Sequential Sampling for Selection: The Undiscounted Case

Stephen E. Chick¹
Peter I. Frazier²

¹Technology & Operations Management, INSEAD
²Operations Research & Information Engineering, Cornell University

Monday November 8, 2010
INFORMS Annual Meeting
Austin
We have a discrete event simulator that can simulate the consequences of alternative real-world decisions, e.g.,

- Designs of a queuing network.
- Inventory policies for a supply chain.
- Pricing strategies for a revenue management problem.

Goal: find an alternative that works well, according to the simulator.

Our simulator needs significant time to accurately characterize an alternative, and we do not have enough time to do so for each one.

Which alternatives should we simulate and for how long?
We have a discrete event simulator that can simulate the consequences of alternative real-world decisions, e.g.,
- Designs of a queuing network.
- Inventory policies for a supply chain.
- Pricing strategies for a revenue management problem.

Goal: find an alternative that works well, according to the simulator.

Our simulator needs significant time to accurately characterize an alternative, and we do not have enough time to do so for each one.

Which alternatives should we simulate and for how long?

We study this problem using \textbf{Bayesian decision theory}, using \textbf{economic costs} of simulation and alternative selection.
We have $k$ alternatives. Alternative $x \in \{1, \ldots, k\}$ has true value, $U_x$.

There is also a standard option with known value $U_0$, e.g., $U_0 = 0$ is the value of “doing nothing.”

Sampling alternative $x$ gives a noisy observation of $U_x$,

$$y \sim \text{Normal}(U_x, \sigma_x^2),$$

where we suppose the measurement variance $\sigma_x^2$ is known.

To describe our belief about $U_1, \ldots, U_k$, we assume an independent normal prior distribution. Let $\vec{\Theta}_0$ be a vector containing the means and variances of the prior.

After observing a sequence of samples, we will have a posterior distribution that is also normal. Let $\vec{\Theta}_t$ be this posterior.
A fully sequential policy $\pi$ is a rule for adaptively choosing which alternative to sample at each point in time, and when to stop.

**Notation:**
- $T$ is the total number of samples.
- $I(T)$ is the alternative selected as the best.
- $c$ is the cost of one sample (can allow dependence on the alternative).

Sampling incurs a direct cost, but improves our eventual choice $I(T)$.

The value of a policy $\pi$ given the information in the posterior $\Theta$ is

$$V^\pi(\Theta) = E_{\pi} \left[ -cT + U_{I(T)} \mid \Theta_0 = \Theta \right].$$

**Goal:** find the policy with optimal value, $V^*(\Theta) = \sup_{\pi} V^\pi(\Theta)$. 
Previous Literature

This work builds on two related sections of the literature.

- **Economics of simulation:** [Chick and Gans, 2009] considers a discounted version of our problem. We extend this work by considering the undiscounted case, and by developing a **new and improved policy**.

- **Knowledge-gradient:** [Gupta and Miescke, 1996, Frazier et al., 2008] derive an allocation rule based on a single-step expected value of information calculation. [Frazier and Powell, 2008] extends this idea to stopping rules. We extend this work by considering **multi-step valuations of information**.
Consider the special case of comparing a **single** alternative against a known standard:

\[
V^*(\Theta) = \sup_{\pi} E_\pi \left[ -cT + \max\{U_0, \mu_T\} \mid \Theta_0 = \Theta \right]
\]

where \( \mu_T = E[U_1 \mid \Theta_T] \) is the posterior mean of the alternative.

- This is an optimal stopping problem.
- We relax this problem by allowing \( T \) to take real values, instead of just integers.
- Then the posterior mean \( \mu_T \) becomes a diffusion, and the optimal stopping problem becomes a **free-boundary problem**.
If we solve for the optimal stopping boundary for standard values \( c = 1, \sigma = 1, U_0 = 0 \), a simple algebraic transformation provides the optimal stopping boundary for any values \( c, \sigma, U_0 \).

Let \( \pm b(t) \) be the optimal boundary for the standard problem.

We use this approximation to \( b(t) \), with little loss in performance.

\[
b(t) \approx \begin{cases} 
.233s^2 & \text{if } s \leq 1 \\
.00537s^4 - .06906s^3 + .3167s^2 - .02326s & \text{if } 1 < s \leq 3 \\
.705s^{1/2}\ln(s) & \text{if } 3 < s \leq 40 \\
.642(s(2\ln(s))^{1.4} - \ln(32\pi))^{1/2} & \text{if } 40 < s,
\end{cases}
\]

where \( s = 1/t \).

This approximation is easy to compute, and does not require solving the free-boundary problem.
We now consider multiple alternatives, and derive or re-derive stopping and allocation rules using the idea of \textit{value of information}.

