Proceedings of the 2021 Winter Simulation Conference
S. Kim, B. Feng, K. Smith, S. Masoud, Z. Zheng, C. Szabo, and M. Loper, eds.

GETTING TO "RATE-OPTIMAL" IN RANKING & SELECTION

Harun Avci
Barry L. Nelson
Andreas Wächter

Department of Industrial Engineering & Management Sciences
2145 Sheridan Road
Northwestern University
Evanston, IL 60208 USA

ABSTRACT

In their 2004 seminal paper, Glynn and Juneja formally and precisely established the rate-optimal, probability-of-incorrect-selection, replication allocation scheme for selecting the best of k simulated systems. In the case of independent, normally distributed outputs this allocation has a simple form that depends in an intuitively appealing way on the true means and variances. Of course the means and (typically) variances are unknown, but the rate-optimal allocation provides a target for implementable, dynamic, data-driven policies to achieve. In this paper we compare the empirical behavior of four related replication-allocation policies: mCEI from Chen and Rzyhov and our new gCEI policy that both converge to the Glynn and Juneja allocation; AOMAP from Peng and Fu that converges to the OCBA optimal allocation; and TTTS from Russo that targets the rate of convergence of the posterior probability of incorrect selection. We find that these policies have distinctly different behavior in some settings.

1 INTRODUCTION

Ranking and selection (R&S) is one of the fundamental methods for solving stochastic simulation optimization problems. In the canonical version of the R&S problem, the aim is to identify the single best among a finite number (k) of systems, where the performance of each system can only be estimated using simulation output; here "best" means the maximum or minimum expected value of performance. The ideal R&S procedure either (a) allocates a limited simulation budget so as to maximize the likelihood that the best is identified, or (b) allocates simulation effort as efficiently as possible until a prespecified likelihood is obtained. This paper addresses formulation (a).

The R&S literature contains many policies for version (a) that sequentially obtain replications from systems and adapt as more and more output data are obtained. These policies tend to be Bayesian or Bayesian-inspired, and include versions of optimal computing budget allocation (OCBA, Chen et al. 2000), expected improvement (EI, Jones et al. 1998), knowledge gradient (KG, Frazier and Powell 2007), and multi-armed bandits (MAB, Jamieson and Nowak 2014). In this paper, an *allocation* is the fraction of a fixed budget of replications that is assigned to each simulated system, while a *policy* is an algorithm for sequentially, and usually adaptively, allocating individual replications to systems.

Glynn and Juneja (2004) derived an expression for the asymptotically optimal *static* replication allocation by using large-deviations theory. They represented the replications allocated to system i as $\alpha_i R$, where R is the total budget of replications, $\alpha_i > 0$, and $\sum_{i=1}^{k} \alpha_i = 1$. The policy is "optimal" in a sense that the probability of incorrect selection (PICS) decays exponentially with the best possible exponent as R increases; incorrect selection means choosing any of the $k-1$ inferior systems. Unfortunately, the optimal allocation depends on the underlying output distributions and their parameters, which are typically unknown, and naive plug-in strategies tend not to work well. This leads to the idea of having *adaptive* policies that aggressively pursue the best system in small samples but converge to an "optimal" allocation, such as that of Glynn and Juneja (2004), in the limit.

Recently, several such policies have been proposed. The empirical allocation of the *modified complete expected improvement* (mCEI) policy of Chen and Ryzhov (2019b) converges to the rate-optimal allocation of Glynn and Juneja (2004) under certain conditions. The *asymptotically optimal myopic allocation policy* (AOMAP) of Peng and Fu (2017) converges to the OCBA limiting allocation. And finally, the *top-two Thompson sampling* (TTTS) policy of Russo (2020) seeks the optimal rate of convergence of the posterior probability of incorrect selection to 0; TTTS is one of three "top-two" policies identified by Russo (2020). Notice that AOMAP and TTTS do *not* converge to the Glynn and Juneja (2004) optimal allocation but instead to limits that are arguably desirable.

To this list we add our new *gradient of CEI* (gCEI) policy that attains the same limit as mCEI. We then empirically evaluate the fixed-budget behavior of all four policies under assumptions that support all four: independent, normally distributed output with known variances. Fixed-budget, as opposed to asymptotic behavior is what an analyst actually experiences in practice. It is worth stating that mCEI, gCEI, AOMAP and TTTS can all be "beaten" in some sense by policies specifically designed for good finite-sample performance, especially when the number of systems k is very large; such policies fully eliminate apparently inferior systems quickly, while mCEI, gCEI, AOMAP and TTTS keep all systems in play until the budget is consumed. Nevertheless, they are building blocks for more sophisticated policies so the comparison is relevant.

The remainder of the paper is organized as follows: We provide a brief literature review in Section 2 and formulate the R&S problem in Section 3. We state the AOMAP, mCEI and TTTS policies in Section 4, then introduce gCEI and sketch a proof of its convergence in Section 5. Empirical performance of all four policies is given in Section 6. Finally, Section 7 concludes the paper.

2 ESSENTIAL LITERATURE

The EI criterion was first introduced by Jones et al. (1998) for Bayesian optimization of deterministic simulations. Adapting EI to the R&S problem with independent normal observations, Ryzhov (2016) derived the asymptotic sampling allocation implied by EI and showed that it is related to the OCBA allocation of Chen et al. (2000). Since EI does not achieve an exponential convergence rate, Peng and Fu (2017) proposed a variant of EI, called AOMAP. In the known-variance case, AOMAP achieves the OCBA allocation of Chen et al. (2000) in the limit. Peng and Fu (2017) note that an adjustment can be made to EI to achieve *any* well-defined limiting allocation, including that of Glynn and Juneja (2004), but this requires solving for the limiting allocation on each iteration based on plug-in estimates.

Since EI was originally created for deterministic simulations, it does not directly account for the uncertainty in the output from a stochastic simulation. To incorporate this uncertainty, Salemi et al. (2019) proposed complete expected improvement (CEI). For the R&S problem with independent normal observations, Chen and Ryzhov (2019b) presented a modified CEI policy. Similarly, under more general sampling distributions, Chen and Ryzhov (2019a) proposed the balancing optimal large deviations policy that evaluates the approximate individual large-deviation rate functions and balances them iteratively. Both policies asymptotically achieve the optimal allocations of Glynn and Juneja (2004) when the variances are known, or when variances are unknown but continually updated via plug-in estimators.

