Proceedings of the 2006 Winter Simulation Conference
L. F. Perrone, F. P. Wieland, J. Liu, B. G. Lawson, D. M. Nicol, and R. M. Fujimoto, eds.

SIMULATION SELECTION PROBLEMS: OVERVIEW OF AN ECONOMIC ANALYSIS

Stephen E. Chick

INSEAD
Technology and Operations Management Area
Boulevard de Constance
77305 Fontainebleau FRANCE

Noah Gans

OPIM Department – Wharton School
University of Pennsylvania
3730 Walnut Street, Suite 500
Philadelphia, PA 19104-6340 U.S.A.

ABSTRACT

This paper summarizes a new approach that we recently proposed for ranking and selection problems, one that maximizes the expected NPV of decisions made when using stochastic or discrete-event simulation. The expected NPV models not only the economic benefit from implementing a selected system, but also the marginal costs of simulation runs and discounting due to simulation analysis time. Our formulation assumes that facilities exist to simulate a fixed number of alternative systems, and we pose the problem as a "stoppable" Bayesian bandit problem. Under relatively general conditions, a Gittins index can be used to indicate which system to simulate or implement. We give an asymptotic approximation for the index that is appropriate when simulation outputs are normally distributed with known but potentially different variances for the different systems.

1 INTRODUCTION

We summarize recent work from Chick and Gans (2005). That work proposes a new approach to the simulation selection problem, one that focuses on the expected net present value of decisions that are supported by simulation, as opposed to more typical approaches that focus on statistical aspects involved with ranking and selection approaches to discrete optimization with simulation.

The premise is that managers must decide the operating characteristics of their firm's manufacturing, supply chain, or service delivery systems. Often the decision reflects the choice of one among a number of competing designs. To aid their decision-making managers may use stochastic or discrete event simulation. Simulation represents a widely-used and relatively low cost 'insurance' mechanism to estimate the performance of alternative systems and to improve the chances that the best system is implemented.

If there is a fixed set of k alternative designs, one must decide how long to simulate each alternative and, given the simulation results, which design to implement.

The most common approach for selecting the best of a finite set of simulated systems is ranking and selection, and the last 10 years have seen exciting progress. Both Bayesian and frequentist approaches are possible (Chick 2005, Chen et al. 2000, Kim and Nelson 2005), and large-scale numerical comparisons have identified the strengths and weaknesses of each approach (Branke et al. 2005). The application of these types of procedures has typically considered the time and financial costs of simulation separately from the value of the output: good procedures minimize the mean number of replications that a procedure needs to reach a desired level of evidence for correct selection. This is a flexible approach which allows one to assess a wide variety of operational and other measures of system performance. The assessment does not usually consider the financial costs of the analysis itself.

When system and simulation results are themselves financial measures, as when simulation is used at a tactical or strategic rather than operational level, a more direct economic approach may be appropriate. If the manager's goal is to maximize the expected net present value (NPV) of high-level system design choices, then the manager is faced with two countervailing costs. On the one hand, uncertainty about the expected NPV of each alternative compels the manager to simulate more to reduce the opportunity costs associated with an incorrect selection. On the other, a lengthy simulation analysis incurs direct costs and reduces the NPV of the system that is ultimately implemented, due to discounting.

We can formulate and solve a simulation selection problem in which the manager seeks to maximize expected NPV. Our formulation is Bayesian: we assume that the manager has prior beliefs concerning the distribution of the NPV of each of the alternatives and that she uses simulation output to update these beliefs. The system which the manager ultimately chooses to implement maximizes expected NPV with respect to the posterior distributions of her beliefs, as well as analysis costs and discounting costs.

Section 2 defines the problem, and Section 3 identifies links to existing simulation and probability literature. Section 4 observes that the simulation selection problem can be expressed as a variant of the multi-armed bandit problem, a so-called stoppable bandit process. As a result, a variant of the well known Gittins-index policy is an optimal way to decide which system to simulate next and when to stop simulating in favor of implementing a system. At each step of the sequential procedure, the system with the largest Gittins index is identified. If the statistics of the system with the largest Gittins index fall into a continuation set, then that system is simulated one more time and its Gittins index is updated. If the statistics fall outside of the continuation set, then simulation stops and that system is implemented. The boundary of the continuation set therefore determines whether one should "learn" (by simulating) or "earn" (by implementing a system). Section 4 requires few distributional assumptions other than joint independence and bounded expectations. Simulation run times of the different systems are initially assumed to be equal.

Gittins indices are typically difficult to compute exactly in Bayesian problems. Section 5 motivates asymptotic approximations for the Gittins index associated with normally distributed simulation output. The approximations apply when simulation output is independent across systems with unknown means and known, potentially different, variances. The approximation is determined by the solution of a free boundary problem for a heat equation that shares characteristics with financial options. The selection procedure that is implied is presented in Section 6.

The theory is applied to examples in Section 7. Section 8 indicates that some of the strong sampling assumptions made in Section 5 can be relaxed. See Chick and Gans (2005) for a fuller description, proofs, and an explanation of how these results can be implemented.