- In general, the optimal stopping rule is

\[ T = \inf \left\{ t \geq 0 : V^*(\vec{\Theta}_t) - \max_{x=0,...,k} \mu_{Tx} = 0 \right\}. \]

- \( \max_x \mu_{Tx} \) is the value obtained by taking no more samples.
- \( V^*(\vec{\Theta}_t) \) is the maximal value that can be extracted given \( \vec{\Theta}_t \).
- \( V^*(\vec{\Theta}_t) - \max_x \mu_{Tx} \geq 0 \) is the \textbf{net value of continuing to sample} in an optimal way.
- \( V^*(\vec{\Theta}_t) \) is hard to calculate for \( k > 1 \). We approximate it.
The optimal stopping rule is

\[ T = \inf \left\{ t \geq 0 : V^*(\Theta_t) - \max_{x=0,\ldots,k} \mu_{T_x} = 0 \right\}. \]

\( V^*(\Theta_t) \) is hard to compute, so we approximate it as the maximum of the value functions \( V^*_x(\Theta_t) \) for "single-alternative problems".

\[ V^*(\Theta_t) \approx \max_{x=0,\ldots,k} V^*_x(\Theta_t) \]

In the single-alternative problem for \( x \), we may only sample \( x \), and upon stopping we can select either \( x \) or the best of the rest. \( V^*_x(\Theta_t) \) can be computed using the approximation for the \( k = 1 \) problem.

We call this the **PDE stopping rule**, and it is easy to compute.
We compared PDE against several other stopping rules derived using approximations to the value of information.

- **PDE**: single alternative, adaptive sample size. (This talk, optimal for $k = 1$)

- **KG$_1$**: single alternative, single sample. [Frazier et al., 2008]

- **KG$_*$**: single alternative, deterministic sample size. [Frazier and Powell, 2010]

- **EOC$_{c,k}$**: multiple alternatives, deterministic sample size. [Chick and Inoue, 2001]
These approximations to the value of information also imply allocation rules.

- **PDE**: Sample the alternative whose posterior mean is furthest from the $k = 1$ stopping boundary.

- **KG$_1$**: Sample the alternative with the largest expected value of information (EVI). [Frazier et al., 2008]

- **KG$_*$**: Sample the alternative with the largest average EVI per sample (over deterministic rules). [Frazier and Powell, 2010]

- **Sequential LL (based on EOC)**: Sample the alternative to which the most samples are allocated by the allocation with the best net EVI. [Chick and Inoue, 2001]
## Numerical Results \((k > 1)\)

- Table shows expected loss \(E[cT + OC]\) for pairs of stopping and allocation rules. **Lower is better.**
- PDE stopping with KG\(^\ast\) allocation is the best policy.
- It is better than LL, EOC\(_{c,1}\), which was best in the large empirical study [Branke et al., 2007].

<table>
<thead>
<tr>
<th>Alloc,Stop</th>
<th>(k=3)</th>
<th>10</th>
<th>20</th>
<th>50</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>KG(_1),KG(_1)</td>
<td>3508 ± 12</td>
<td>7140 ± 18</td>
<td>8767 ± 19</td>
<td>10862 ± 67</td>
<td>12500 ± 72</td>
</tr>
<tr>
<td>KG(<em>{\ast}),KG(</em>{\ast})</td>
<td>674 ± 4</td>
<td>1445 ± 6</td>
<td>1761 ± 6</td>
<td>2245 ± 23</td>
<td>2666 ± 25</td>
</tr>
<tr>
<td>Equal,EOC(_{c,k})</td>
<td>433 ± 2</td>
<td>1040 ± 3</td>
<td>1815 ± 3</td>
<td>4220 ± 16</td>
<td>8425 ± 29</td>
</tr>
<tr>
<td>LL,EOC(_{c,k})</td>
<td>429 ± 2</td>
<td>821 ± 4</td>
<td>1095 ± 4</td>
<td>1577 ± 17</td>
<td>2168 ± 22</td>
</tr>
<tr>
<td>KG(<em>{\ast}),EOC(</em>{c,k})</td>
<td>424 ± 2</td>
<td>799 ± 3</td>
<td>1057 ± 4</td>
<td>1489 ± 16</td>
<td>2027 ± 19</td>
</tr>
<tr>
<td>KG(<em>1),EOC(</em>{c,k})</td>
<td>419 ± 2</td>
<td>728 ± 3</td>
<td>916 ± 3</td>
<td>1223 ± 11</td>
<td>1577 ± 11</td>
</tr>
<tr>
<td>KG(_1),PDE</td>
<td>348 ± 2</td>
<td>694 ± 3</td>
<td>875 ± 3</td>
<td>1158 ± 10</td>
<td>1516 ± 10</td>
</tr>
<tr>
<td>PDE,PDE</td>
<td>344 ± 2</td>
<td>700 ± 3</td>
<td>856 ± 3</td>
<td>1075 ± 11</td>
<td>1308 ± 12</td>
</tr>
<tr>
<td>KG(_{\ast}),PDE</td>
<td>327 ± 2</td>
<td>600 ± 2</td>
<td>722 ± 3</td>
<td>905 ± 8</td>
<td>1111 ± 9</td>
</tr>
</tbody>
</table>
Consider the effect of the stopping rule.

