac{\tau\zeta_2}{2w_1(\zeta_1+1)^{3/2}};\\
a_2 &\doteq \frac{1}{\sqrt{\zeta_1+\zeta_2}}, \quad b_2 \doteq \frac{\tau}{\sqrt{\zeta_1+1}};\\
a_3 &\doteq \frac{(\tau-1)\zeta_1+\tau\zeta_2}{\sqrt{(\zeta_1+\zeta_2)(\zeta_1+\zeta_2+\zeta_1\zeta_2)}}, \quad b_3 \doteq \frac{(1-\tau)\zeta_1+1}{\sqrt{(\zeta_1+1)(\zeta_1+\zeta_2+\zeta_1\zeta_2)}};\\
e_1 &\doteq \frac{\zeta_1}{w_1\sqrt{2\pi(\zeta_1+\zeta_2+\zeta_1\zeta_2)}}\left(\frac{\zeta_2}{\zeta_2+\zeta_1}+\frac{1}{1+\zeta_1}\right),\\
e_2 &\doteq \sqrt{\left(\frac{1}{\zeta_1+\zeta_2}+\frac{\tau^2}{\zeta_1+1}\right)\Big/\left(1+\arcsin\left(\frac{\zeta_1}{\sqrt{\zeta_1+\zeta_2}\sqrt{\zeta_1+1}}\right)\right)}.
\end{aligned} \tag{3}$$

It is easy to see $a_1,a_2,b_1,b_2,e_1,e_2>0$. We have the following conclusion for the determining the sign of the lower bound given in (1) based on Ψ_1 in (2).

Theorem 2 (Peng et al. 2015b)

(i) There exist $M_1>\varepsilon_1>0$ such that for $\alpha\in[0,\varepsilon_1]$, $\Psi_1(\alpha)>0$, and for $\alpha\in[M_1,\infty)$, $\Psi_1(\alpha)<0$;

(ii) there exist $M_2>\varepsilon_2>0$ such that for $\alpha^{'}\in[0,\varepsilon_2]$, $\Psi_2(\alpha^{'})>0$, and for $\alpha^{'}\in[M_2,\infty)$, $\Psi_2(\alpha^{'})<0$, where Ψ_2 can be defined similarly as Ψ_1 in (2).

Remark. The proof of the theorem can be found in Peng et al. (2015b). This theorem tells us that for $k=3$ when α (or $\alpha^{'}$) is large enough, i.e., the CVs-DISMAL scenario, the necessary condition for $\frac{d}{dN_1}PCS<0$ is satisfied; on the other hand, if α (or $\alpha^{'}$) is small enough, the sufficient condition for $\frac{d}{dN_1}PCS>0$ is satisfied.

We provide a graphical illustration on the effect of one additional replication allocated to the first alternative for the counterexample provided in Section 2. The parameters in (3) for this example are given by $\tau=1$ and

$$a_1=b_1=\frac{1}{4\sqrt{2}}; \quad a_2=b_2=\frac{1}{\sqrt{2}}; \quad a_3=b_3=\frac{1}{\sqrt{6}};$$
$$e_1=\frac{1}{\sqrt{6\pi}}, \quad e_2=\sqrt{1\Big/\left(1+\frac{\pi}{6}\right)}.$$

For these values, the lower bound given in (1) changes sign at $\alpha^*\approx 1/0.94$.

Figure 1 shows the two effects of one additional replication allocated to the best alternative: first, it moves the center of the density further towards the first quadrant (shadowed area); second, it changes the