More recently, Russo (2020) proposed three different Bayesian policies for adaptively allocating measurement effort in stochastic decision problems including simulation. On every iteration, these policies use the posterior distribution of the output parameter (e.g., mean) to identify the top-two alternatives; one of them is randomly chosen to measure (simulate). The selection probability is a tuning parameter, although Russo (2020) found $1/2$ had robust empirical performance. Top-two probability sampling identifies the two alternatives with the largest posterior probabilities of being optimal. Similarly, top-two value sampling considers the posterior expected value of the difference between the mean of each system and the best of the others. The third version is TTTS; see Thompson (1933) for the origins of Thompson sampling. We employ TTTS with selection probabililty $1/2$ in this paper, and describe it fully below. For the known variances case, Russo (2020) showed that these policies attain the best exponential rate of convergence of the posterior probability of incorrect selection for the true best system when the tuning parameter is set optimally or adjusted adaptively toward the optimal value.

In addition to AOMAP, mCEI and TTTS, we propose a new policy called gCEI that makes replication-allocation decisions based on the gradient of CEI with respect to the number of replications obtained from each system, treating "number of replications" as if it were continuous-valued. Like mCEI, it achieves the optimal allocation of Glynn and Juneja (2004). However, gCEI attains this limit without the need to directly enforce the balance between simulating the best system and the inferior systems, as mCEI does.

3 PRELIMINARIES

Let $\mathscr{S} = \{1,2,\ldots,k\}$ be the set of systems. Each system $i \in \mathscr{S}$ has an unknown mean μ_i. Bigger is better, and unknown to us $\mu_1 \leq \mu_2 \leq \ldots \leq \mu_{k-1} < \mu_k$. Our goal is to find system k which is the unique best.

From a Bayesian perspective, the mean of each system i has a prior distribution $\mu_i \sim \mathrm{N}(\bar{\mu}_i(0), 1/\theta_i(0))$ where $\bar{\mu}_i(0)$ and $\theta_i(0)$ are the prior mean and precision, respectively. The prior mean $\bar{\mu}_i(0)$ represents the initial belief about the true value of μ_i whereas the prior precision $\theta_i(0)$ quantifies the confidence in this belief. We assume that μ_i's are independent of each other under this prior. Notice that we use μ_i to denote the true, fixed means of the systems, and $\bar{\mu}_i(t)$ to denote the posterior mean through iteration t, which is a random variable.

We consider a finite horizon problem with a fixed simulation budget: Let R be the length of our finite horizon. At each iteration $t = 0,1,\ldots,R$, we obtain a single replication $Y(t+1)$ by simulating $x(t)$, an independent and identically distributed $\mathrm{N}(\mu_{x(t)}, \sigma^2_{x(t)})$ random variable with $\sigma_i^2 > 0$ being the variance inherent to the stochastic simulation output for system i. In this paper we assume that σ_i^2's are known and that each system is simulated independently of the others (no common random numbers).

Let $\mathscr{F}^t$ be the sigma-algebra generated by $\{x(\tau), Y(\tau+1)\}_{\tau=0}^{t-1}$. Using the recursive approach in De Groot (1970), the posterior parameters for systems $i \in \mathscr{S}$ at iteration t are

$$\bar{\mu}_i(t+1) = \begin{cases} \dfrac{\bar{\mu}_i(t)\theta_i(t) + Y(t+1)/\sigma_i^2}{\theta_i(t) + 1/\sigma_i^2} & \text{if } x(t) = i \text{ (i.e., if system } i \text{ is simulated at iteration } t) \\ \bar{\mu}_i(t) & \text{if } x(t) \neq i, \end{cases}$$

$$\theta_i(t+1) = \begin{cases} \theta_i(t) + 1/\sigma_i^2 & \text{if } x(t) = i \\ \theta_i(t) & \text{if } x(t) \neq i. \end{cases}$$

Let $r_i(t)$ denote the total number of replications that have been obtained by simulating system i up to iteration t, i.e., $r_i(t) = \sum_{\tau=0}^{t-1} \mathbb{I}_{\{x(\tau)=i\}}$ where $\mathbb{I}_{\{\cdot\}}$ is the indicator function. We employ a non-informative prior (i.e., $\theta_i(0) = 0$). Thus, we have $\bar{\mu}_i(t) = \bar{Y}_i(t)$ and $\theta_i(t) = r_i(t)/\sigma_i^2$ where

$$\bar{Y}_i(t) = \frac{1}{r_i(t)} \sum_{\tau=0}^{t-1} \mathbb{I}_{\{x(\tau)=i\}} Y(\tau+1)$$

is the sample mean of system i.

Let $k(t)$ be the sample-best system at iteration t, $k(t) = \mathrm{argmax}_{i \in \mathscr{S}} \{\bar{\mu}_i(t)\} = \mathrm{argmax}_{i \in \mathscr{S}} \{\bar{Y}_i(t)\}$. We define the (frequentist) probability of correct selection (PCS) at iteration t as $\mathbb{P}\{k(t) = k\}$; thus, the probability of incorrect selection (PICS) is $\mathbb{P}\{k(t) \neq k\}$. These quantities are with respect to the fixed, true means. We can also define corresponding quantities for the posterior probability that system i is or is not the best, which is relevant for TTTS.

A generic adaptive policy is given in Algorithm 1. AOMAP, mCEI, gCEI and TTTS differ in how they decide $x(t)$ in Step 3 to obtain good finite-R and asymptotic $R \to \infty$ performance.

Algorithm 1 Generic Adaptive Policy

1: Let $x(0) = i$ for some $i \in \mathscr{S}$. Obtain $Y(1)$ and update $\mathscr{F}^1$. Also, let $t \leftarrow 1$.
2: **while** $t < R$ **do**
3: Decide to simulate $x(t)$.
4: Obtain $Y(t+1)$ by simulating $x(t)$, update $\mathscr{F}^{t+1} \leftarrow \mathscr{F}^t \cup \{x(t), Y(t+1)\}$ and $t \leftarrow t+1$.
5: **end while**
6: Return $k(R) = \text{argmax}_{i\in\mathscr{S}}\{\bar{\mu}_i(R)\}$ as the selected best system.

4 POLICIES

In this section we summarize three different policies in the recent literature.