2 PROBLEM DESCRIPTION

A manager seeks to develop one of k projects, labelled $i = 1, \ldots, k$. The net present value (NPV) of each of the i projects is not known with certainty, however. The manager wishes to develop the project which maximizes her expected NPV, or to do nothing if the expected present value of all projects is negative. We represent the "do nothing" option as $i = 0$ with a sure NPV of zero.

2.1 Uncertain Project NPV's

Let X_i be the random variable representing the NPV of project i, where $X_0 \equiv 0$. If the manager is risk neutral and the distributions of all X_i's are known to her, then she will select the project with the largest expected NPV, $i^* = \arg\max_i\{\mathrm{E}[X_i]\}$.

Although we model NPVs as simple random variables, the systems that generate them may be quite complex. For example, a particular project's sequence of cash flows may involve the composition of several interrelated random processes describing the evolution of investments, $\mathcal{I}(t)$, revenues, $\mathcal{R}(t)$, and operating costs, $\mathcal{O}(t)$, over time. Nevertheless, given a continuous-time discount rate $\delta > 0$, each realization of these processes, ω_i, yields a sample $X(\omega_i) = \int_0^\infty [\mathcal{R}(t,\omega_i) - \mathcal{O}(t,\omega_i) - \mathcal{I}(t,\omega_i)]e^{-\delta t}dt$.

It may also be the case that the distributions of the X_i's are not known with certainty by the manager. Rather, she may believe that a given X_i may come from one of a family of probability distributions, $P_{X_i|\theta_i}$, indexed by parameter θ_i. We model her belief as taking the form of a probability distribution on θ_i, which we call P_{Θ_i}. For example, the manager may believe that X_i is normally distributed with a known variance, σ_i^2, but unknown mean. Then P_{Θ_i} represents a probability distribution for the mean. To ease notation, we will sometimes refer to the distribution as Θ_i. In this case, the expected NPV of project $i > 0$ is $\mathrm{E}[X_i] = \mathrm{E}[X(\Theta_i)] \triangleq \iint X(\theta_i) dP_{X_i|\theta_i} dP_{\Theta_i}$. We denote the vector of distributions for the projects by $\mathbf{\Theta} = (\Theta_1, \ldots, \Theta_k)$.

Remark 1 *This notation is consistent with the literature on the Bayesian bandit problem (Chang and Lai 1987, for example). It is not consistent with much of the literature on simulation selection. In the latter, capital Θ_i would refer to a random variable, rather than to a distribution function.*

2.2 Simulation to Select the Best Project

If the distributions of the X_i's are not known, then the manager may be able to use simulation as a tool to reduce distributional uncertainty, before having to decide which project to develop. She may decide to simulate the outcome of project i a number of times, and she views the result of each run as a sample of X_i. She uses Bayes' rule to update her beliefs concerning Θ_i.

We model the running of simulations as occurring at sequence of discrete stages $t = 0, 1, 2, \ldots$, and we represent Bayesian updating of prior beliefs and sample outcomes, $\{(\mathbf{\Theta}_t, \mathbf{X}_t) \,|\, t = 0, 1, \ldots\}$ as follows. If project $i > 0$ is simulated at stage t with sample outcome $x_{i,t}$, then $X_{i,t} = x_{i,t}$ and $X_{j,t} = 0$ for all $j \neq i$. In turn, Bayes' rule is used to determine $\mathbf{\Theta}_{t+1}$:

$$dP_{\Theta_{i,t+1}}(\theta_i \,|\, x_{i,t}, \Theta_{i,t}) = \frac{dP_{X_i\,|\,\theta_i}(x_{i,t} \,|\, \theta_i)\, dP_{\Theta_{i,t}}(\theta_i)}{\int_{\theta_i} dP_{X_i\,|\,\theta_i}(x_{i,t} \,|\, \theta_i)\, dP_{\Theta_{i,t}}(\theta_i)}$$

for all $\theta_i \in \Omega_{\Theta_i}$, while $\Theta_{j,t+1} = \Theta_{j,t}$ for all $j \neq i$. So the evolution of the manager's beliefs regarding the distribution of outcomes of each project is Markovian. We also assume that simulation results, hence the evolution of the manager's beliefs, are independent from one project to the next.

If, in theory, simulation runs could be performed at zero cost and in no time, then the manager might simulate each of the k systems infinitely, until all uncertainty regarding the θ_i's was resolved. At this point the problem would revert to the original case in which the distributions and means of the X_i are known.

But simulation runs do take time and cost money. We assume that each run of system i costs \$$c_i$ and takes η_i units of time to complete. Thus, given a continuous-time discount rate of $\delta > 0$, the decision to simulate system i costs the manager c_i plus a reduction of $\Delta_i = \int_0^{\eta_i} e^{-\delta s} ds < 1$ times the expected NPV of the (unknown) project that is eventually chosen.

There may also be associated up-front costs associated with the development of the simulation tool, itself. For example it may cost time and money to develop the underlying simulation platform, independent of which projects end up being evaluated. Additional costs may be required to be able to simulate particular projects. Furthermore, these project-specific costs may be inter-related.