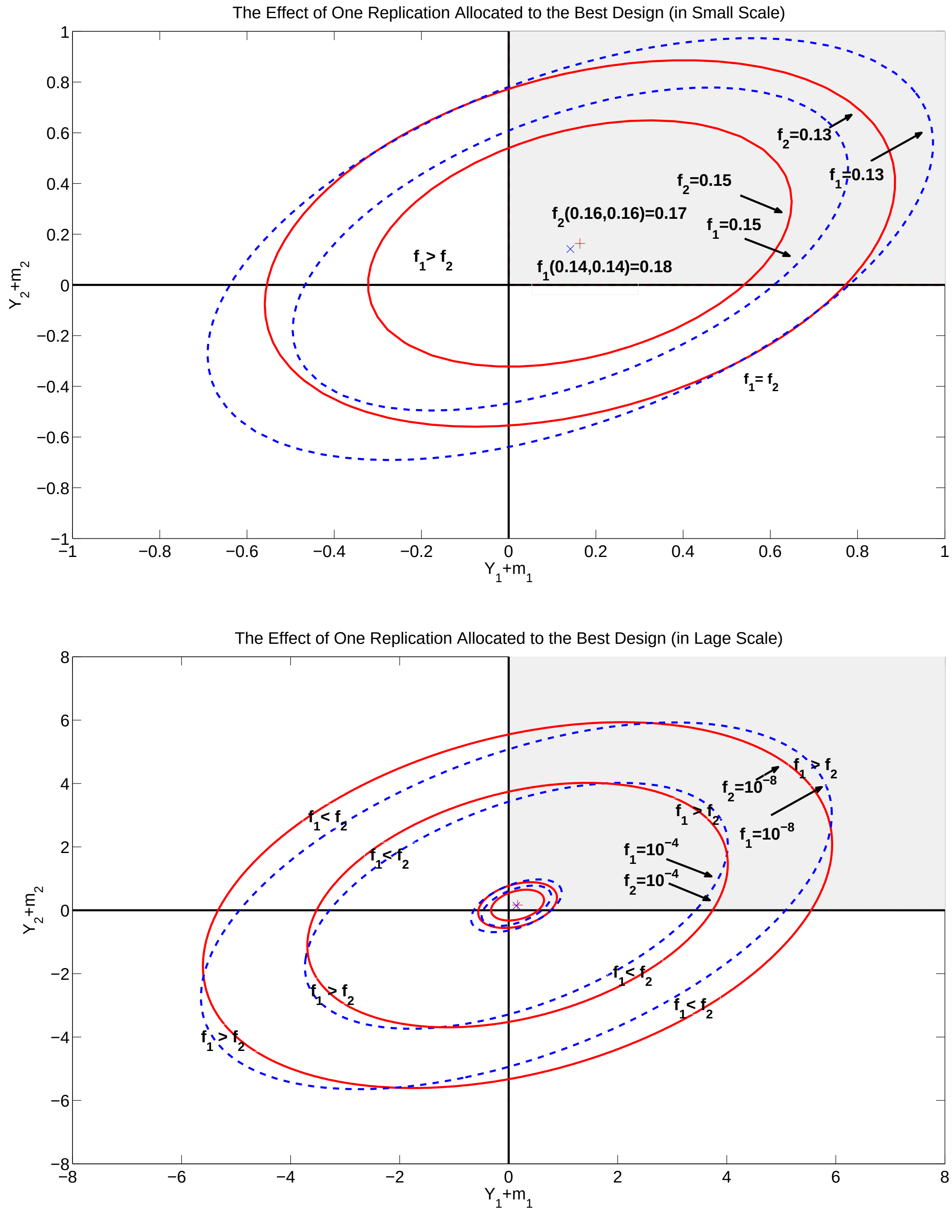

Figure 1: $\Delta = 0.2$. The ellipses with dotted (blue) lines are the contours of the density (i.e., f_1) of (Y_1+m_1, Y_2+m_2) with $N_1 = N_2 = N_3 = 1$; the ellipses with solid (red) lines are the contours of the density (i.e., f_2) of (Y_1+m_1, Y_2+m_2) with $N_1 = 2$ and $N_2 = N_3 = 1$.

shape of the elliptic contours of the density of

$$(Y_1+m_1, Y_2+m_2) = \left(\frac{\bar{X}_1-\bar{X}_2}{\sigma\sqrt{1/N_1+1/N_2}}, \frac{\bar{X}_1-\bar{X}_3}{\sigma\sqrt{1/N_1+1/N_3}}\right),$$

which is governed uniquely by the correlation c_{23} since the variances are normalized to be 1. The dotted (blue) line ellipses are the contours of the density $f_1 = f\left(x_1, x_2;\ 0.2/\sqrt{2}, 0.2/\sqrt{2}, 1/2\right)$, where f is defined by (1), and the solid (red) line ellipses are are the contours of the density $f_2 = f\left(x_1, x_2;\ 0.2/\sqrt{3/2}, 0.2/\sqrt{3/2}, 1/3\right)$.