4.1 AOMAP

For our R&S problem, the EI for system i at iteration t is

$$\text{EI}_i(t) = \mathbb{E}\left[\max\{\mu_i - \bar{\mu}_{k(t)}(t), 0\}|\mathscr{F}^t\right] = \sqrt{1/\theta_i(t)}\ f\left(\frac{\bar{\mu}_i(t) - \bar{\mu}_{k(t)}(t)}{\sqrt{1/\theta_i(t)}}\right)$$

where $f(z) = z\Phi(z) + \phi(z)$ with ϕ and Φ being the standard normal probability density and cumulative distribution functions, respectively. Ryzhov (2016) shows that EI does not precisely achieve the OCBA allocation as the allocations to inferior systems converge to zero. Under a Bayesian framework, Peng and Fu (2017) propose a myopic allocation policy, called AOMAP, as a new variant of EI, and show that AOMAP does achieve the OCBA allocation when the variances are known. Under this policy Step 3 becomes

$$x(t) = \text{argmax}_{i\in\mathscr{S}}\{\mathbb{E}\left[\max\{\mu_i - A_i(t), 0\}|\mathscr{F}^t\right]\} = \text{argmax}_{i\in\mathscr{S}}\left\{\sqrt{1/\theta_i(t)}\ f\left(\frac{\bar{\mu}_i(t) - A_i(t)}{\sqrt{1/\theta_i(t)}}\right)\right\}$$

where $A_i(t) = \bar{\mu}_{k(t)}(t)\ \mathbb{I}_{\{\bar{\mu}_i(t)\neq\bar{\mu}_{k(t)}(t)\}} + (\bar{\mu}_{k(t)}(t) + \xi_{k(t)}(t)\sigma_{k(t)})\ \mathbb{I}_{\{\bar{\mu}_i(t)=\bar{\mu}_{k(t)}(t)\}}$, and

$$\xi_{k(t)}(t) = \left(\sum_{i\in\mathscr{S}\setminus\{k(t)\}} \frac{\sigma_{k(t)}^2\sigma_i^2}{[\bar{\mu}_i(t) - \bar{\mu}_{k(t)}(t)]^4}\right)^{-1/4}.$$

Notice that if $\xi_{k(t)}(t) = 0$, then $A_i(t) = \bar{\mu}_{k(t)}(t)$, and thus the expectation becomes $\text{EI}_i(t)$. Since EI is too greedy in allocating to the best system, the additional term adjusts the allocation to the best system to make it less favorable as the number of iterations approaches infinity. This adjustment enables AOMAP to achieve the OCBA limiting allocation.

4.2 mCEI Policy

Since EI does not fully capture the uncertainty in the output from a stochastic simulation, Salemi et al. (2019) introduced CEI in a Gaussian Markov random field framework for discrete simulation optimization. For our R&S problem, CEI for system $i \neq k(t)$ at iteration t is

$$\text{CEI}_i(t) = \mathbb{E}\left[\max\{\mu_i - \mu_{k(t)}, 0\}|\mathscr{F}^t\right] = \sqrt{1/\theta_i(t) + 1/\theta_{k(t)}(t)}\ f\left(\frac{\bar{\mu}_i(t) - \bar{\mu}_{k(t)}(t)}{\sqrt{1/\theta_i(t) + 1/\theta_{k(t)}(t)}}\right).$$

Chen and Ryzhov (2019b) present the mCEI policy for R&S, which is a modified version of the original CEI policy of Salemi et al. (2019). Under mCEI, $x(t) = k(t)$ if

$$\left(\frac{r_{k(t)}(t)}{\sigma_{k(t)}}\right)^2 < \sum_{i\in\mathscr{S}\setminus\{k(t)\}}\left(\frac{r_i(t)}{\sigma_i}\right)^2. \tag{1}$$

Otherwise, $x(t) = c(t)$ where $c(t) = \text{argmax}_{i \neq k(t)} \text{CEI}_i(t)$. Condition (1) enforces, in the limit, the balance condition in the optimal allocation of Glynn and Juneja (2004).

4.3 TTTS Policy

As TTTS involves more than a simple substitution for Step 3 in Algorithm 1, we provide the new Step 3 as Algorithm 2. The asymptotically best performance of TTTS is obtained by tuning β toward an optimal value. However, Russo (2020) obtained good empirical performance by the simple choice of $\beta = 1/2$.

Algorithm 2 TTTS Step 3

Sample $\widehat{\mu}_i \sim \text{N}(\bar{\mu}_i(t), 1/\theta_i(t))$ for $i \in \mathscr{S}$ and set $I \leftarrow \text{argmax}_{i \in \mathscr{S}} \widehat{\mu}_i$. ▷ Thompson sampling
Sample $B \sim \text{Bernoulli}(\beta)$.
if $B = 1$ **then**
  $x(t) = I$.
else
  repeat
    Sample $\widehat{\mu}_j \sim \text{N}(\bar{\mu}_j(t), 1/\theta_j(t))$ for $j \in \mathscr{S}$ and set $J \leftarrow \text{argmax}_{j \in \mathscr{S}} \widehat{\mu}_j$.
  until $J \neq I$
  $x(t) = J$.
end if

5 gCEI POLICY

EI has been shown to be an effective search strategy in Bayesian optimization of deterministic simulations; CEI extends EI to stochastic simulation; and mCEI tailors CEI to obtain optimal asymptotic performance in R&S by insuring that the necessary balance between simulating the best system and the rest is achieved in the limit; a pure CEI policy never simulates the current sample best in the next iteration.

One feature of CEI-based simulation-optimization methods such as GMIA in Salemi et al. (2019) is that CEI identifies promising solutions, but not how many replications to expend on them. gCEI grew out of an ongoing investigation of employing CEI for that purpose by exploiting its gradient with respect to the number of replications treating the number of replications as if it was continuous. Here we use it simply to decide how to allocate the next *single* replication, as with the other policies.

To derive an expression for the gradient of CEI, first notice that the derivative of f with respect to z is $f'(z) = \Phi(z)$. Then, for $i \neq k(t)$,

$$\frac{\partial}{\partial r_i(t)} \left(\frac{1}{\theta_i(t)} \right) = -\frac{\sigma_i^2}{(r_i(t))^2} \text{ and } \frac{\partial \bar{\mu}_i(t)}{\partial r_i(t)} = \frac{\partial \bar{Y}_i(t)}{\partial r_i(t)} = 0.$$

To simplify notation, let $\nu_i = 1/\theta_i(t) + 1/\theta_{k(t)}(t)$. Then we have

$$\frac{\partial}{\partial r_i(t)} \left(\frac{\bar{\mu}_i(t) - \bar{\mu}_{k(t)}(t)}{\sqrt{\nu_i}} \right) = \frac{\left(\bar{Y}_i(t) - \bar{Y}_{k(t)}(t)\right)}{2\nu_i\sqrt{\nu_i}} \frac{\sigma_i^2}{(r_i(t))^2}.$$