For the moment, we make two simplifying assumptions regarding the costs of simulation. First, we ignore all up-front costs for the simulation tool, assuming that the necessary facilities exist to simulate all k projects. Second, we assume that $\eta_i \equiv \eta$ for all k projects. This allows us to define a common $\Delta \equiv \Delta_i$ for the projects as well. Section 8 argues that these assumptions may be relaxed.

Even with these simplifications, the availability of a simulation tool to sample project outcomes makes the manager's problem much more complex. Rather than simply choosing the project that maximizes expected NPV, she must choose a sequence of simulation runs and, ultimately select a project, so that the discounted stream of costs and terminal expected value, together, maximize expected NPV.

We define a number of indices in order to track the manager's choices as they proceed. Let $T \in \{t = 0, 1, 2, \ldots\}$ be the stage at which the manager selects a system to implement. For $t < T$, define $i(t) \in \{1, \ldots, k\}$ to be the index of the project simulated at time t, and define $I(T) \in \{0, \ldots, k\}$ to be the ultimate choice of project.

A *selection policy* is the choice of a sequence of simulation runs, a stopping time, and a final project. Define Π to be the set of all *non-anticipating* selection policies, whose choice at time $t = 0, 1, \ldots$ depends only on system history up to t: $\{\mathbf{\Theta}_0, \mathbf{X}_0, \ldots, \mathbf{\Theta}_{t-1}, \mathbf{X}_{t-1}, \mathbf{\Theta}_t\}$. Given prior distributions $\mathbf{\Theta} = (\Theta_1, \ldots, \Theta_k)$ and a policy $\pi \in \Pi$, the expected discounted value of the future stream of rewards is

$$V^\pi(\mathbf{\Theta}) = \mathrm{E}_\pi \left[\sum_{t=0}^{T-1} -\Delta^t c_{i(t)} + \Delta^T X_{I(T),T} \mid \mathbf{\Theta}_0 = \mathbf{\Theta} \right], \tag{1}$$

where $X_{I(T),T}$ is the unknown NPV of the selected system, $I(T)$, when a system is selected (at time T).

Formally, we define the manager's *simulation selection problem* to be the to choice of a selection policy $\pi^* \in \Pi$ that maximizes $V^{\pi^*}(\mathbf{\Theta}) = \sup_{\pi \in \Pi} V^\pi(\mathbf{\Theta})$.

3 RELATED LITERATURE

Two broad classes of research are related to this work. One is the ranking and selection literature, the other is the bandit and optimal stopping literature. Both have substreams.

None of the ranking and selection literature has explicitly accounted for discounting costs due to elapsed simulation times. Still, two lines of thought in ranking and selection are related to this work, either through their use of sampling costs, or of diffusion approximations to simulation output.

Chick and Inoue (2001) provided two-stage procedures whose second stage allocation can trade off the cost of sampling with an approximation to the Bayesian expected value of information (EVI) of sampling. That approximation assumes a large number of samples (small sampling costs), and a Bonferroni-like bound for the EVI. The EVI is measured with respect to either the posterior expected opportunity cost (EOC) of a potentially incorrect selection, or the posterior probability of incorrect selection (PICS).

The indifference-zone (IZ) approach provides a frequentist guarantor of selection procedure effectiveness (Kim and Nelson 2005). Almost all IZ procedures focus on probability of correct selection (PCS) guarantees for each problem instance within a given class, in which probability is defined as the (frequentist) probability the best system is correctly selected for a given problem. Recent innovations for IZ procedures (Kim and Nelson 2001) theoretically justify and use diffusion approximations to improve efficiency in a sequential screening procedure.

In another stream of literature, Gittins (1979) offers an early account of optimal of dynamic allocation indices (later called Gittins indices) for infinite horizon, discounted multi-armed bandit problems. Glazebrook (1979) provides sufficient conditions under which these index results apply to reward streams derived from stoppable arms. Gittins (1989) shows that the results of Glazebrook (1979) hold under a slightly weaker set of assumptions.

Gittins indices are difficult to compute exactly. Chang and Lai (1987) derives approximations for the Gittins index for the infinite horizon discounted "Bayesian bandit" problem. Brezzi and Lai (2002) uses a diffusion approximation for the Gittins index of a Bayesian bandit, motivated by ground breaking work on composite hypothesis tests (Chernoff 1961, Breakwell and Chernoff 1964).

Chick and Gans (2005) provide a new approach to the simulation-selection problem by directly accounting for the economics of its discounting and simulation expenditures. The idea is to link the selection problem to the bandit literature, and to develop diffusion approximations for Gittins indices for special distributions.

4 SIMULATION SELECTION AND BANDITS

The simulation selection problem is closely related to a class of sequential decision problem known as the multi-armed bandit problem. Chick and Gans (2005) demonstrate that simulation selection problems can be reduced to multi-armed bandits. The result implies that a class of simple, index-based policies is optimal for simulation selection.

In the discounted multi-armed bandit problem, a decision-maker chooses repeatedly among a finite set of mutually-independent Markov chains that are indexed $i = 1, \ldots, k$. A choice of chain i at stage t yields an expected reward that is specific to the state of chain i, and it initiates a state transition for chain i. The $k-1$ chains not chosen at stage t remain in their current states and earn no rewards. The objective is to maximize the expected sum of discounted rewards over an infinite horizon (Gittins 1989).

