

THE CONJUNCTION OF THE KNOWLEDGE GRADIENT AND THE ECONOMIC APPROACH TO SIMULATION SELECTION

Stephen E. Chick

Technology & Operations Management Area
INSEAD
Boulevard de Constance
77300 Fontainebleau, FRANCE

Peter Frazier

Dept. Operations Research & Financial Engineering
Princeton University
Engineering Quadrangle, Olden St.
Princeton, NJ, 08544, U.S.A.

ABSTRACT

This paper deals with the selection of the best of a finite set of systems, where best is defined with respect to the maximum mean simulated performance. We extend the ideas of the knowledge gradient, which accounts for the expected value of one stage of simulation, by accounting for the future value of the option to simulate over multiple stages. We extend recent work on the economics of simulation, which studied discounted rewards, by balancing undiscounted simulation costs and the expected value of information from simulation runs. This contribution results in a diffusion model for comparing a single simulated system with a standard that has a known expected reward, and new stopping rules for fully sequential procedures when there are multiple systems. These stopping rules are more closely aligned with the expected opportunity cost allocations that are effective in numerical tests. We demonstrate an improvement in performance over previous methods.

1 INTRODUCTION

This paper focuses on the use of sampling to select the best of a finite set of simulated systems, where ‘best’ is defined in terms of the maximum expected value of the simulation output. Quite a bit has been written on this topic from both the frequentist (e.g., see [Kim and Nelson 2006](#) for a review) and Bayesian perspectives (e.g., see [Chick 2006](#) for a review). This paper takes a Bayesian perspective, as this perspective has shown a high degree of effectiveness in sampling to quickly identify the best system ([Branke, Chick, and Schmidt 2007](#)).

Selection procedures use allocation rules to identify how many and from which system simulation samples should be drawn next, stopping rules to indicate whether sampling should continue or should stop, and selection rules that identify which system to pick when sampling has stopped. Early work on Bayesian selection procedures focuses on allocation rules to improve sampling efficiency by seeking optimal allocations for two-stage procedures, where possible, or on finding useful heuristics or asymptotic properties ([Gupta and Miescke 1996](#), [Chen 1996](#), [Chick and Inoue 2001](#)). Later work explores a greater variety of stopping rules, and finds that the choice of stopping rule can strongly influence the performance of an allocation ([Branke et al. 2007](#)). Stopping rules include sampling to exhaust a finite budget and adaptive stopping rules that continue to sample, as a function of the data, until a statistical threshold for the posterior expected loss is achieved.

[Frazier, Powell, and Dayanik \(2008\)](#) employ a dynamic programming framework to analyze a greedy allocation rule introduced by [Gupta and Miescke \(1996\)](#), referring to it as the knowledge gradient (KG) to emphasize the way in which it compares the difference in expected value between two different quantities of information or knowledge, and also to emphasize its membership in a family of one-step lookahead expected value of information rules. [Frazier and Powell \(2008\)](#) introduce a new stopping rule that, when used with the KG allocation rule, improves its selection efficiency in several numerical tests, as compared with the best combinations of allocation rules and stopping rules in [Branke et al. \(2007\)](#).

Both [Branke, Chick, and Schmidt \(2007\)](#) and [Frazier and Powell \(2008\)](#) observe that procedures tend to perform best when the allocation rule and the stopping rule are derived with a similar criterion, such as an expected opportunity cost (EOC) criterion or a probability of good selection criterion. The stopping rules from those papers compare the expected reward from a single stage of additional sampling with either statistical thresholds or with sampling cost information.

This paper seeks to further improve the performance of such procedures. We develop new stopping rules that can be used to compare the value of the option to continue simulating (i.e., the expected reward from not just one stage of simulating, but

from the option to simulate, observe the data, and still have the option to continue simulating) with the costs of simulating. The stopping rules below therefore do a better job, in theory, of balancing the cost of simulating with the value of the option to continue simulating before deciding. This balance is sought with the goal of better aligning these stopping rules with the EOC-based allocations that will be used.

Chick and Gans (2009) also assess the value of the option to continue sampling before selecting. That paper examines the special case of one system that can be simulated, then either accepted or rejected as being better or worse than a system with a known expected reward. It does so by solving for the optimal stopping time for a particular diffusion process. That paper extends the idea to multiple simulated systems by deriving upper and lower bounds on the expected value of the option to simulate. Those bounds are shown to be close to each other, and so a diffusion approximation is fortunately not needed when there are multiple systems. Those bounds result in stopping rules that effectively balance the expected value of information from the option to further simulate (not just the information from one stage), and the costs of simulation. The costs of simulation in Chick and Gans (2009) include a cost per sample $c \geq 0$ and a positive discount rate $\delta > 0$, both of which incur costs when more samples are taken. That paper does not show, however, the solution when there are positive sampling costs ($c > 0$) but no discounting ($\delta = 0$).

The main contribution of this paper is to extend the expected value of information approach by focusing on the expected value of the option to continue sampling, not on just the expected value of a single stage of sampling. We handle the case of positive sampling costs and no discounting costs. We use dynamic programming ideas and treat simulation selection as an option to gain information. We also develop easily-implemented stopping rules for determining when to select a system rather than continuing to simulate. Those rules are based on the costs of simulation runs, rather than statistical thresholds.

2 PROBLEM STATEMENT

An analyst seeks to implement one of k systems, labeled $i = 1, 2, \dots, k$, whose expected reward is not known with certainty, or to implement an alternative course of action, labeled $i = 0$, whose expected reward is the known value m . For example, if the analyst has the option to implement one of k simulated systems or to reject them all and ‘do nothing’, then $m = 0$. Prior to selecting a system ($i = 0, 1, 2, \dots, k$), the analyst can choose to sequentially simulate one or more of the k systems until a system is chosen for implementation. The analyst would like to maximize the expected reward from this process, including the cost of simulation replications and the expected reward from implementing a system.

Section 2.1 formalizes the general problem, then Section 2.2 specializes the analysis to the case of normally distributed simulation output with unknown means (but known sample variances).

2.1 Assumptions and Notation

Let X_i be the random variable representing the reward of system i , where $X_0 \equiv m$. If the analysis is risk neutral and the distributions of all X_i ’s are known to him, then he will select the system with the largest expected reward, $i^* = \arg \max_i \{E[X_i]\}$.

Here, we assume that the distributions of the X_i ’s are not known with certainty by the analyst, for $i = 1, 2, \dots, k$. Rather, he believes that a given X_i comes from one of a family of probability distributions, $P_{X_i|\theta_i}$, indexed by a parameter θ_i from a parameter space Ω_{θ_i} . We model his belief with a probability distribution on θ_i , which we call P_{θ_i} . The expected reward of system $i > 0$ is then $E[X_i] = \iint X_i dP_{X_i|\theta_i} dP_{\theta_i}$. We assume integrability of X_i under $dP_{X_i|\theta_i} dP_{\theta_i}$.