Thus, the derivative of $\text{CEI}_i(t)$ with respect to $r_i(t)$ is

$$\begin{aligned} &\frac{\partial \text{CEI}_i(t)}{\partial r_i(t)} \\ &= \quad -\frac{\sigma_i^2}{2\sqrt{\nu_i}(r_i(t))^2} \, f\left(\frac{\bar{Y}_i(t) - \bar{Y}_{k(t)}(t)}{\sqrt{\nu_i}} \right) + \sqrt{\nu_i} \left[\frac{\left(\bar{Y}_i(t) - \bar{Y}_{k(t)}(t)\right)}{2\nu_i\sqrt{\nu_i}} \frac{\sigma_i^2}{(r_i(t))^2} \right] \, f'\left(\frac{\bar{Y}_i(t) - \bar{Y}_{k(t)}(t)}{\sqrt{\nu_i}} \right) \end{aligned}$$

$$= \quad -\frac{\sigma_i^2}{2\sqrt{\nu_i}(r_i(t))^2}\left[\frac{\bar{Y}_i(t)-\bar{Y}_{k(t)}(t)}{\sqrt{\nu_i}}\;\Phi\left(\frac{\bar{Y}_i(t)-\bar{Y}_{k(t)}(t)}{\sqrt{\nu_i}}\right)+\phi\left(\frac{\bar{Y}_i(t)-\bar{Y}_{k(t)}(t)}{\sqrt{\nu_i}}\right)\right.$$
$$\left. -\frac{\left(\bar{Y}_i(t)-\bar{Y}_{k(t)}(t)\right)}{\sqrt{\nu_i}}\;\Phi\left(\frac{\bar{Y}_i(t)-\bar{Y}_{k(t)}(t)}{\sqrt{\nu_i}}\right)\right]$$
$$= \quad -\frac{\sigma_i^2}{(r_i(t))^2}\;\frac{1}{2\sqrt{\nu_i}}\;\phi\left(\frac{\bar{Y}_i(t)-\bar{Y}_{k(t)}(t)}{\sqrt{\nu_i}}\right)\leq 0.$$

Proceeding similarly,

$$\frac{\partial \mathrm{CEI}_i(t)}{\partial r_{k(t)}(t)} = -\frac{\sigma_{k(t)}^2}{(r_{k(t)}(t))^2}\;\frac{1}{2\sqrt{\nu_i}}\;\phi\left(\frac{\bar{Y}_i(t)-\bar{Y}_{k(t)}(t)}{\sqrt{\nu_i}}\right)\leq 0$$

whereas $\partial \mathrm{CEI}_i(t)/\partial r_j(t) = 0$ for $j \in \mathscr{S}\setminus\{i,k(t)\}$ because $\partial(\theta_i(t))^{-1}/\partial r_j(t) = 0$.

Since simulating the sample best $k(t)$ at iteration t affects all CEI's, the *total* impact of simulating $k(t)$ is $\sum_{i\in\mathscr{S}\setminus\{k(t)\}}\partial \mathrm{CEI}_i(t)/\partial r_{k(t)}(t)$. As lower CEI values are better, then from among the systems other than $k(t)$, simulating

$$g(t) = \mathrm{argmin}_{i\in\mathscr{S}\setminus\{k(t)\}}\frac{\partial \mathrm{CEI}_i(t)}{\partial r_i(t)}$$

has potentially the most improvement. To make a decision as to which system to simulate next, $k(t)$ or $g(t)$, we propose the following condition:

$$\sum_{i\in\mathscr{S}\setminus\{k(t)\}}\frac{\partial \mathrm{CEI}_i(t)}{\partial r_{k(t)}(t)} \stackrel{?}{\leq} \min_{i\in\mathscr{S}\setminus\{k(t)\}}\frac{\partial \mathrm{CEI}_i(t)}{\partial r_i(t)} = \frac{\partial \mathrm{CEI}_{g(t)}(t)}{\partial r_{g(t)}(t)}.$$

If this condition holds, then the total impact of simulating $k(t)$ is potentially greater than simulating $g(t)$, and thus we prefer simulating $k(t)$ to $g(t)$, i.e., $x(t) = k(t)$. On the other hand, if the condition does not hold, then we prefer simulating $g(t)$, i.e., $x(t) = g(t)$. This leads to the gCEI policy in Algorithm 3.

Algorithm 3 gCEI Policy

Let $x(0) = i$ for some $i \in \mathscr{S}$. Obtain $Y(1)$ and update $\mathscr{F}^1$. Also, let $t \leftarrow 1$.
while $t < R$ **do**
  if

$$\sum_{i\in\mathscr{S}\setminus\{k(t)\}}\partial \mathrm{CEI}_i(t)/\partial r_{k(t)}(t) \leq \min_{i\in\mathscr{S}\setminus\{k(t)\}}\{\partial \mathrm{CEI}_i(t)/\partial r_i(t)\}$$

  then
    $x(t) = k(t)$.
  else
    $x(t) = g(t)$ where $g(t) = \mathrm{argmin}_{i\in\mathscr{S}\setminus\{k(t)\}}\{\partial \mathrm{CEI}_i(t)/\partial r_i(t)\}$.
  end if
  Obtain $Y(t+1)$ by simulating $x(t)$, update $\mathscr{F}^{t+1} \leftarrow \mathscr{F}^t \cup \{x(t), Y(t+1)\}$ and $t \leftarrow t+1$.
end while
Return $k(R) = \mathrm{argmax}_{i=1,\ldots,k}\{\bar{\mu}_i(R)\}$ as the selected best system.

Here we provide a sketch of the proof that gCEI converges to the Glynn and Juneja (2004) rate-optimal allocation. First, it is easy to show that as $R \to \infty$ the gCEI policy will simulate all systems infinitely often.

Mimicing the analysis in Ryzhov (2016), we consider the deterministic behavior of gCEI with the true means inserted for the estimates. This implies that $k(t) = k$ for all t and

$$\frac{\partial \mathrm{CEI}_i(t)}{\partial r_i(t)} = -\frac{\sigma_i^2}{(r_i(t))^2}\,\frac{1}{2\sqrt{\nu_i}}\,\phi\left(\frac{\mu_i-\mu_k}{\sqrt{\nu_i}}\right) \text{ and } \frac{\partial \mathrm{CEI}_i(t)}{\partial r_k(t)} = -\frac{\sigma_k^2}{(r_k(t))^2}\,\frac{1}{2\sqrt{\nu_i}}\,\phi\left(\frac{\mu_i-\mu_k}{\sqrt{\nu_i}}\right).$$

Consider the empirical allocation $\{r_i(t)/t,\ i = 1,2,\ldots,k\}$. We know that it must have a convergent subsequence, $r_i(t)/t \overset{t\to\infty}{\longrightarrow} \alpha_i$; we show that any such subsequence must converge to the rate-optimal allocation (a complete proof includes showing that the limit of the subsequence is not 0). Let $\nu_i' = (\sigma_i^2/\alpha_i + \sigma_k^2/\alpha_k) = \lim_{t\to\infty} t\nu_i$.