For the case in which expected one-period rewards are bounded, so that $R_i(\Theta_i) < \infty$ for almost all $\Theta_i \in \Omega_{\Theta_i}$, $i = 1, \ldots, k$, Gittins and co-workers proved two important sets of results which are relevant for our problem. First, Gittins and Jones (1974) demonstrated that there exists a state-dependent index for each arm, $G_i(\Theta_i)$, which is independent of all other arms, such that it is optimal to choose at each stage, t, the arm whose index is the greatest among all arms. Second, Gittins and Glazebrook (1977) and Gittins (1979) demonstrated that this so-called Gittins index has an appealing form.

The structure of the simulation selection problem defined in Section 2 is clearly close to that of the multi-armed bandit. Both have discrete-time discounting, independent projects, and Markovian state transitions.

At the same time, the simulation selection problem includes a stopping time, T, which changes the problem's decision structure. If, as in the simulation selection problem, a "zero" arm is included, then the bandit problem has $k+1$ actions available for all $t = 0, 1, \ldots$. In contrast, for $t \leq T$ the simulation selection problem has $2k+1$ actions available — decide $t < T$ and choose arm $i(t) \in \{1, \ldots, k\}$ to simulate, or decide $t = T$ and choose an arm $I(t) \in \{0, \ldots, k\}$ to implement — and for $t > T$ no actions are available.

In fact, the added stopping decision makes the simulation selection problem an example of what Glazebrook (1979) calls a *stoppable family of alternative bandit processes*. Chick and Gans (2005) apply the results of Glazebrook (1979) to show that the simulation selection problem can be effectively reduced to a multi-armed bandit, so that a Gittins indices result can be used to solve the simulation selection problem in Equation (1). The idea is to modify the simulation stopping problem so that the optimal value function satisfies the so-called Bellman equation:

$$
\begin{aligned}
& V_i^{\pi_i^*}(\Theta_{i,t}) \\
&\quad = \max\{-c_i + \Delta \mathrm{E}[V_i^{\pi_i}(\Theta_{i,t+1}) \mid \Theta_{i,t}, t \neq T_i], \\
&\qquad\quad (1-\Delta)\mathrm{E}[X(\Theta_{i,t})] + \Delta \mathrm{E}[V_i^{\pi_i^*}(\Theta_{i,t})]\Big\} \\
&\quad = \max\{-c_i + \Delta \mathrm{E}[V_i^{\pi_i}(\Theta_{i,t+1}) \mid \Theta_{i,t}, t \neq T_i], \\
&\qquad\quad \mathrm{E}[X(\Theta_{i,t})]\}\,. \qquad (2)
\end{aligned}
$$

Chick and Gans (2005) show how to transform optimal policies for the transformed stopping problems into an optimal policy for the original simulation stopping problem. They also prove that there is a Gittins index policy that is optimal, *and* that the Gittins index for each arm is proportional to the optimal value function in Equation (2). The optimal value function, if it can be computed for each project individually, can therefore serve as a Gittins index, as is needed to optimally solve the simulation selection problem.

5 GITTINS INDEX APPROXIMATION FOR NORMAL OUTPUT WITH KNOWN VARIANCE

At a high level, the optimal policy is straightforward. At each t it compares the expected discounted value of optimal stopping for each project and selects the one with the highest $V_i^{\pi_i^*}(\Theta_{i,t})$. If that project is $i = 0$, then abandonment is most favorable and no project is further simulated or implemented. If the best project is some $i > 0$ and $t = T_i^*$ then project i is implemented. Otherwise, project i is simulated, Bayes' rule is used to calculate $\Theta_{i,t+1}$, $V_i^{\pi_i^*}(\Theta_{i,t+1})$ is determined, and the comparison begins again.

The foundation of the optimal policy is the repeated determination of the various $V_i^{\pi_i^*}(\Theta_{i,t})$'s, as in Equation (2). The calculation of each $V_i^{\pi_i^*}(\Theta_{i,t})$, what we call the optimal expected discounted reward (OEDR), is itself a difficult task for which exact solutions are not available.

Chick and Gans (2005) develop diffusion approximations for these indices in the spirit of Chernoff (1961) and Breakwell and Chernoff (1964). The diffusion approximations are asymptotically appropriate when the discount rate over the duration of a simulation replication is small, as is usually the case in simulation. Repeated sampling leads to realizations of a scaled Brownian motion with drift. This section summarizes the diffusion approximations of Chick and Gans (2005).

The calculation of the OEDR involves the solution of a so-called free boundary problem for a heat equation that is obtained from the diffusion approximation. The boundary is "free" since it is determined by equating the two maximands in the value function, rather than on a known, pre-specified boundary. A comparison of the maximands in the continuous-time analogue of Equation (2) determines the free boundary between a continuation set, $\mathcal{C}$, in which it is optimal to continue simulating a project, and a stopping set, in which it is optimal to stop simulating and implement the project. If the boundary is never reached, then the NPV of simulating forever is better than the expected NPV of implementing a poor system.