When P_{θ_i} can be uniquely represented by a finite set of sufficient statistics Θ_i , such as for distributions that are in a regular exponential family of distributions, then P_{θ_i} can be referred to by Θ_i without ambiguity. For example, the analyst may believe that X_i is normally distributed with a known variance, σ_i^2 , but unknown mean. Then P_{θ_i} is a probability distribution for the unknown mean. If P_{θ_i} itself is a normal distribution with mean μ_{0i} and variance σ_{0i}^2 , then P_{θ_i} can be identified with $\Theta_i = (\mu_{0i}, \sigma_{0i}^2)$. When sufficient statistics are available for each system, we set $\vec{\Theta} = (\Theta_1, \Theta_2, \dots, \Theta_k)$.

If the distributions of the X_i ’s are not known, then the analyst has the option to use simulation as a tool to reduce uncertainty about the distribution before selecting which system to develop outside the simulator. He may decide to simulate the outcome of system i a number of times, and he views the result of each replication as a sample X_i from $P_{X_i|\theta_i}$, where θ_i is the (unknown) true parameter value for system i . We also assume that simulation results, and hence the evolution of the analyst’s beliefs, are independent from one system to the next.

We model simulations as occurring at a sequence of discrete stages $t = 0, 1, 2, \dots$. Let \mathbf{X}_t be the set of all outputs seen through stage t . We represent Bayesian updating of prior beliefs and sample outcomes through time, $\{(\vec{\Theta}_t, \mathbf{X}_t) | t = 0, 1, \dots\}$, as follows. If system $i > 0$ is simulated at stage t with sample outcome $x_{i,t}$, then $X_{i,t} = x_{i,t}$ and Bayes’ rule determines the posterior distribution $P_{\theta_i,t+1}$. The density of that posterior distribution is described as a function $dP_{\theta_i|\Theta_{i,t+1}}$ of the parameter

$\theta_i \in \Omega_{\Theta_i}$:

$$dP_{\theta_i|\Theta_{i,t+1}} = \frac{dP_{X_i|\theta_i} dP_{\theta_i|\Theta_{i,t}}}{\int_{\Theta_i} dP_{X_i|\theta_i} dP_{\theta_i|\Theta_{i,t}}} \quad \forall \theta_i \in \Omega_{\Theta_i},$$

while $\Theta_{j,t+1} = \Theta_{j,t}$, and $X_{j,t}$ need not be defined for all $j \neq i$. Thus, the evolution of the analyst's beliefs regarding the distribution of outcomes of each system, $\Theta_{i,t}$, is Markovian.

We assume that the incremental cost of each run of system i is c_i . Unlike [Chick and Gans \(2009\)](#), we do not consider discounting costs associated with delays from sampling. In the absence of sampling costs or a discount rate, one could imagine pathological scenarios such as simulating forever to resolve all uncertainty about the mean performance of each system before selecting one. Sampling costs avoid such pathological scenarios.

The analyst must choose a sequence of simulation runs and ultimately select a system, so that the stream of costs and terminal expected reward, together maximize the expected reward. We define indices to track the analyst's choices. Let $T \in \{t = 0, 1, \dots\}$ be the stage at which the analyst selects a system to implement. For $t < T$, let $i(t) \in \{1, 2, \dots, k\}$ be the index of the system simulated at time t . Let $I(T) \in \{0, 1, \dots, k\}$ be the index of the system that is ultimately chosen.

Then a *selection policy* $\pi = (i(\cdot), T, I(T))$ is the choice of a sequence of simulation runs, a stopping time, and a final system. We define Π to be the set of all *non-anticipating* selection policies. For such policies, the function $i(t)$, the indicator function of the event $\{T = t\}$ and the function $I(T)$ (given the event $\{T = t\}$) each depend only on the history up to t : $\{\vec{\Theta}_0, \mathbf{X}_0, \vec{\Theta}_1, \mathbf{X}_1, \dots, \vec{\Theta}_{t-1}, \mathbf{X}_{t-1}, \vec{\Theta}_t\}$, for $t = 0, 1, \dots$.

Given prior distributions $\vec{\Theta} = (\Theta_1, \Theta_2, \dots, \Theta_k)$ and a policy $\pi \in \Pi$, the expected value of the future stream of rewards is

$$V^\pi(\vec{\Theta}) = E_\pi \left[\sum_{t=0}^{T-1} -c_{i(t)} + X_{I(T),T} \mid \vec{\Theta}_0 = \vec{\Theta} \right]. \quad (1)$$

Formally, we define the analyst's (undiscounted) *simulation selection problem* to be the choice of a selection policy $\pi^* \in \Pi$ that maximizes this (undiscounted) expected reward: $V^{\pi^*}(\vec{\Theta}) = \sup_{\pi \in \Pi} V^\pi(\vec{\Theta})$, when the supremum is attained. Note that, conditional on stopping to select, the system $I(T)$ that is selected gives an expected reward

$$\max\{m, E[X_{1,T} \mid \vec{\Theta}_T], E[X_{2,T} \mid \vec{\Theta}_T], \dots, E[X_{k,T} \mid \vec{\Theta}_T]\}.$$

2.2 Specialization To Normal Random Variables With Conjugate Prior Distributions

This section specializes (1) to the case of normally distributed realizations that have unknown means and known variances. Let X_{ij} be a random variable whose realization x_{ij} is the output of the $(j+1)$ th sample from system i , for $i = 1, \dots, k$ and $j = 0, 1, 2, \dots$. Let U_i be the unknown mean and let σ_i^2 be the known variance of system i . Given a realization of the unknown means, we denote their order statistics by $u_{[1]} \leq u_{[2]} \leq \dots \leq u_{[k]}$, with ties broken randomly. We assume that, conditioned on the unknown means U_i ,

$$\{X_{ij} : j = 0, 1, 2, \dots\} \mid U_i \stackrel{iid}{\sim} \text{Normal}(u_i, \sigma_i^2) \text{ for } i = 1, 2, \dots, k.$$

Let $\mathbf{U} = (U_1, U_2, \dots, U_k)$, $\mathbf{u} = (u_1, u_2, \dots, u_k)$ and $\boldsymbol{\sigma}^2 = (\sigma_1^2, \sigma_2^2, \dots, \sigma_k^2)$.

To describe uncertainty about the mean expected rewards, we assume the following conjugate prior distributions,

$$U_i \stackrel{iid}{\sim} \text{Normal}(\mu_{0i}, \sigma_i^2/t_{0i}) \text{ for } i = 1, 2, \dots, k.$$

The value t_{0i} can be interpreted as the effective number of replications of data that is embodied by the prior distribution for the unknown mean of system i .

When specializing (1) to the case of normally distributed output with an unknown mean whose distribution is normally distributed itself, P_{Θ_i} is the Normal $(\mu_{0i}, \sigma_i^2/t_{0i})$ distribution, Θ_i can be represented by the vector $(\mu_{0i}, \sigma_i^2/t_{0i})$ or an equivalent set of information, such as (μ_{0i}, t_{0i}) when σ_i^2 is known, and the realization of the parameter θ_i is $U_i = u_i$.