First consider the sub-subsequence on which the inequality in Step 3 holds. For such iterations

$$\sum_{i\neq k} \frac{\sigma_k^2/(r_k(t))^2}{\sigma_j^2/(r_j(t))^2}\sqrt{\frac{\nu_j}{\nu_i}} \times \exp\left\{-\frac{1}{2}\left(\frac{(\mu_i-\mu_k)^2}{\nu_i} - \frac{(\mu_j-\mu_k)^2}{\nu_j}\right)\right\} \geq 1 \text{ for any } j\neq k. \tag{2}$$

However, as $t\to\infty$, the exponential term will go to 0 or ∞ unless

$$\frac{(\mu_i-\mu_k)^2}{\nu_i'} = \frac{(\mu_j-\mu_k)^2}{\nu_j'},\ i\neq j\neq k. \tag{3}$$

Thus, for Equation (2) to hold for any j, Equation (3) must hold, which is the first of two conditions for the rate-optimal allocation of Glynn and Juneja (2004). Therefore, as t increases, Inequality (2) becomes (after some manipulation)

$$\sum_{i\neq k} \frac{\sigma_k^2/(r_k(t))^2}{\sigma_j^2/(r_j(t))^2}\sqrt{\frac{1}{\nu_i}} \geq \sqrt{\frac{1}{\nu_j}} \text{ for any } j\neq k. \tag{4}$$

Summing both sides over $j = 1,2,\ldots,k-1$, dividing out the common term, and letting $t\to\infty$ gives

$$\sum_{j\neq k}\left(\frac{\sigma_k}{\alpha_k}\right)^2\left(\frac{\alpha_j}{\sigma_j}\right)^2 \geq 1. \tag{5}$$

Next consider the sub-subsequence on which inferior system $j\neq k$ is chosen in Step 3. This reverses the inequality in (4), and must be true for each $j\neq k$. Then a similar argument shows that the left-hand side of (5) must be ≤ 1. Therefore, equality is required, which is the second condition of Glynn and Juneja (2004).

6 EMPIRICAL PERFORMANCE

We ran 16 experiments in total, including four different values of number of systems $k\in\{5,10,20,30\}$. For each k, we set $\mu_i = cm_i$ where the m_i's are prescaled true mean values provided in Table 1 and c is a scaling constant we explain below. In the slippage and ascending means configurations, the systems have equal variances. In the other two configurations the means are ascending, but the variances are proportional to, and inversely proportional to, the prescaled mean values. Notice that Figure 1 cited below is found in the body of the paper, while Figures 2—5 are in Appendix A.

In each experiment we first allocate 2 replications to each system before applying any policy. To create sensible cases, we scaled the true means so that at least r_0 replications will be consumed before the difference between the best and second-best systems is one standard error of their estimated difference under the Glynn and Juneja (2004) rate-optimal policy. Specifically,

$$\mu_k - \mu_{k-1} = c(m_k - m_{k-1}) = \sqrt{\frac{\sigma_{k-1}^2}{r_0\alpha_{k-1}^\star} + \frac{\sigma_k^2}{r_0\alpha_k^\star}}.$$

Table 1: Configurations for experiments.

Configuration	Prescaled true mean values	True standard deviations
Slippage	$m_i = -1$ for $i \neq k$ and $m_k = 0$	$\sigma_i = 1$
Ascending mean	$m_i = \log(i)$	$\sigma_i = 1$
Ascending variance	$m_i = \log(i+1)$	$\sigma_i = \sqrt{m_i}$
Descending variance	$m_i = \log(i+1)$	$\sigma_i = 1/\sqrt{m_i}$

Thus, we control how quickly the best system becomes distinguishable from the others. To find c satisfying the equation, we first calculate the $\alpha_i^\star$ by solving the expression for the rate-optimal allocation of Glynn and Juneja (2004) with the m_i's from Table 1. The constant c does not change the optimal allocation because scaling all μ_i's or all σ_i's does not have any impact. At the same time, we want our total simulation budget R to be large enough so that we can observe the convergence behavior of the policies. We set $r_0 = 20k$ and $R = 100k$. Lastly, we set the number of macro-replications M to 5000 to be able to estimate PICS to two decimal places over a range of values. To measure the performance of each policy, we report $\widehat{\text{PICS}}(t) = \sum_{\tau=1}^{t} \mathbb{I}_{\{k(\tau)\neq k\}}/t$, $\widehat{\alpha}_k(t) = r_k(t)/t$, and the mean and standard deviation of $\mu_k - \mu_{k(t)}$, the optimality gap, at each iteration t.

Figures 1–2 exhibit results for the slippage configurations with $k = 5$ and $k = 30$, respectively. To observe the tail behavior more closely, Figure 3 shows the results with a larger budget of $R = 5000$ for $k = 5$. We summarize our key observations: Under each policy, PICS converges to zero as expected. However, the convergence behaviors are not always the same, and it appears that gCEI performs as well or better than the other policies. Both gCEI and mCEI first overshoot the asymptotically optimal allocation of Glynn and Juneja (2004) for the best system, but then converge as expected in the long run. gCEI tends to allocate less to the best system than mCEI. On the other hand, TTTS allocates much more to the best system, which makes sense given its heritage in MAB and minimizing regret. Remember that the limiting allocations for AOMAP and TTTS are *not* those of Glynn and Juneja, so we do not expect the same allocations.

For the slippage configuration the mean optimality gap is a scaled version of PICS because the gap is the same whenever any inferior system is selected as the best. The standard deviation of the optimality gap shows what might be considered unexpected behavior as it first increases and then decreases. This is because the best system is not distinguishable with a small budget and the inferior systems all have the same mean values. As the best system becomes more recognizable, the variability increases up to a point, then decreases as each policy becomes more sure of the identity of the best system.