The following approximation assumes that each project's simulation output is normally distributed with a known variance. While this assumption may not satisfy the uniform boundedness condition, the analysis below results in a well-defined finite OEDR when the initial prior distributions for the unknown means are proper.

We will calculate the OEDR of a single project, so we drop the system's index, i, from subscripts to simplify notation. Specifically, simulation replications X_j are i.i.d. $\texttt{Normal}\left(\theta, \sigma^2\right)$ for $j = 1, 2, \ldots$, with a known finite variance σ^2 and unknown mean θ. We suppose that θ has a $\texttt{Normal}\left(\mu_0, \sigma_0^2\right)$ prior distribution. Some of these restrictive assumptions will be relaxed in Section 8.

Define $n_0 = \sigma^2/\sigma_0^2$, and redefine $t = n_0 + n$, where n is the number of simulation observations seen so far for the single system in question. Set $Y_t = n_0\mu_0 + \sum_{j=1}^{n} X_j$. This transformation conveniently makes the posterior distribution of θ a $\texttt{Normal}\left(Y_t/t, \sigma^2/t\right)$ distribution, and will help to find an optimal stopping time when there is $k = 1$ system. (When we return to the original problem with $k \geq 1$ in Section 6 below, each system will have its own time progression, t_i, and the time index will be $t = \sum_{i=1}^{k} t_i$.)

Chick and Gans (2005) show the following points.

- The Gittins index can be approximated by solving the following partial differential equation:

$$0 = -c - \delta B + \frac{y}{t} B_y + B_t + \frac{\sigma^2}{2} B_{yy} \quad (3)$$

  on the continuation set $\mathcal{C}$, along with the condition that the (free) boundary, $\partial\mathcal{C}$, of $\mathcal{C}$ be determined by equating the value function with the value of stopping to implement,

$$B(y,t) \quad = \quad D(y,t), \text{ on } \partial\mathcal{C}. \quad (4)$$

  Here, $B(y,t)$ is a diffusion approximation to the OEDR, where y is the realization of Y_t at time t, and subscripts on B represent partial derivatives.
- When $c = 0$, Equations (3-4) represents what might be called a perpetual American call option on regular (not geometric) Brownian motion, with unknown drift that is inferred through time.
- The diffusion in Equation (3) has three parameters. One can reparameterize the problem so that only *one standardized* problem must be solved to handle *any* values for $\delta > 0$, $c \geq 0$, and $\sigma > 0$.
- The diffusion approximation is reasonable under conditions that are typically valid in simulation.

Chick and Gans (2005) describe numerical techniques, based on ideas from Chernoff and Petkau (1986), that implement the necessary calculations in a tractable amount of time. The boundary depends upon a function that they call $b_1(\cdot)$, whose argument is a function of the number of samples that have been observed.

The OEDR of the standardized problem is defined by a function $B_1(\cdot,\cdot)$ with two arguments that are functions of the sample mean of the simulation output and the number of replications that have been observed so far.

They denote the OEDR of the original (not standardized) problem by $\mathcal{V}_{\ell,i}$ for each alternative $i = 0, 1, \ldots, k$, where $\ell = 1$ when $c = 0$ (case 1), and $\ell = 2(\kappa)$ when $c > 0$ (case 2), and the parameter κ depends upon the values of δ, c and σ. The continuation set $\mathcal{C}$ and OEDR $\mathcal{V}_{\ell,i}$ are readily-computed functions of $b_1(\cdot)$, $B_1(\cdot,\cdot)$, and the parameters

$$\alpha = \delta^{1/2}\sigma^{-1}, \beta = \delta^{-1/2}\sigma^{-1}, \gamma = \delta, \text{ and } \kappa = \delta^{3/2}\sigma c^{-1}.$$

6 SIMULATION SELECTION PROCEDURE

Table 1 is the simulation selection procedure that results from the above analysis, which assumes independent, normally distributed output with known variances.

Table 1: Simulation Selection Procedure

1. Identify economic parameters (discount factor $\delta > 0$ and costs $c_i \geq 0$ per replication). Provide prior distributions for each of the k alternative systems and initialize y_i, t_i for each system (see below). Include system 0 as an option ('do nothing' option with a guaranteed NPV of $\mathcal{V}_0 = 0$) if appropriate.
2. Compute the OEDR $\mathcal{V}_{\ell,i}$ for each alternative $i = 0, 1, \ldots, k$, and set $t = \sum_{i=1}^{k} t_i$, where $\ell \in \{1, 2(\kappa)\}$, depending on the values of c_i, δ, σ_i (see the last paragraph of Section 5).
3. While (a system is not yet implemented):
   (a) Increment $t \leftarrow t + 1$.
   (b) Identify the system with largest index, $i(t) = \arg\max_{i=0,1,\ldots,k} \mathcal{V}_{\ell,i}$ (break ties randomly).
   (c) If $(y_{i(t)}, t_{i(t)})$ is not in the continuation set, $\mathcal{C}_{\ell,i}$, then stop simulating and implement the appropriate system, otherwise run a simulation for that system to get output $x_{i,t_{i(t)}+1}$.
   (d) Update $y_{i(t)} \leftarrow y_{i(t)} + x_{i,t_{i(t)}+1}$; $t_{i(t)} \leftarrow t_{i(t)} + 1$ and $\mathcal{V}_{\ell,i(t)}$ for system $i(t)$.