3 SPECIAL CASE OF ONE SYSTEM

This section examines the special case of $k = 1$ simulated system and deciding whether to implement it or not after simulations have been run. We drop $i(t)$ and $I(T)$ in subscripts for notational convenience in this section since there is only one simulated system. Then the optimization of (1) becomes the following optimal stopping problem,

$$V^{\pi^*}(\vec{\Theta}) = \sup_{\pi} E_{\pi} \left[-cT + \max\{m, E[X_T | \vec{\Theta}_T]\} \mid \vec{\Theta}_0 = \vec{\Theta} \right]. \quad (2)$$

We describe a diffusion approximation to estimate $V^{\pi^*}(\vec{\Theta})$ and numerically solve it to find the optimal stopping boundary.

3.1 Diffusion Approximation for Simulation Selection when $k=1$

We approach the optimal stopping problem in (2) by using a diffusion approximation (analogous to innovations by Chernoff 1961 and used later by Chick and Gans 2009). We note that the process $\{(\mu_0 t_0 + \sum_{i=t}^n X_i, t_0 + n) : n = 0, 1, 2, \dots\}$ is a random walk with Gaussian and independent increments. This discrete time process has the same distribution as the continuous time process $\{(Y_t, t) : t \geq t_0\}$ at points t for which $t - t_0$ is an integer, where, conditioned on W , $\{Y_t : t \geq t_0\}$ is a Brownian motion with $Y_{t_0} = \mu_0 t_0$, drift W , and volatility σ^2 . From this construction, we have

$$Y_t \mid U \sim \text{Normal}(\mu_0 t_0 + U(t - t_0), \sigma^2(t - t_0)) \text{ and } U \sim \text{Normal}(\mu_0, \sigma^2/t_0).$$

Define $t = t_0 + n$ and $y_{t_0} = \mu_0 t_0$, so that (y_{t_0}, t_0) is sufficient to describe the prior distribution for U . If we set $y_t = y_{t_0} + \sum_{i=1}^n x_i$ (the continuous time and the discrete time processes have the same statistics, so we adopt this abuse of notation here), then

$$U \mid (Y_s, s) \text{ for } s \in [t_0, t] \sim \text{Normal}(y_t/t, \sigma^2/t).$$

That is, the posterior distribution of the unknown expected reward from implementing a system, U , conditional on x_1, x_2, \dots, x_n , is easily described by t and y_t . We can therefore refer to the posterior distribution $\vec{\Theta}_t$ by (y_t, t) and we will do so below.

We now convert the above diffusion (Y_t, t) to a standardized stopping problem for a diffusion that can be used to approximate the optimal stopping boundary and the expected reward for the simulation selection problem with $k = 1$. We do this by adapting techniques in Chick and Gans (2009) to the context here.

This is done in two steps. First, we rescale the variables in the problem with the goal of being able to solve for any combination of m , c and σ^2 with the solution to a single optimal stopping problem. We introduce constants α and γ , which are arbitrary for now but will be defined later in terms of c and σ^2 . Let $r = \gamma t$ be a scaled time and $Z_r = \alpha Y_t$ be a scaled effective sum of observations. Set $\beta = \alpha/\gamma$ so that $Z_r/r = \beta Y_t/t$.

Let the continuous time stopping time for the diffusion, \tilde{T} , take the role of the discrete time stopping time, $t_0 + T$, shifted by the effective number of observations in the prior distribution. Then the continuous time analog of (2) is

$$\begin{aligned} V^{\pi^*}(y_{t_0}, t_0) &= \sup_{\pi} E_{\pi} \left[-cT + \max\{m, E[X_T | \vec{\Theta}_T]\} \mid \vec{\Theta}_0 = \vec{\Theta} \right] \\ &\approx \sup_{\tilde{T} \geq t_0} E_{\pi} \left[-c(\tilde{T} - t_0) + \max\{m, Y_{\tilde{T}}/\tilde{T}\} \mid (y_{t_0}, t_0) \right] \text{ (diffusion approximation)}. \end{aligned} \quad (3)$$

Note that the diffusion approximation in (3) can be replaced with a \leq because the diffusion approximation allows for all real-valued stopping times including integers whereas the original problem only allows for stopping at integer-valued times.

Second, we change to a reverse-time coordinate system so that the statistics of the Gaussian process are simplified. To this end, let $s = 1/r = 1/(\gamma t)$, $W_s = \alpha Y_t/(\gamma t) = \beta Y_t/t$ and $\tilde{m} = \beta m$. Set $s_0 = 1/(\gamma t_0)$ and $w_{s_0} = \alpha y_{t_0}/(\gamma t_0)$. If we set $\alpha^2 \sigma^2 = \gamma$ (to control the variance of W_s), then (W_s, s) is a Brownian motion without drift in the $-s$ scale from reverse time s_0 back to 0 (Chernoff 1961). Then $S = 1/(\gamma \tilde{T})$ is a stopping time in the $-s$ scale.

With a bit of algebra that is not shown, one can seek to standardize the optimization problem in (3) as follows.

$$B(m, y_{t_0}, t_0) \triangleq \sup_{\tilde{T} \geq t_0} E_{\pi} \left[-c(\tilde{T} - t_0) + \max\{m, Y_{\tilde{T}}/\tilde{T}\} \mid Y_{t_0} = y_{t_0} \right] \tag{4}$$

$$= m + \sup_{s \in [0, s_0]} \beta^{-1} E \left[-\frac{c\alpha}{\gamma^2} \left(\frac{1}{S} - \frac{1}{s_0} \right) + \max\{0, W_S\} \mid W_{s_0} = w_{s_0} - \tilde{m} \right]. \tag{5}$$

If, in addition to the constraints $\beta = \alpha/\gamma$ and $\alpha^2\sigma^2 = \gamma$ that result in W_s being a Brownian motion in the $-s$ scale, we add the constraint $c\alpha = \gamma^2$, then (5) simplifies even further. In particular, these three constraints imply that

$$\alpha = c^{1/3}\sigma^{-4/3}, \beta = c^{-1/3}\sigma^{-2/3} \text{ and } \gamma = c^{2/3}\sigma^{-2/3}, \tag{6}$$

and with these parameter values, the supremum of the expectation in (5) is

$$\tilde{B}(w_{s_0} - \tilde{m}, s_0) \triangleq \sup_{s \in [0, s_0]} E \left[-\left(\frac{1}{S} - \frac{1}{s_0} \right) + \max\{0, W_S\} \mid W_{s_0} = w_{s_0} - \tilde{m} \right]. \tag{7}$$

This development leads to the following proposition. It implies that only the standardized problem ($c = 1, \sigma = 1, m = 0$) in (7) need be solved in order to get the diffusion approximation $V^{\pi^*}(y_{t_0}, t_0) \approx B(m, y_{t_0}, t_0)$ for any $m, c > 0$ and $\sigma > 0$.