To understand how the dynamic policies behave relative to employing the *asymptotically* optimal allocation of Glynn and Juneja (2004) from the beginning, we compare gCEI with an unrealistic policy where the Glynn and Juneja allocation is known and is applied starting from the first iteration in the slippage configuration with $k = 5$. More specifically, under this unrealistic policy, two replications should be allocated to the best system for each replication allocated to an inferior system. Figure 4 exhibits the result of this comparison where we only report iterations that are a multiple of six. gCEI performs better than this unrealistic policy even though it overallocates to the best system for a while. This emphasizes that the rate-optimal allocations address large-sample, not small-sample, behavior.

Lastly, Figure 5 exhibits results for the ascending variance configuration with $k = 10$. We do not report the results for the other configurations as they are so similar to this one. Here all policies perform similarly, based on our metrics. The only difference appears in their allocations to the best system. Similar to the slippage configuration, gCEI allocates less to the best system than mCEI. However, in contrast to the slippage configuration, mCEI and gCEI do not overshoot the asymptotically optimal allocation of Glynn and Juneja (2004).

The slippage configuration is certainly unrealistic, but it represents a situation in which there are many close competitors to the best. In this setting gCEI seems to have some advantages. When the means are ascending it appears to be easier for all policies to control the PICS and optimality gap because the inferior systems are easier to identify; we found that all policies tended to allocate the majority of their replications to the top two systems in these settings.

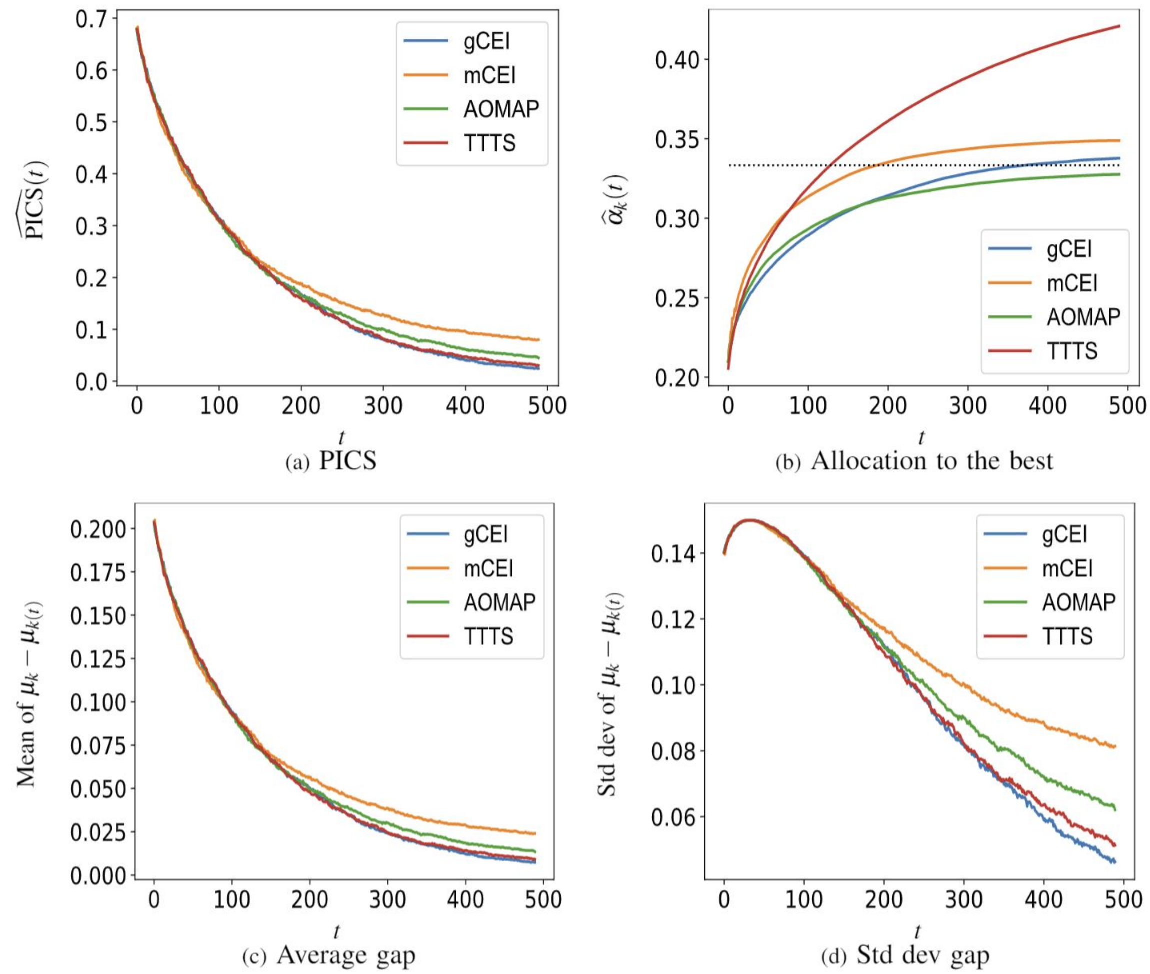


Figure 1: The slippage configuration with $k = 5$, $R = 500$ and $M = 5000$. The dotted line in (b) is the Glynn & Juneja optimal allocation to the best system.