The PCS is the integral of the density over the first quadrant. From the figure, it is easy to see that moving the center of the density towards the first quadrant would increase the PCS if the shape of the density is kept fixed. However, by analyzing the contours of the density with Y_1+m_1 and Y_2+m_2 coordinates shown in the graph at the top of Figure 1, we know $f_1 > f_2$ inside the ellipse given by the contour $f_2 = 0.13$. This means the density flattens in the center with the additional replication allocated to the first design. Also, by analyzing the contours of the density with Y_1+m_1 and Y_2+m_2 coordinates shown in the graph at the bottom of Figure 1, we know $f_2 > f_1$ in the second and fourth quadrants, while $f_1 > f_2$ on a large proportion of the first and third quadrants. Summarizing, f_1 is greater than f_2 on the area near the center of both densities where the density is highly concentrated because of the exponentially decreasing rate of normal distribution, and on a large proportion of the first quadrant where the PCS is integrated. The decline of the PCS can be explained by the decrease of the induced correlation between Y_1+m_1 and Y_2+m_2.

5 NUMERICAL EXPERIMENTS

We test the following sampling allocation policies: OCBA (Chen et al. 2000) implemented sequentially by the "most starving" rule, which allocates the next sample to the alternative whose current fraction is most below the recommended fraction (Chen and Lee 2011); sequential EVI with 0-1 loss function (Chick et al. 2010); knowledge gradient (KG) with uninformative prior (Frazier et al. 2008); PTV, for which the number of allocated replications to each alternative is proportional to its sample variance, implemented sequentially by the "most starving" rule; equal allocation (EA) sequentially from the first to the last alternative in a cyclical manner. In a two-stage sampling procedure, the parameters are estimated using a fraction of the simulation budget in the first stage. In the following experiments, each alternative has $n_0 = 10$ first-stage replications; in the second stage, for a given allocation policy, we allocate additional replications sequentially and report the PCS as a function of the number of additional replications. For each example, we run 10^6 independent macro experiments, so the PCS is accurate to about three decimal points with 90% confidence.

5.1 Example 1

There are three alternatives with least favorable configuration of true mean, $\mu_1 = 0.001$, $\mu_2 = \mu_3 = 0$, and variances $\sigma_1^2 = 2$, $\sigma_2^2 = \sigma_3^2 = 1$. This example falls into CVs-DISMAL scenario, and the best alternative has the largest variance. From Figure 2, we can see that for OCBA, EVI and PTV, the PCSs are decreasing with the number of additional replications from 1 through 10. After that, the PCSs for KG and PTV continue to decrease, whereas the PCS for OCBA and EVI tilt but the slope is almost flat. The trajectories of OCBA and EVI are indistinguishable in this example. From the trajectory of EA, we can see that each time a replication is allocated to the first (best) alternative, the PCS decreases, and when a replication is allocated to the second or third alternative, the PCS increases.