Proposition 1 *If B is as in (4), \tilde{B} is as in (7) and $\alpha, \beta,$ and γ are as in (6), then $B(m, y_{t_0}, t_0) = m + \beta^{-1}\tilde{B}(\beta(y_{t_0}/t_0 - m), t_0)$.*

The key to the optimal solution when $k = 1$, then, is to solve for the optimal stopping problem in (7). Standard results (Chernoff 1961, Bather 1970) can be used to show that the function \tilde{B} in (7) satisfies a free-boundary problem for a heat equation. Such a problem is a partial differential equation whose boundary is not fixed. The boundary is determined by a condition that expresses an indifference between the options of stopping and of continuing to sample. The solution to such a problem determines a boundary that defines a continuation region, \mathcal{C} , within which it is optimal to continue sampling. Outside of \mathcal{C} , it is optimal to stop sampling and to choose the better of taking a known m or implementing the simulated system. See Chick and Frazier (2009) for details on how to derive and numerically solve that free boundary problem.

3.2 Empirical Results for Diffusion Approximation when $k=1$

We implemented a finite differences (FD) scheme with Matlab in order to solve the free boundary problem that corresponds to (7). This section describes the results from that FD scheme and applies it to some numerical examples.

Figure 1 displays the optimal stopping boundary for the standardized problem with hatched lines for two ranges of s . Chick and Frazier (2009) gives a ‘Quick Approximation’ for the optimal stopping boundary, $\tilde{b}(s)$, for the range $s \in (0, 6300)$. Such an approximation obviates the need for other users to implement and solve the PDE for cases with $k = 1$. The figure also displays the contours of the expected benefit for continuing to sample when it is optimal to do so, $\tilde{B}(w, s) - \max(0, w)$.

The continuation set in Figure 1 has both an upper and a lower stopping boundary. This differs from the solution to the optimal continuation set for the selection problem when the discount rate is positive. If the discount rate is positive but the cost of sampling is 0, then there is only an upper stopping boundary and one would only simulate for a finite time if the default reward m exceeds the NPV of simulating forever (Chick and Gans 2009). Here, the penalty for simulating forever is infinite (unlike the case of a positive discount rate) and the positive sampling cost implies that if a system is sufficiently worse than m , one would prefer to stop sampling altogether and accept a known expected reward of m .

For small values of s (left panel of Figure 1) the upper boundary of the stopping region appears to be convex in s and is narrow. Small values of s indicates that a large number of samples have been observed, so the value of the mean is likely to be close to the scaled posterior mean, w . Only a narrow range of possible values of w merit additional sampling.

For larger values of s (right panel of Figure 1) the upper boundary of the stopping region appears to be concave in s and includes a wider range of values of w in the continuation region. This is because there is still a great deal of uncertainty about w when s is large, and sampling has more value.

The functional form for the upper and lower boundaries in $(y_t/t, t)$ coordinates is $m \pm \beta^{-1}b(1/(\gamma t))$. That is, one continues sampling as long as the posterior mean, y_t/t , is in the interval $[m - c^{1/3}\sigma^{2/3}b(\sigma^{2/3}/(c^{2/3}t)), m + c^{1/3}\sigma^{2/3}b(\sigma^{2/3}/(c^{2/3}t))]$.

Figure 2 shows contours that demonstrate the expected benefit of continuing to sample, given that it is optimal to do so, relative to stopping prematurely. It does so in coordinates that are more natural for a simulation study (y_t, t) rather than

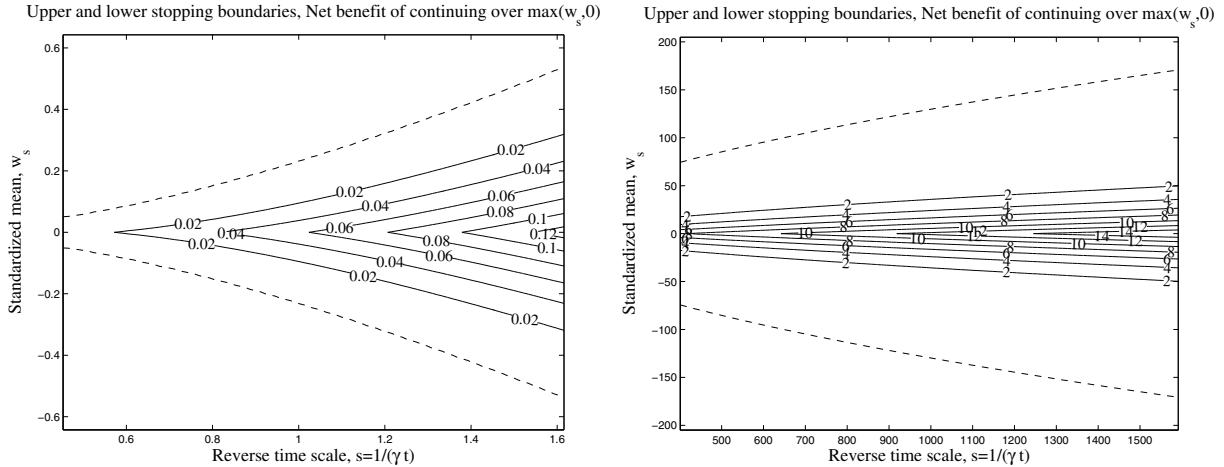


Figure 1: The optimal continuation set of the standardized free boundary problem that corresponds to (7) is between the dashed lines. The solid lines give level sets of the expected reward of continuing when it is optimal to do so, $\tilde{B}(w, s) - \max(0, w)$.

w_s, s). Mathematically, that benefit is $B(m, y_t, t) - \max\{m, y_t/t\}$. Above the upper dashed line and below the lower dashed line it is optimal to stop immediately. In those stopping regions, the expected benefit of following the optimal policy is 0.

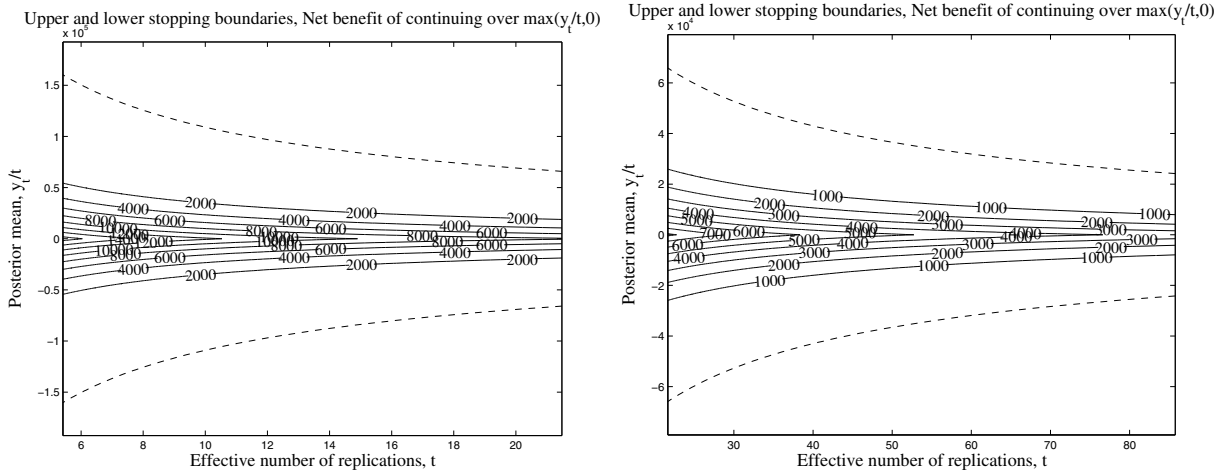


Figure 2: Expected benefit of continuing when optimal to do so, $B(m, y_t, t) - \max\{m, y_t/t\}$, with $c = 1$, $\sigma = 10^5$, $m = 0$.