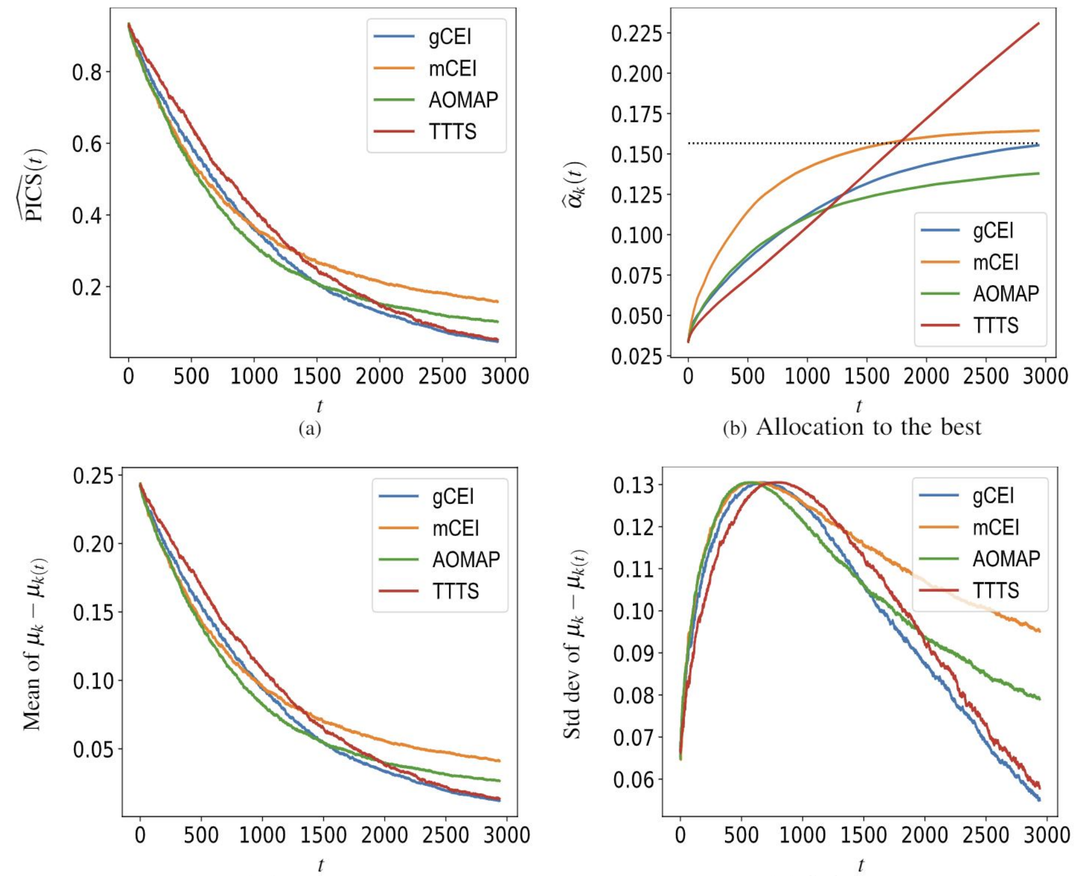
7 CONCLUSIONS

In this paper we examined three recent policies, and one new policy, for assigning replications to systems in the fixed-budget R&S problem. All of the policies adapt as they obtain additional simulation outputs, and each policy achieves a form of optimal allocation as the budget increases; they differ in their definition of "optimal" and their small-sample behavior. Looking at PICS, and the mean and standard deviation of the optimality gap at termination, gCEI appears to perform as well or better than AOMAP, mCEI and TTTS. Our comparisons did not consider computational effort (other than replications) or the ability to stop with a prespecified PCS, measures that also distinguish R&S procedures.

ACKNOWLEDGEMENTS

This research was partially supported by National Science Foundation Grant Number DMS-1854562.

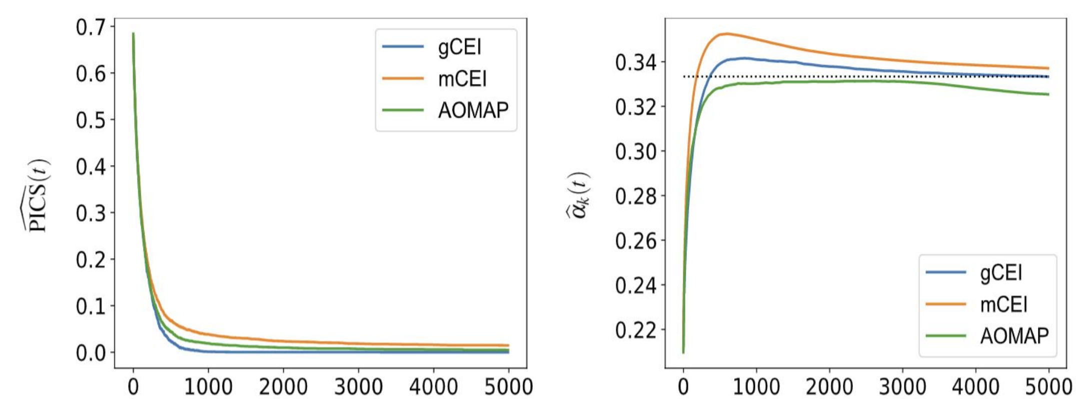
A FIGURES



(a) (b) Allocation to the best

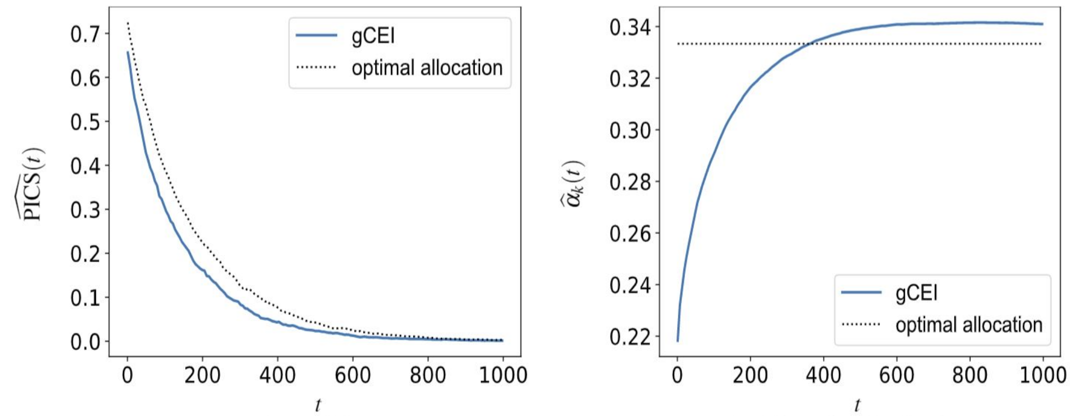
(c) Average gap (d) Std dev gap

Figure 2: The slippage configuration with $k = 30$, $R = 3000$ and $M = 5000$. The dotted line in (b) is the Glynn & Juneja optimal allocation to the best system.