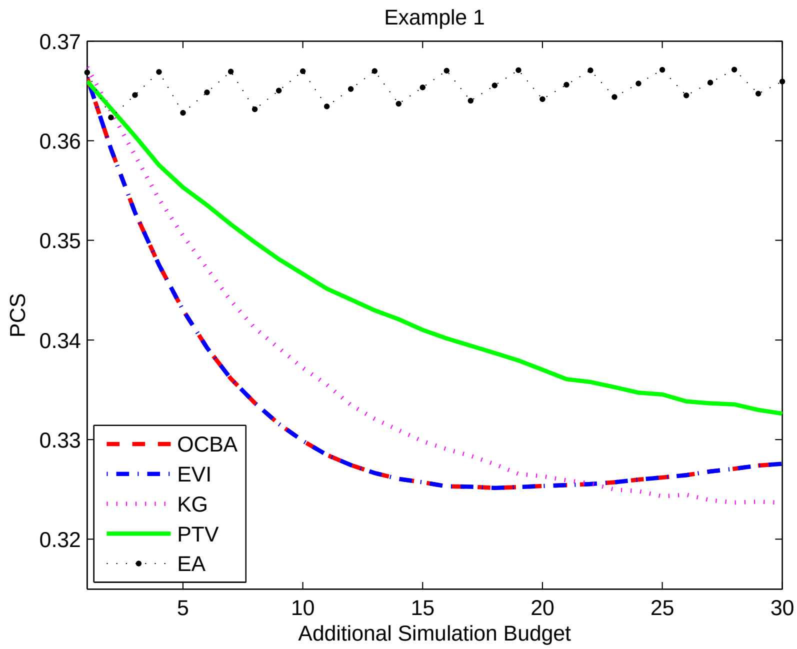

Figure 2: Three alternatives with true means $\mu_1 = 0.001$, $\mu_2 = \mu_3 = 0$, and variances $\sigma_1^2 = 2$, $\sigma_2^2 = \sigma_3^2 = 1$. 10 initial replications are allocated to each alternative to estimate the parameters.

5.2 Example 2

There are ten alternatives with linear structure of true mean, $\mu_i = 10^{-2} - 10^{-3} \times i$, $i = 1,..,10$, and variances $\sigma_i^2 = 2$, $\sigma_{i+5}^2 = 1$, $i = 1,..,5$. This example falls into CVs-DISMAL scenario, and the best alternative has the largest variance. From Figure 3, we can see that for OCBA, EVI and PTV, the PCSs are decreasing with the number of additional replications from 1 through 20. After that, the PCSs for KG and PTV continue to decrease, whereas the PCS for OCBA and EVI tilt upward slowly. Again for EA, each time a replication is allocated to the first (best) alternative, the PCS decreases, and when a replication is allocated to any of the other alternatives, the PCS increases.

5.3 Example 3

There are ten alternatives with least favorable configuration of true means $\mu_1 = 0.001$, $\mu_i = 0$, $i = 2,..,10$, and variances $\sigma_i^2 = 2$, $\sigma_{i+5}^2 = 1$, $i = 1,..,5$. This example also falls into CVs-DISMAL scenario, and the best alternative is among the alternatives that have relatively large variances. The conclusions from the numerical results shown in Figure 4 are similar to those in the previous example.

6 CONCLUSION

In this paper, we investigate and characterize scenarios where the PCS increases with the number of replications allocated to alternatives other than the best. For the best alternative, we provide a necessary condition for the PCS to be decreasing, and the condition can be determined by one factor for the special case of three alternatives. This necessary condition is generally satisfied in the CVs-DISMAL scenario. Numerical experiments confirm that in the CVs-DISMAL scenario, the PCS usually decreases with a replication allocated to the best alternative; furthermore, the PCS of some well-known existing sampling

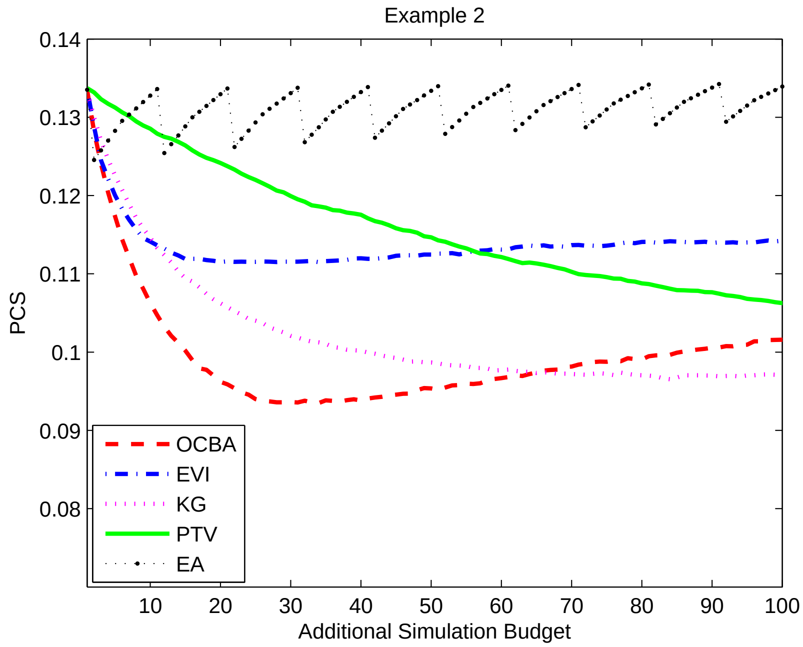

Figure 3: Ten alternatives with true means $\mu_i = 10^{-2} - 10^{-3} \times i$, $i = 1,..,10$, and variances $\sigma_i^2 = 2$, $\sigma_{i+5}^2 = 1$, $i = 1,..,5$. 10 initial replications are allocated to each alternative to estimate the parameters.