The optimal stopping boundary in this section is useful for two reasons. First, the PDE need not be implemented to re-solve the problem optimally when $k = 1$. Instead, the optimal stopping boundary for the standard problem is readily approximated with an easy to compute function (Chick and Frazier 2009), and this standard stopping boundary is easily transformed to the stopping boundary appropriate to the problem that one wishes to solve. Second, it allows assessing the effectiveness and calculating the suboptimality of the stopping rules that are developed below in Section 4.2. Section 5 will pursue this with numerical experiments.

4 MULTIPLE SYSTEMS: STOPPING RULES AND ALLOCATIONS

The approach in Section 3 for $k = 1$ simulated system, that of deriving a diffusion model and numerically solving a corresponding free boundary problem to approximate the optimal solution, would be unwieldy for a large number of systems. We therefore seek an alternative approach to developing a sequential selection procedure for $k > 1$. The key elements of a sequential procedure are (i) a stopping rule that determines when to stop sampling, (ii) an allocation rule that identifies

which system to sample when continuing, and (iii) a selection rule that chooses a system to implement when sampling has stopped.

This section presents background useful for deriving several different stopping and allocation rules. It then presents the rules themselves. Although the question of which stopping and allocation rules are best is open, the optimal selection rule is well understood and easy to compute. It is known that, conditional on being forced to select, the selection rule that minimizes the expected regret of a potentially incorrect selection is the decision that picks the system with the largest posterior mean reward (Chick and Inoue 2001). In the context of the present work, this means it is optimal to pick the system with the maximal expected reward when stopping. We use this selection rule uniformly.

4.1 Background Results for Stopping Rules and Allocations

The optimal expected reward (OER) $V^{\pi^*}(\vec{\Theta})$ of the simulation selection problem is at least as large as the expected reward of any given policy. That includes so-called *one-stage allocation* policies. We use the idea of one-stage allocation policies to develop bounds for the OER and stopping rules when $k > 1$.

Formally, a one-stage allocation policy maps each sampling budget $\beta \geq 1$ to an allocation of those samples across the k systems, with a total of $\tau_i = \tau_i(\beta) \geq 0$ replications to be run for system i , so that $\sum_{i=1}^k c_i \tau_i = \beta$. We set $\tau = (\tau_1, \tau_2, \dots, \tau_k)$ and allow β and the τ_i to be real-valued for mathematical convenience. In implementations, the values of τ_i will be rounded to integer values. We require a one-stage allocation to have the property that all of the τ_i are all nondecreasing as β increases. For example, the equal allocation sets $\tau_i = \beta/k$ (rounding to retain the integer budget constraint if needed, while retaining the monotonicity of the τ_i in β). After observing those samples, a one-stage allocation policy selects the system with the largest posterior expected reward, if that reward exceeds

$$\mu_{00} \triangleq E[X_0] = m \tag{8}$$

and otherwise selects the default system, which has reward m .

Suppose further that samples from system i are normally distributed with known variance σ_i^2 , but unknown mean. The prior distribution of this unknown mean is Normal $(\mu_{0i}, \sigma_i^2/t_{0i})$, with $\mu_{0i} = y_{0i}/t_{0i}$. Then the posterior mean of our belief about this sampling mean that will result after the future sampling is performed is the random variable (de Groot 1970)

$$Z_i \sim \text{Normal} \left(\mu_{0i}, \frac{\sigma_i^2 \tau_i}{t_{0i}(t_{0i} + \tau_i)} \right). \tag{9}$$

If we consider the allocation to be a function of β and vary β over all feasible sampling budgets, we obtain the following lower bound. The bound refers to order statistics (i) for $i = 0, 1, \dots, k$, which are defined so that $\mu_{0(0)} \leq \mu_{0(1)} \leq \dots \leq \mu_{0(k-1)} \leq \mu_{0(k)}$. These statistics are updated as samples are observed.

Proposition 2 *Let $V^{\pi^*}(\vec{\Theta})$ maximize (1) and $\tau = \tau(\beta)$ be a one-stage allocation. Let Z_i be the (random) posterior mean given that τ_i replications for system i will be run. Then*

$$V^{\pi^*}(\vec{\Theta}) \geq \underline{\text{OER}}(\vec{\Theta}) \triangleq \sup_{\beta \geq 1} E[\max\{\mu_{00}, Z_1, Z_2, \dots, Z_k\}] - \beta. \tag{10}$$

Complementing $\underline{\text{OER}}(\vec{\Theta})$, we also define $\overline{\text{OER}}(\vec{\Theta}) \triangleq E[\max\{\mu_{00}, U_1, U_2, \dots, U_k\}]$ as the the expected reward of having perfect information about each system at no cost and selecting the best system.

The expectation on the right hand side of (10) has some easy-to-compute bounds.

Proposition 3 *Let τ be a one-stage allocation, assume that output is jointly independent and normally distributed with a known variance, and normal prior distribution so that (9) is valid for each i , let $\Psi[s] = \int_s^\infty (\xi - s)\phi(\xi)d\xi = \phi(s) - s(1 - \Phi(s))$ be the standard normal linear loss function, let $\Delta_{0i} = |\mu_{0i} - \max_{j \neq i} \mu_{0j}|$ be the difference in expected value between system i and the best of the other systems, let $\sigma_{Z,0,\tau}^2 = 0$, $\sigma_{Z,i,\tau}^2 = \frac{\sigma_i^2 \tau_i}{t_{0i}(t_{0i} + \tau_i)}$, and $\sigma_{Z,i,(k),\tau}^2 \sigma_{Z,i,\tau}^2 + \sigma_{Z,(k),\tau}^2$. Then*

$$\mu_{0(k)} + \max_{i:i \neq (k)} \sigma_{Z,i,(k),\tau} \Psi \left[\frac{\Delta_{0i}}{\sigma_{Z,i,(k),\tau}} \right] \leq E[\max\{\mu_{00}, Z_1, Z_2, \dots, Z_k\}] \leq \mu_{0(k)} + \sum_{i:i \neq (k)} \sigma_{Z,i,(k),\tau} \Psi \left[\frac{\Delta_{0i}}{\sigma_{Z,i,(k),\tau}} \right]. \tag{11}$$

Proposition 4 notes that the expectation on the right hand side of (10) is available analytically when $k = 1$ or when all samples are allocated to a single system. The lemma and the development below refer to the notation in Proposition 3.