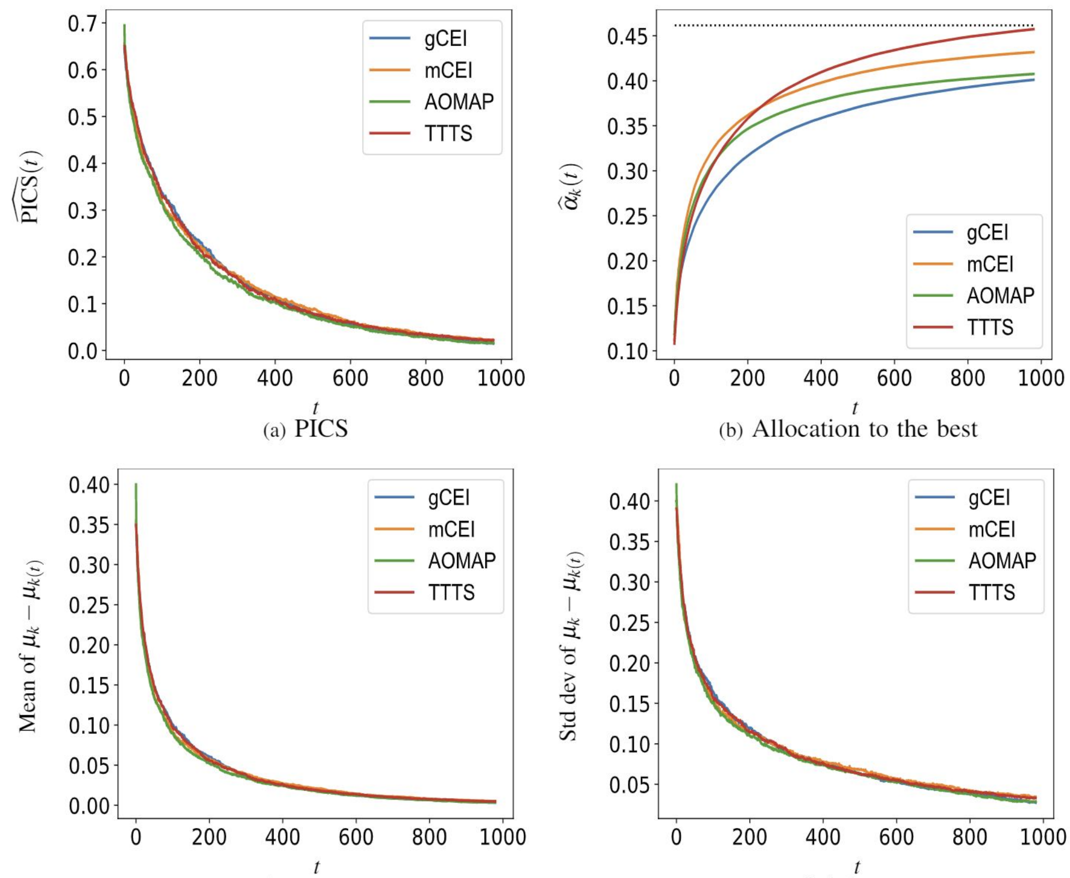
(a) PICS (b) Allocation to the best

Figure 3: The slippage configuration with $k = 5$, $R = 5000$ and $M = 5000$. The dotted line in (b) is the Glynn & Juneja optimal allocation to the best system.



(a) PICS (b) Allocation to the best

Figure 4: The slippage configuration with $k = 5$, $R = 1000$ and $M = 5000$. The dotted line in (b) is the Glynn & Juneja optimal allocation to the best system.



(a) PICS (b) Allocation to the best

(c) Average gap (d) Std dev gap

Figure 5: The ascending variance configuration with $k = 10$, $R = 1000$ and $M = 5000$. The dotted line in (b) is the Glynn & Juneja optimal allocation to the best system.

REFERENCES

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.

Chen, Y., and I. O. Ryzhov. 2019a. "Balancing Optimal Large Deviations in Ranking and Selection". In *Proceedings of the 2019 Winter Simulation Conference*, edited by N. Mustafee, K.-H. G. Bae, S. Lazarova-Molnar, M. Rabe, C. Szabo, P. Haas, and Y.-J. Son, 3368–3379. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Chen, Y., and I. O. Ryzhov. 2019b. "Complete Expected Improvement Converges to an Optimal Budget Allocation". *Advances in Applied Probability* 51(1):209–235.

De Groot, M. H. 1970. *Optimal Statistical Decisions*. New York: John Wiley & Sons.

Frazier, P., and W. Powell. 2007. "The Knowledge Gradient Policy for Offline Learning with Independent Normal Rewards". In *2007 IEEE International Symposium on Approximate Dynamic Programming and Reinforcement Learning*, 143–150. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Glynn, P., and S. Juneja. 2004. "A Large Deviations Perspective on Ordinal Optimization". In *Proceedings of the 2004 Winter Simulation Conference*, edited by R. G. Ingalls, M. D. Rossetti, J. S. Smith, and B. A. Peters, 577–585. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Jamieson, K., and R. Nowak. 2014. "Best-Arm Identification Algorithms for Multi-Armed Bandits in the Fixed Confidence Setting". In *2014 48th Annual Conference on Information Sciences and Systems (CISS)*, 1–6. Princeton, New Jersey: Princeton University.

Jones, D. R., M. Schonlau, and W. J. Welch. 1998. "Efficient Global Optimization of Expensive Black-Box Functions". *Journal of Global Optimization* 13(4):455–492.

Peng, Y., and M. C. Fu. 2017. "Myopic Allocation Policy with Asymptotically Optimal Sampling Rate". *IEEE Transactions on Automatic Control* 62(4):2041–2047.

Russo, D. 2020. "Simple Bayesian Algorithms for Best-Arm Identification". *Operations Research* 68(6):1625–1647.

Ryzhov, I. O. 2016. "On the Convergence Rates of Expected Improvement Methods". *Operations Research* 64(6):1515–1528.

Salemi, P. L., E. Song, B. L. Nelson, and J. Staum. 2019. "Gaussian Markov Random Fields for Discrete Optimization via Simulation: Framework and Algorithms". *Operations Research* 67(1):250–266.

Thompson, W. R. 1933. "On the Likelihood That One Unknown Probability Exceeds Another in View of the Evidence of Two Samples". *Biometrika* 25(3/4):285–294.

AUTHOR BIOGRAPHIES

HARUN AVCI is a PhD candidate in the Department of Industrial Engineering and Management Sciences at Northwestern University. His research interests include simulation optimization and simulation methodology. His e-mail address is harun.avci@u.northwestern.edu.

BARRY L NELSON is the Walter P. Murphy Professor in the Department of Industrial Engineering & Management Sciences at Northwestern University. He is a Fellow of INFORMS and IIE, and is the author of *Foundations and Methods of Stochastic Simulation: A First Course*, from Springer. His e-mail and web addresses are nelsonb@northwestern.edu and http://www.iems.northwestern.edu/~nelsonb, respectively.

ANDREAS WÄCHTER is a Professor in the Department of Industrial Engineering and Management Sciences at Northwestern University. His research centers on the design, analysis, implementation, and application of numerical algorithms for nonlinear optimization. He is a recipient of the J. H. Wilkinson Prize for Numerical Software for the Ipopt open-source optimization package, and a Fellow of SIAM. His e-mail address is andreas.waechter@northwestern.edu.