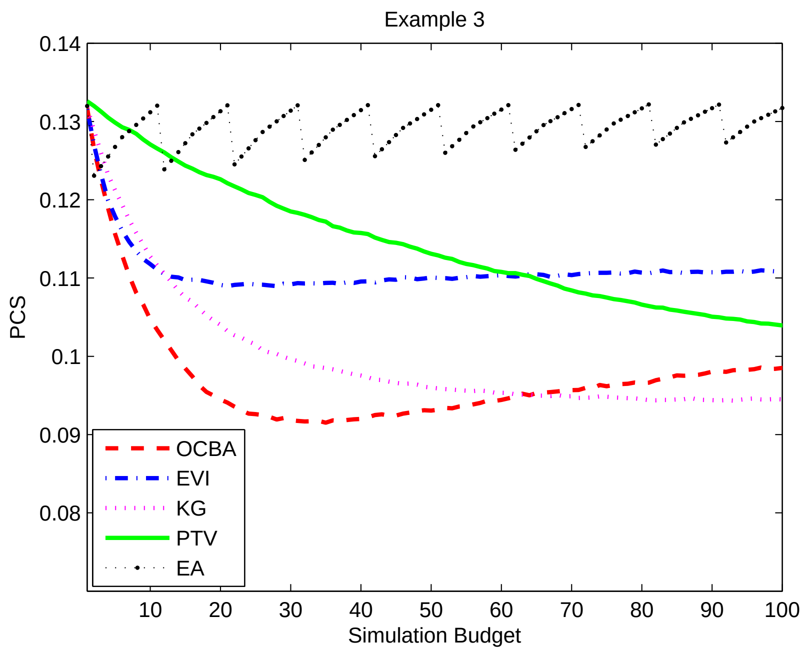

Figure 4: Ten alternatives with true means $\mu_1 = 0.001$, $\mu_i = 0$, $i = 2,..,10$, and variances $\sigma_i^2 = 2$, $\sigma_{i+5}^2 = 1$, $i = 1,..,5$. 10 initial replications are allocated to each alternative to estimate the parameters.

allocation policies, such as OCBA, EVI, KG and PTV, actually decrease in this scenario. More details can be found in Peng et al. (2015b).

REFERENCES

Bechhofer, R. E., T. J. Santner, and D. M. Goldsman. 1995. *Design and Analysis for Statistical Selection, Screening, and Multiple Comparisons*. New York: John Wiley and Sons.

Berger, J. O., and J. Deely. 1988. "A Bayesian approach to ranking and selection of related means with alternatives to analysis-of-variance methodology". *Journal of the American Statistical Association* 83 (402): 364–373.

Branke, J., S. E. Chick, and C. Schmidt. 2007. "Selecting a selection procedure". *Management Science* 53 (12): 1916–1932.

Chen, C.-H., and L. H. Lee. 2011. *Stochastic Simulation Optimization: An Optimal Computing Budget Allocation*. World Scientific Publishing Company.

Chen, C.-H., J. Lin, E. Yücesan, and S. E. Chick. 2000. "Simulation budget allocation for further enhancing the efficiency of ordinal optimization". *Discrete Event Dynamic Systems* 10 (3): 251–270.

Chick, S. E., J. Branke, and C. Schmidt. 2010. "Sequential sampling to myopically maximize the expected value of information". *INFORMS Journal on Computing* 22 (1): 71–80.

Chick, S. E., and K. Inoue. 2001. "New two-stage and sequential procedures for selecting the best simulated system". *Operations Research* 49 (5): 732–743.

Frazier, P. I., W. B. Powell, and S. Dayanik. 2008. "A knowledge-gradient policy for sequential information collection". *SIAM Journal on Control and Optimization* 47 (5): 2410–2439.

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

Gupta, S. S., and K. J. Miescke. 1986. "On the Problem of Finding the Largest Normal Mean under Heteroscedasticity.". Technical report, DTIC Document.