Proposition 4 *If the allocation τ allocates its entire budget to system i , that is, if $\tau_i = \beta/c_i$ and $\tau_j = 0$ for all $j \neq i$, then*

$$E[\max\{\mu_{00}, Z_1, Z_2, \dots, Z_k\}] = \mu_{0(k)} + \sigma_{Z,i,\tau} \Psi \left[\frac{\Delta_{0i}}{\sigma_{Z,i,\tau}} \right].$$

4.2 Stopping Rules for Multiple Systems

When $k = 1$, the optimal stopping rule is known. When $k \geq 1$, the bounds and approximations in Section 4.1 provide additional mechanisms to assess whether the expected reward of continuing to sample exceeds the expected cost of continuing to sample. The following stopping rules codify some of these tradeoffs.

KG $_{\beta}$ stopping rule. Continue to sample if and only if there is a system i such that using the β exclusively for that system produces a net expected benefit, as computed by Proposition 4. Thus, the KG $_{\beta}$ stopping rule calls for sampling to continue if only if $\max_{i \in \{0,1,\dots,k\}} \sigma_{Z,i,\tau} \Psi[\Delta_{0i}/\sigma_{Z,i,\tau}] > \beta$, where the maximand checks, for system i , the one-stage allocation with $\tau_i = \beta/c_i$. Equivalently, sampling continues if and only if $\max_i v_i^{KG,\beta} > 1$, where $v_i^{KG,\beta}$ may be understood as an average benefit per unit cost of measurement and is defined by

$$v_i^{KG,\beta} = \frac{1}{\beta} \sigma_{Z,i,\tau} \Psi \left[\frac{\Delta_{0i}}{\sigma_{Z,i,\tau}} \right].$$

This extends an earlier KG stopping rule that required $\beta = 1$ and $c_i = 1$ (Frazier and Powell 2008).

Instead of considering a fixed β , we can consider the expected benefit obtained and cost incurred from allocating any budget β entirely to a single system. This idea gives rise to the following stopping rule.

KG $_{*}$ stopping rule. Continue to sample if and only if there exists a budget $\beta \geq 1$ such that the KG $_{\beta}$ stopping rule calls for continuing. That happens when $v_i^{KG,*} > 1$ for some i , where

$$v_i^{KG,*} = \sup_{\beta \geq 1} \frac{1}{\beta} \sigma_{Z,i,\tau} \Psi \left[\frac{\Delta_{0i}}{\sigma_{Z,i,\tau}} \right].$$

The KG $_{\beta}$ and KG $_{*}$ stopping rules test if allocating samples to a single system can justify continuing. The next stopping rule allows for a specified one-stage allocation to sample from multiple systems when checking whether to continue sampling or not. Further, it further increases sampling by using an upper bound on the expected reward of sampling.

EOC $_{c,k}$ stopping rule. Recall (11) and (10). Continue sampling if and only if there is a budget $\beta \geq 1$ such that

$$\mu_{0(k)} + \sum_{i \in \{0,1,\dots,k\} \setminus \{(k)\}} \sigma_{Z,i,(k),\tau} \Psi \left[\frac{\Delta_{0i}}{\sigma_{Z,i,(k),\tau}} \right] - \beta > \mu_{0(k)}. \tag{12}$$

4.3 Allocations for Multiple Systems

The stopping rule indicates whether or not to continue sampling. If sampling is to continue, we must also decide how to allocate the next sample to one of the k systems. Previous work and the stopping rules in Section 4.2 suggest several allocations.

KG $_1$ /LL $_1$ Allocation. The KG $_1$ /LL $_1$ allocation greedily allocates one sample at a time (e.g., if $\beta = c_i$ for all i) to a single system in order to maximize the expected value of information (EVI), were the procedure to collect only one more sample. That EVI, when the sample is allocated to system i , is $EVI_i = \sigma_{Z,i,\tau} \Psi[\Delta_{0i}/\sigma_{Z,i,\tau}]$. The KG policy of Frazier et al. (2008) for known variances and the LL $_1$ allocation of Chick et al. (2009) for unknown variances work on the same principle. This allocation is also the (R1,R1,...,R1) rule of Gupta and Miescke (1996).

KG $_{\beta}$ Allocation. One sample is allocated to the system that maximizes $v_i^{KG,\beta}$. Although a measurement is valued according to the average value obtained over a budget β worth of samples, only one sample is allocated at a time.

KG* Allocation. While the KG_{β} allocation values a measurement according to the average benefit obtained by measuring it β times, KG_* values it according to the best average benefit obtainable, which is $v_i^{KG,*}$. The KG_* allocation allocates one sample to the system that maximizes $v_i^{KG,*}$.

The one-stage LL allocation asymptotically maximizes the right hand side of (11), subject to the constraint that the total number of samples is β , for large values of β (see Chick and Gans 2009 for the variation with known sample variances). The sequential LL allocation takes advantage of this and is therefore aligned with the $EOC_{c,k}$ stopping rule. It works as follows.

Sequential LL Allocation. In (12), there is an optimal budget β^* . With that budget, the one-stage LL allocation allocates $\tau_1(\beta^*), \tau_2(\beta^*), \dots, \tau_k(\beta^*)$ samples to the k systems. Sequential LL uses this batch allocation to choose a single sample to allocate next, allocating one sample to system $\arg \max \{c_i \tau_i(\beta^*)\}$.

Equal Allocation. The equal allocation allocates one replication at a time in round robin fashion to the k systems, so that each system has roughly the same number of samples at any time until sampling stops.

4.4 Implementation

The above stopping and allocation rules allow for some flexibility in their specification and some approximations or tricks can speed up the computation of the quantities they use. There are also numerical stability issues to resolve when implementing the conceptual models above. For example, the optimal budget for the KG_* stopping rule can be solved approximately with a readily-computed $\tilde{\beta}^*$ rather than by a brute force search for the optimal β^* . Chick and Frazier (2009) provide more details.

5 SIMULATION RESULTS

Simulation results are presented for the case of $k = 1$ and then for the case $k > 1$. With $k = 1$ we can compare the known optimal result for the diffusion with the approximations in the above stopping rules, in order to assess their degree of suboptimality. With $k > 1$, an optimal diffusion result is not as tractable, and so we assess the performances of the various procedures by comparing them with bounds on their performance.

In each Monte Carlo (MC) experiment, 10^6 random problem instances were generated from the prior distribution $\vec{\Theta}_0$ for the unknown means of each system, and the average performance was assessed by applying each simulation selection procedure to each problem instance.

5.1 Numerical Results: One Simulated System

Figure 3 shows the stopping boundaries that delimit the continuation sets for several stopping rules. The figure presumes that $c = 1$, $\mu_0 = 0$, $\sigma^2 = 10^5$, $t_0 = 1$ and $m = 0$. The optimal stopping boundary somewhat exceeds the KG_* stopping boundary and the relative error percentage increases somewhat as t increases. The KG_* continuation set coincides, theoretically, with the $EOC_{c,k}$ continuation set for the special case of $k = 1$. The KG_* continuation sets in turn contain the continuation sets of the KG_{β} stopping rules. Those continuation sets are not nested for $\beta = 1, 10$ and 100 . The figure also shows that the KG_* boundary, as computed with a brute force search for the optimal β^* , is closely approximated by stopping boundary when β^* is approximated by $\tilde{\beta}^*$.