Step 1 of the procedure requires prior distributions for the unknown means. There are at least two options for generating these initial priors. Expert judgment may provide a prior distribution, $\texttt{Normal}\left(\mu_{0i}, \sigma_{0i}\right)$, for the unknown means θ_i, for $i = 1, 2, \ldots, k$. If that is done, initialize $t_i = \sigma_i^2/\sigma_{0i}^2$ and $y_i = t_i\mu_{0i}$. Alternatively, default assessments can be implemented by running n_0 replications for each system and setting $y_i = \sum_{j=1}^{n_0} x_{i,j}$ and $t_i = n_0$.

Chick and Gans (2005) describe how to compute the OEDR $\mathcal{V}_{\ell,i}$ and boundary, $b_1(\cdot)$. The continuation set in Step 3c corresponds to $y_{i(t)}/t_{i(t)} < \beta_{i(t)}^{-1} b_1(1/\gamma t_{i(t)}) - c_{i(t)}/\delta = \sigma_{i(t)}\sqrt{\delta} b_1(1/\delta t_{i(t)}) - c_{i(t)}/\delta$. This formulation allows for different sampling costs c_i for each system.

It is possible that the system being implemented is the 'do nothing' option, which has a known NPV of $\mathcal{V}_0 = 0$. Alternative systems with a known, positive expected NPV can be included by replacing the 0-arm with the option of implementing that better alternative if the maximum OEDR is $\mathcal{V}_0 =$ known expected NPV (e.g., for comparing mutually exclusive alternatives with an existing system).

7 SIMULATION SELECTION EXAMPLES

This section comments on what the theory developed above implies for realistic scenarios. Chick and Gans (2005) indicate how to numerically approximate the Gittins indices, and provide more graphs to supplement the examples here.

The first example shows how large the simulation output mean must be before one stops to implement a system. Assume that a firm uses a discount rate of 10%/year, that the output of replications of a single simulated alternative has standard deviation $\sigma = \$10^6$ and has no marginal cost ($c = 0$), but requires about 5.3 minutes to run. That simulated time makes the discount rate per replication equal to $\delta = 10^{-6} (= 5.256 \times 0.10/365/24)$, so 10^6 replications are required to get to scaled time $\tau = 1$. Figure 1 indicates that simulation should stop after 16 replications if the sample mean is $y_t/t = \$10^6$ (which corresponds to a z-score of $z = (y_t/t)/(\sigma/\sqrt{t}) = 4$).

If the simulated system is implemented (with $y_t/t > 0$), then the posterior probability of incorrect selection, PICS, is the probability that the unknown mean is less than the value of not implementing any system (NPV = 0). Recall that the posterior probability for the unknown mean is $\texttt{Normal}\left(y_t/t, \sigma^2/t\right)$, with density $p_t(\theta) = (\sqrt{t}/\sqrt{2\pi\sigma^2}) e^{-(\theta - y_t/t)^2 t/2\sigma^2}$. If $z = 4$ when $t = 16$, then PICS $= \int_{-\infty}^{0} p_t(\theta) d\theta = \Phi[-z] = 3 \times 10^{-5}$. If the simulated system is selected as best, but the mean turns out to be $\theta < 0$, then the opportunity cost is $0 - \theta$. If $z = 4$ when $t = 16$, the posterior expected opportunity cost of potentially incorrect selection is EOC $= \int_{-\infty}^{0} (0 - \theta) p_t(\theta) d\theta = \Psi[z]\sigma/\sqrt{t} = 2$. One stops after 855 replications (3.12 days) if $y_t/t = \$10^5$ ($z = 2.92$; PICS $= 1.7 \times 10^{-3}$; EOC $= 17$), and after 29830 replications (108 days) if $y_t/t = \$10^4$ ($z = 1.73$; PICS $= 4.2 \times 10^{-2}$; EOC $= 99$). In this example, a greater potential upside means that one is willing to 'stop simulating and start building' sooner, but a more stringent level of evidence for correct selection is required (a higher z-score, meaning a lower PICS and EOC).

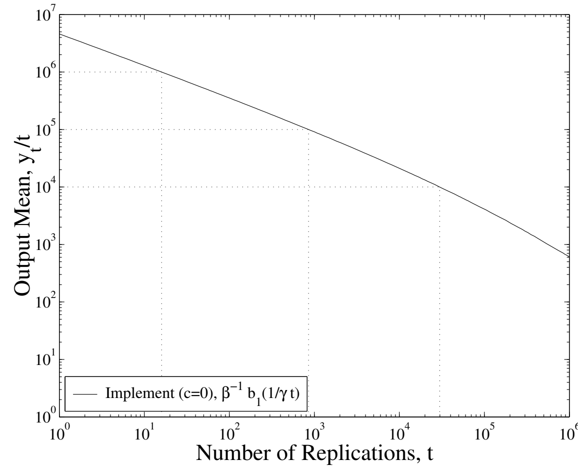

Figure 1: Threshold for Stopping Simulation in Favor of Implementing a Single Project ($k = 1$, $\sigma = \$10^6$; $\delta = 10^{-6}$, or 10% Per Year; $c = 0$).