Gupta, S. S., and K. J. Miescke. 1989. "On finding the largest normal mean and estimating the selected mean". Technical report, DTIC Document.

Gupta, S. S., and K. J. Miescke. 1996. "Bayesian look ahead one-stage sampling allocations for selection of the best population". *Journal of Statistical Planning and Inference* 54 (2): 229–244.

Hong, L. J., and B. L. Nelson. 2005. "The tradeoff between sampling and switching: New sequential procedures for indifference-zone selection". *IIE Transactions* 37 (7): 623–634.

Kim, S.-H., and B. L. Nelson. 2006. "Selecting the best system". In *Chapter 17 in Handbooks in Operations Research and Management Science: Simulation*, Volume 13, 501–534. Elsevier.

Pasupathy, R., S. R. Hunter, N. A. Pujowidianto, L. H. Lee, and C.-H. Chen. 2014. "Stochastically constrained ranking and selection via SCORE". *ACM Transactions on Modeling and Computer Simulation (TOMACS)* 25 (1): 1.

Peng, Y., C.-H. Chen, M. C. Fu, and J.-Q. Hu. 2015a. "Dynamic sampling allocation and design selection". *submitted to INFORMS Journal on Computing*:under review.

Peng, Y., C.-H. Chen, M. C. Fu, and J.-Q. Hu. 2015b. "Probability of correct selection: is more sampling always better?". *submitted to Operations Research*:under review.

Rinott, Y. 1978. "On two-stage selection procedures and related probability inequalities". *Communications in Statistics Theory and Methods* 7 (8): 799–811.

Xu, J., C. C. Huang, C.-H. Chen, and L. H. Lee. 2015. "Simulation optimization: a review and exploration in the new era of cloud computing and big data". *Asia-Pacific Journal of Operational Research* 32 (3).

AUTHOR BIOGRAPHIES

YIJIE PENG is a postdoctoral scholar in the School of Management at Fudan University. He received his B.S. degree in applied mathematics from Wuhan University and Ph.D. degree in management science from Fudan University. His research interests lie in simulation methodology, modeling, analysis, and optimization. His email address is pengy10@fudan.edu.cn.

CHUN-HUNG CHEN is a Professor of Systems Engineering and Operations Research at George Mason University and is also affiliated with National Taiwan University. Dr. Chen has led research projects in stochastic simulation and optimization, sponsored by the NSF, NSC, FAA, Air Force, and NASA. He served as Co-Editor of the *Proceedings of the 2002 Winter Simulation Conference* and Program Co-Chair for 2007 Informs Simulation Society Workshop. He has served on the editorial boards of *IEEE Transactions on Automatic Control*, *IEEE Transactions on Automation Science and Engineering*, *IIE Transactions*, *Journal of Simulation Modeling Practice and Theory*, and *International Journal of Simulation and Process Modeling*. He received his Ph.D. degree from Harvard University in 1994. His email address is cchen9@gmu.edu.

MICHAEL C. FU is Ralph J. Tyser Professor of Management Science in the Robert H. Smith School of Business, with a joint appointment in the Institute for Systems Research and affiliate faculty appointment in the Department of Electrical and Computer Engineering, all at the University of Maryland. His research interests include simulation optimization and applied probability, with applications in supply chain management and financial engineering. He has a Ph.D. in applied math from Harvard and degrees in math and EECS from MIT. He served as WSC2011 Program Chair, NSF Operations Research Program Director, *Management Science* Stochastic Models and Simulation Department Editor, and *Operations Research* Simulation Area Editor. He is a Fellow of **INFORMS** and **IEEE**. His email address is mfu@umd.edu.

JIAN-QIANG HU is a Professor with the Department of Management Science, School of Management, Fudan University. He received his B.S. degree in applied mathematics from Fudan University, China, and M.S. and Ph.D. degrees in applied mathematics from Harvard University. He was an Associate Professor with the Department of Mechanical Engineering and the Division of Systems Engineering at Boston University before joining Fudan University. His research interests include discrete-event stochastic systems, simulation, queueing network theory, stochastic control theory, with applications towards supply chain management, risk management in financial markets and derivatives, and communication networks. His e-mail addresses is hujq@fudan.edu.cn.