Table 1 shows MC simulation results for these stopping rules when $c = 0, \mu_0 = 0, \sigma = 10^5$ and $m = 0$ for both $t_0 = 1$ and $t_0 = 100$. The table reports estimates (\pm one standard error) of the expected number of samples, $E[T]$, the expected opportunity cost, $E[U_{[k]} - U_{(k)} | \Theta_T]$, and the expected reward, $E[\max\{m, E[X_T | \Theta_T]\} - cT | \Theta_0]$, where T depends on the specified stopping rule. The PDE solution, as calculated via a first-order bias corrected FD scheme, is slightly higher than the MC estimates based on discrete time sampling and the estimated optimal stopping boundary, as expected. The relative level of sub-optimality for each stopping rule is calculated in a comparison with the MC value of the reward with the estimated optimal stopping boundary.

For both $t_0 = 1$ and $t_0 = 100$ we observe that there is an ordering of the stopping rules, which is decreasing in the expected number of samples, increasing in the expected opportunity cost, and decreasing in the expected reward. The optimal stopping rule (as estimated with the FD approximation) performs best, with $EOC_{c,k}$ second best, followed by KG_* and KG in this order. This ordering in reward, opportunity cost, and number of samples is explained by the nesting of the continuation regions, with the optimal continuation region containing the $EOC_{c,k}$ continuation region, which contains the KG_* continuation region, which contains the KG_1 continuation region.

The poor performance of KG_1 when $t_0 = 100$ can be explained by recalling Figure 3. When t is in the range from 80 to 3000, the KG_1 stopping boundary is quite a bit lower than the optimal stopping boundary. But that is the range of t that is most likely going to lead to stopping when $t_0 = 100$, given the value of $E[T]$. When $t < 80$ or $t > 3000$, then the KG_1

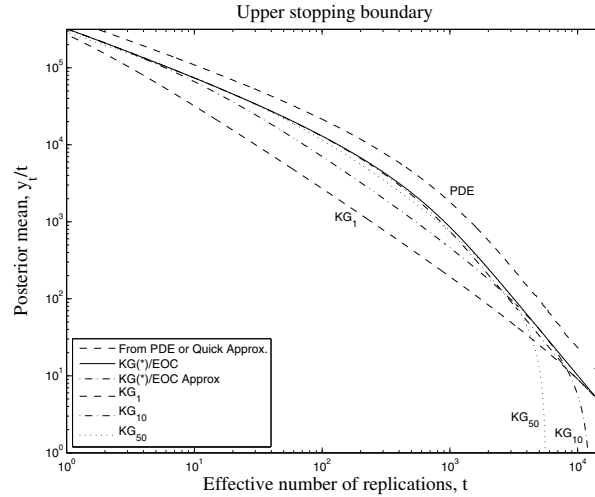


Figure 3: Stopping boundaries with $k = 1$ system and parameter values $c = 1$, $\sigma = 10^5$, $m = 0$. The optimal continuation set (as approximated by the finite difference solution) contains the KG_* continuation set (which is the same as the $EOC_{c,k}$ continuation set when $k = 1$), which in turn contains the KG_{β} continuation sets.

stopping boundary is a relatively better estimate of the optimal stopping boundary, and so the KG_1 stopping rule performs relatively better, as compared to the optimal, when $t_0 = 1$. A similar argument explains why the KG_* and $EOC_{c,k}$ stopping rules perform less well when $t_0 = 100$ as compared with when $t_0 = 1$ (although these rules are both better than KG_1).

While the $EOC_{c,k}$ and KG_* stopping rules are equivalent, if implemented without any approximations, in our implementation the KG_* stopping rule uses an approximation $\tilde{\beta}^*$ to the optimal budget for KG , as noted in Section 4.4. The implementation of $EOC_{c,k}$ checks the budget $\tilde{\beta}^*$ to see if it merits continuing as well as other values of the budget. Thus its implemented continuation region is slightly larger.

5.2 Numerical Results: Multiple Systems

Table 2 summarizes the performance of various allocation and stopping rules for $k > 1$. As we observed in the $k = 1$ case, the $EOC_{c,k}$ stopping rule performed the best, with other stopping rules stopping earlier, incurring larger opportunity costs, and lower rewards. KG_* performed second best, taking more samples than KG_1 and similarly achieving lower opportunity costs and higher rewards.

The KG_1 allocation with the KG_1 stopping rule performs the poorest of these allocations. It is even worse than the optimal one-stage allocation (whose relevant values are in the rows labeled OER) for the expected total reward and expected opportunity cost of an incorrect selection (but the one-stage allocation requires, by far, more sampling). Thus, the KG_1 stopping rule is myopic and leads to suboptimal performance due to early stopping, even though the KG_1 (aka LL_1) allocation has been shown to have attractive properties and performance in other settings. The use of the $EOC_{c,k}$ stopping rule is less myopic, and leads to an improved performance of the KG_1 allocation with that stopping rule.

The fully sequential KG_* allocation with the KG_* stopping rule performs somewhat better than the optimal two-stage allocation, whose reward is given by OER.

In the table we observe a real improvement with the $EOC_{c,k}$ stopping rule. When using the $EOC_{c,k}$ stopping rule, it does not seem to affect the expected net reward whether we use the sequential LL or KG_* allocation rule. The difference between those rules for this experiment is that LL samples slightly more and achieves a slightly smaller expected opportunity cost.

The differences in performance between the policies grows with k . As k grows, it seems likely that the approximations made by KG_* and KG_1 (to consider allocations to only a single system) become more restrictive and less justified, causing the difference in value between the best single system allocation and the best allocation overall to become larger. The $EOC_{c,k}$ stopping rule accounts for the benefits of allocations that split observations over multiple systems. This would explain the increased tendency to early stopping (not absolutely, but relative to $EOC_{c,k}$) observed in these rules with increased k .

Table 1: Performance of stopping rules for $k = 1$, $c = 1$ and $\sigma = 10^5$ calculated using Monte Carlo simulation with 10^6 samples. In the upper table, $t_0 = 1$, and in the lower table, $t_0 = 100$. Note that the KG_* and $EOC_{c,k}$ stopping rules are identical for $k = 1$, except for the more thorough way in which $EOC_{c,k}$ searches the set of budgets.

Stopping Rule ($\mu_0, t_0) = (0, 1)$	E[T]	E[Opportunity Cost]	E[Reward]	% suboptimality
estimated optimal boundary	64.33 ± 0.13	32.8 ± 0.4	39797.1 ± 0.4	—
$EOC_{c,k}$	29.90 ± 0.06	127.0 ± 1.2	39737.3 ± 1.2	0.15%
KG_* ($\tilde{\beta}^*$ stopping)	29.39 ± 0.06	128.6 ± 1.2	39736.2 ± 1.2	0.15%
KG_1	6.88 ± 0.01	1103.1 ± 5.4	38784.3 ± 5.4	2.54%

Note: PDE estimate is $V^{\pi^*}(0, t_0) \approx B(0, 0, t_0) = 39800$, bounds are $\overline{OER} = 39894$ and $\underline{OER} = 39613$.