Figure 1 shows that a long delay before implementing a system can occur if variable costs of simulation are ignored. Our second example shows that one stops simulating earlier when these variable costs are included. Suppose now that variable costs attributed to further simulation are \$1/hour (e.g., additional computer time), and all other parameters are as in the first example. This makes the cost per replication $c = 5.256 \times \$1/60 \approx \$0.0876$, and the stopping boundary shifts down by $c/\delta \approx \$87.6$ thousand.

With those sampling costs, Figure 2 shows that the inclusion of positive sampling costs reduces the size of the continuation region. Other things equal, one would stop sampling earlier, the greater the marginal cost of additional samples, c. Figure 2 also plots a lower boundary. Below that lower boundary, one would prefer to stop simulating the single simulated option in order to implement the 0-arm. That is, if the sample mean of the simulation output is sufficiently low, and the variable cost of simulation replications is nonzero, then one would prefer to stop simulating and receive a sure NPV of zero rather than continuing to lose money with a simulation analysis of an unfavorable system.

Chick and Gans (2005) further discuss how σ, c and δ interact to determine the continuation region in the context of stationary simulations.

The last example illustrates how the OEDR can be used to identify which system to simulate or implement when there are multiple systems. Figure 3 plots the OEDR $\mathcal{V}_1(y_{t_i}, t_i) = \beta_i^{-1} B_1(\beta_i(y_{t_i}/t_i + c_i/\delta), 1/\delta t_i) - c_i/\delta$, assuming the basic setup of the first example, which has $c_i = 0$ and $\beta_i = \delta^{-1/2}\sigma_i^{-1} = 10^{-3}$. If the parameters are different for each system, then a different scaling for the OEDR would result for each system.

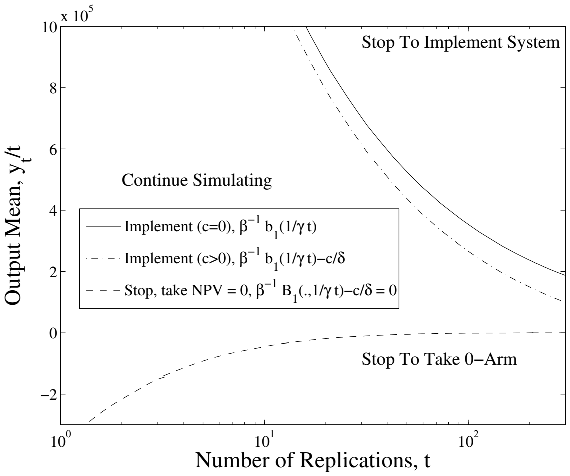

Figure 2: One Stops Sampling Earlier in Favor of Implementing when the Marginal Cost of Sampling is \$1/Hour Rather Than Free ($\sigma = \$10^6$; $\delta = 10^{-6}$, or 10% Per Year).

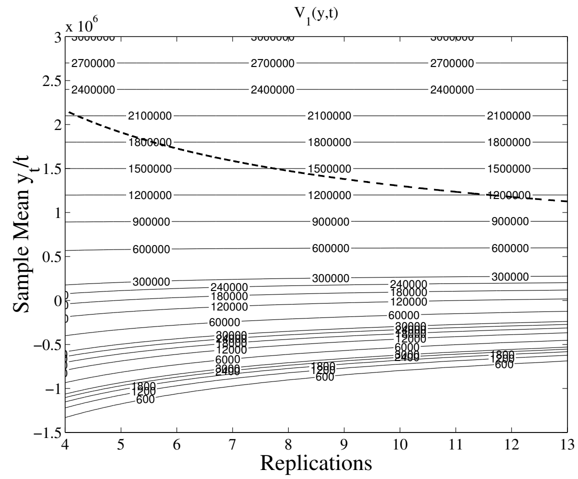

Figure 3: Contours of OEDR, $\mathcal{V}_1(y_{t_i}, t_i)$ for Last Example, with Dashed Stopping Boundary.

If default assessments are used for each unknown mean NPV, then some number (e.g., 6) of replications is run for each system, and the sample average (y_i/t_i) and the number of replications (t_i) are initialized. The system with the highest sample mean is initially the one with the highest OEDR (since the number of replications for each is initially the same). That system is simulated until either the sample mean crosses the implementation boundary (resulting in simulation stopping and a system being implemented), or until its OEDR drops below the OEDR of another system (resulting in a change in which system gets simulated). The process of simulating a single system, updating its OEDR, and choosing whether to continue simulating, to switch which system gets simulated, or to stop and implement a system then repeats until a system is implemented.

8 EXTENSIONS

Section 5 assumed jointly independent Gaussian output with known variances. Chick and Gans (2005) argue that those approximations can be generalized to the following scenarios, if a few additional technical conditions hold.

- Samples from a one-parameter member of the exponential family of distributions can be handled (exponential, Bernoulli, Poisson, …).
- Autocorrelated output, if a "batch means" analysis is appropriate. Such autocorrelation is typical for the analysis of many queueing or inventory systems.
- Different runtime durations across systems.