Stopping Rule ($\mu_0, t_0) = (0, 100)$	E[T]	E[Opportunity Cost]	E[Reward]	% suboptimality
estimated optimal boundary	321.85 ± 0.25	263.0 ± 1.0	3404.6 ± 1.0	—
$EOC_{c,k}$	142.53 ± 0.11	612.1 ± 1.8	3234.8 ± 1.8	4.99%
KG_* ($\tilde{\beta}^*$ stopping)	136.51 ± 0.11	633.9 ± 1.9	3219.0 ± 1.9	5.45%
KG_1	10.53 ± 0.01	2504.5 ± 4.6	1474.4 ± 4.6	56.69%

Note: PDE estimate is $V^{\pi^*}(0, t_0) \approx B(0, 0, t_0) = 3407$, bounds are $\overline{OER} = 3989$ and $\underline{OER} = 3170$.

Table 2: Performance of sampling and stopping rules for $k > 1$, $c = 1$, $\sigma = 10^5$, $m = 0$ and $t_0 = 1$, calculated using Monte Carlo simulation with 10^6 samples.

	k=2	3	5	10	20
E[Reward]					
KG_1, KG_1	65728 ± 9	85324 ± 12	111810 ± 15	146761 ± 18	177983 ± 19
KG_*, KG_* ($\tilde{\beta}^*$ stopping)	67682 ± 3	88137 ± 4	115940 ± 5	152435 ± 6	184973 ± 6
KG_*, KG_* (exact KG_* stopping)	67691 ± 3	88150 ± 4	115944 ± 5	152464 ± 6	184984 ± 6
$KG_*, EOC_{c,k}$	67788 ± 2	88386 ± 2	116386 ± 3	153084 ± 3	185694 ± 4
$LL, EOC_{c,k}$	67792 ± 2	88406 ± 2	116425 ± 2	153157 ± 3	185736 ± 3
\overline{OER}	68104	88815	116970	153887	186748
\underline{OER}	67582	88079	115855	151940	183284
E[N]					
KG_1, KG_1	11.27 ± 0.01	14.57 ± 0.01	19.80 ± 0.01	30.09 ± 0.02	46.21 ± 0.02
KG_*, KG_* ($\tilde{\beta}^*$ stopping)	45.28 ± 0.07	55.45 ± 0.08	69.85 ± 0.09	96.74 ± 0.11	132.73 ± 0.12
KG_*, KG_* (exact KG_* stopping)	45.74 ± 0.07	55.96 ± 0.08	70.78 ± 0.09	97.73 ± 0.11	133.81 ± 0.12
$KG_*, EOC_{c,k}$	51.69 ± 0.08	71.96 ± 0.10	101.03 ± 0.13	140.54 ± 0.15	181.00 ± 0.16
$LL, EOC_{c,k}$	66.44 ± 0.11	94.47 ± 0.13	137.82 ± 0.15	238.58 ± 0.23	448.83 ± 0.43
\overline{OER}	280	420	700	1400	2800
E[OC]					
KG_1, KG_1	2364 ± 9	3477 ± 12	5141 ± 15	7096 ± 18	8719 ± 19
KG_*, KG_* ($\tilde{\beta}^*$ stopping)	376 ± 3	622 ± 4	960 ± 5	1354 ± 6	1642 ± 6
KG_*, KG_* (exact KG_* stopping)	367 ± 3	608 ± 4	956 ± 5	1324 ± 6	1630 ± 6
$KG_*, EOC_{c,k}$	264 ± 2	357 ± 2	484 ± 3	662 ± 3	873 ± 4
$LL, EOC_{c,k}$	245 ± 2	314 ± 2	408 ± 2	491 ± 3	562 ± 3
\overline{OER}	522	736	1116	1947	3463

6 DISCUSSION AND CONCLUSION

Ranking and selection in simulation has typically focused on frequentist or Bayesian rules that have statistical criteria for sampling and stopping. This paper used the cost of sampling to determine when to stop sampling. It extended previous KG/LL₁ work by examining more closely the expected value of the option to continue sampling, and not just the expected value of information of a single stage of sampling. It extended previous work on the economics of simulation by indicating how to handle sampling costs in the absence of discounting costs.

REFERENCES

- Bather, J. A. 1970. Optimal stopping problems for optimal control. *Adv. Appl. Probab.* 2:259–286.
- Branke, J., S. Chick, and C. Schmidt. 2007. Selecting a selection procedure. *Management Sci.* 53 (12): 1916–1932.
- Chen, C.-H. 1996. A lower bound for the correct subset-selection probability and its application to discrete event simulations. *IEEE Trans. on Automatic Control* 41 (8): 1227–1231.
- Chernoff, H. 1961. Sequential tests for the mean of a normal distribution. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Volume 1, 79–91. University of California Press.
- Chick, S., J. Branke, and C. Schmidt. 2009. Sequential sampling to myopically maximize the expected value of information. *INFORMS Journal on Computing*:to appear.
- Chick, S. E. 2006. Subjective probability and Bayesian methodology. In *Handbook in Operations Research and Management Science: Simulation*, ed. S. Henderson and B. Nelson. Elsevier.
- Chick, S. E., and P. Frazier. 2009. Beyond the knowledge gradient in sequential sampling for selection. Working paper.
- Chick, S. E., and N. Gans. 2009. An economic analysis of simulation selection problems. *Management Sci.* 55:to appear.
- 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.
- de Groot, M. H. 1970. *Optimal statistical decisions*. New York: McGraw-Hill.
- Frazier, P., and W. B. Powell. 2008. The knowledge-gradient stopping rule for ranking and selection. In *Proceedings of the 2008 Winter Simulation Conference*, ed. S. Mason, R. Hill, L. Mönch, O. Rose, T. Jefferson, and J. Fowler, 305–312. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- Frazier, P., W. B. Powell, and S. Dayanik. 2008. A knowledge-gradient policy for sequential information collection. *SIAM Journal on Control and Optimization* 47:2410–2439.
- Gupta, S. S., and K. J. Miescke. 1996. Bayesian look ahead one-stage sampling allocations for selecting the best population. *J. Statistical Planning and Inference* 54:229–244.
- Kim, S.-H., and B. L. Nelson. 2006. Selecting the best system. In *Handbook in Operations Research and Management Science: Simulation*, ed. S. G. Henderson and B. L. Nelson, 501–534. Elsevier.

AUTHOR BIOGRAPHIES

STEPHEN E. CHICK is a professor of Technology and Operations Management at INSEAD, and the Novartis Chair of Healthcare Management. He has worked in the automotive and software sectors prior to joining academia, and now teaches operations with applications in manufacturing and services, particularly the health care sector. He enjoys Bayesian statistics, stochastic models, and simulation. His web page is [<faculty.insead.edu/chick/>](http://faculty.insead.edu/chick/).

PETER FRAZIER is an Assistant Professor in the Department of Operations Research and Information Engineering at Cornell University, and received a Ph.D. in Operations Research and Financial Engineering from Princeton University in 2009. His research interest is in the optimal acquisition of information, with applications in simulation, medicine and operations management. His web address is people.orie.cornell.edu/pfrazier and his email address is pf98@cornell.edu.