The analysis above assumes that the cost and time for implementing the simulation tool are zero. More recent results with a PhD student show that the results can also be extended to account for the monetary cost, $\$\bar{g}_i$ and $\bar{h}_i$ units of time required to build the simulation tool for system i under certain conditions (one still has a stoppable bandit problem and computable Gittins indix approximations if the time and money spent building the simulation tool for one system is independent of the costs of the tools for each other system).

The Gittins index results apply when samples are independent and normally distributed with *unknown* variance, but the approximations of Section 5 do not apply when the variance is unknown. A nonoptimal *ad hoc* solution would be to plug in the sample variance for the true variance, or apply some fudge factor (e.g., by plugging in the variance of a Student random variable with the appropriate degrees of freedom, $\nu = t_i - 1$, for the known variance). A better approximation for the Gittins index of simulation selection problems when the variance is unknown would be useful.

9 DISCUSSION AND CONCLUSIONS

This paper responds to the question of how to link financial measures (a firm's discount rate, the marginal cost of simulations) to the optimal control of simulation experiments that are designed to inform operational decisions. The broad answer, that index-based policies are optimal, is theoretically valid over a broad range of distributional assumptions, provided the observations are independent across projects and individual trials.

Asymptotically optimal approximations are appropriate when simulation output is independent and normally distributed with known variance. Some generalizations of that distributional assumption are available.

The paper perhaps raises more questions than it answers, however, from both the business and simulation perspectives. From a business perspective, we did not include the fixed costs of developing simulation models, or even of deciding what types of alternative configurations should be studied, or how many. We did not address the issue of first-mover advantage, or hard project due dates. Finite time horizons would invalidate basic assumptions of the bandit results, but the dynamic programming formulation espoused here may perhaps lead to insights.

From a simulation perspective, we did not account for common random numbers across systems. We also did not account for unknown variances. For that, one could try a plug-in estimator of the sample variance for the variance, but this may cause some inefficiency when the number of observations is small.

This work appears to open a number of research questions in simulation optimization. Much current research focuses on asymptotic convergence guarantees for the probability of correct selection. This paper suggests that an alternative approach may be useful: maximizing the expected discounted NPV of decisions based on simulation analysis, even at the expense of potentially incorrect selections.

ACKNOWLEDGMENTS

The research of Noah Gans was supported by the Fishman-Davidson Center for Service and Operations Management. Both authors appreciate the support of the Wharton-INSEAD Alliance. We appreciate the help of Işıl Yıldırım.

REFERENCES

Branke, J., S. E. Chick, and C. Schmidt. 2005. New developments in ranking and selection, with an empirical comparison of the three main approaches. In *Proceedings of the 2005 Winter Simulation Conference*, ed. M. Kuhl, N. Steiger, F. Armstrong, and J. Joines, 708–717.

Breakwell, J., and H. Chernoff. 1964. Sequential tests for the mean of a normal distribution II (large t). *Annals of Mathematical Statistics* 35:162–163.

Brezzi, M., and T. L. Lai. 2002. Optimal learning and experimenation in bandit problems. *Journal of Economic Dynamics & Control* 27:87–108.

Chang, F., and T. L. Lai. 1987. Optimal stopping and dynamic allocation. *Advances in Applied Probability* 19:829–853.

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: Theory and Applications* 10 (3): 251–270.

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.

Chernoff, H., and A. J. Petkau. 1986. Numerical solutions for Bayes sequential decision problems. *SIAM Journal on Scientific and Statistical Computing* 7 (1): 46–59.

Chick, S. E. 2005. 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 N. Gans. 2005. An economic analysis of simulation selection problems. INSEAD/Wharton Alliance Working Paper, Fontainebleau, France.

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.

Gittins, J. C. 1979. Bandit problems and dynamic allocation indices. *Journal of the Royal Statistical Society, Series B* 41:148–177.

Gittins, J. C. 1989. *Multi-armed bandit allocation indices*. New York: Wiley.

Gittins, J. C., and K. D. Glazebrook. 1977. On Bayesian models in stochastic scheduling. *Journal of Applied Probability* 14:556–565.

Gittins, J. C., and D. M. Jones. 1974. A dynamic allocation index for the sequential design of experiments. In *Progress in Statistics*, ed. J. Gani, K. Sarkadi, and J. Vincze. North-Holland.

Glazebrook, K. D. 1979. Stoppable families of alternative bandit processes. *Journal of Applied Probability* 16:843–854.

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

Kim, S.-H., and B. L. Nelson. 2005. Selecting the best system. In *Handbook in Operations Research and Management Science: Simulation*, ed. S. Henderson and B. Nelson. Elsevier.

AUTHOR BIOGRAPHIES

STEPHEN E. CHICK is an Associate Professor of Technology and Operations Management at INSEAD. 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/>.

NOAH GANS is an Associate Professor in the OPIM Department at the Wharton School. He worked in business consulting before entering academia, and now teaches service operations management. He is interested in call center operations and enjoys stochastic models and applied probability. His email address is <gans@wharton.upenn.edu>.