PDE is the best stopping rule, followed in order by $EOC_{c,k}$, $KG_*$, and $KG_1$. $KG_1$ performed badly because it underestimates the value of information.

<table>
<thead>
<tr>
<th></th>
<th>k=3</th>
<th>10</th>
<th>20</th>
<th>50</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>KG$_1$,KG$_1$</td>
<td>3508 ± 12</td>
<td>7140 ± 18</td>
<td>8767 ± 19</td>
<td>10862 ± 67</td>
<td>12500 ± 72</td>
</tr>
<tr>
<td>Equal,$EOC_{c,k}$</td>
<td>433 ± 2</td>
<td>1040 ± 3</td>
<td>1815 ± 3</td>
<td>4220 ± 16</td>
<td>8425 ± 29</td>
</tr>
<tr>
<td>KG$<em>*$,KG$</em>*$</td>
<td>674 ± 4</td>
<td>1445 ± 6</td>
<td>1761 ± 6</td>
<td>2245 ± 23</td>
<td>2666 ± 25</td>
</tr>
<tr>
<td>LL,$EOC_{c,k}$</td>
<td>429 ± 2</td>
<td>821 ± 4</td>
<td>1095 ± 4</td>
<td>1577 ± 17</td>
<td>2168 ± 22</td>
</tr>
<tr>
<td>KG$<em>*$,$EOC</em>{c,k}$</td>
<td>424 ± 2</td>
<td>799 ± 3</td>
<td>1057 ± 4</td>
<td>1489 ± 16</td>
<td>2027 ± 19</td>
</tr>
<tr>
<td>KG$<em>1$,$EOC</em>{c,k}$</td>
<td>419 ± 2</td>
<td>728 ± 3</td>
<td>916 ± 3</td>
<td>1223 ± 11</td>
<td>1577 ± 11</td>
</tr>
<tr>
<td>KG$_1$,PDE</td>
<td>348 ± 2</td>
<td>694 ± 3</td>
<td>875 ± 3</td>
<td>1158 ± 10</td>
<td>1516 ± 10</td>
</tr>
<tr>
<td>PDE,PDE</td>
<td>344 ± 2</td>
<td>700 ± 3</td>
<td>856 ± 3</td>
<td>1075 ± 11</td>
<td>1308 ± 12</td>
</tr>
<tr>
<td>KG$_*$,PDE</td>
<td>327 ± 2</td>
<td>600 ± 2</td>
<td>722 ± 3</td>
<td>905 ± 8</td>
<td>1111 ± 9</td>
</tr>
</tbody>
</table>
Stopping Boundary

Upper stopping boundary

Posterior mean, $y_t$

Effective number of replications, $t$

- From PDE or Quick Approx.
- KG(*)/EOC
- KG(*)/EOC Approx
- KG
- KG_{10}
- KG_{50}

Stopping boundaries with $k = 1$ alternative and parameter values $c = 1, \sigma = 10, m = 0$. 
Numerical Results ($k > 1$): Allocation Rule

- Consider the effect of the allocation rule.
- $KG_*$ and $KG_1$ are the best allocation rules, despite poor performance as stopping rules. They consistently underestimate the value of information, but this bias cancels when making allocation decisions.
- The PDE allocation rule also performs well.

<table>
<thead>
<tr>
<th>Allocation Rule</th>
<th>$k=3$</th>
<th>10</th>
<th>20</th>
<th>50</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>$KG_1$, $KG_1$</td>
<td>$3508 \pm 12$</td>
<td>$7140 \pm 18$</td>
<td>$8767 \pm 19$</td>
<td>$10862 \pm 67$</td>
<td>$12500 \pm 72$</td>
</tr>
<tr>
<td>$KG_<em>$,$KG_</em>$</td>
<td>$674 \pm 4$</td>
<td>$1445 \pm 6$</td>
<td>$1761 \pm 6$</td>
<td>$2245 \pm 23$</td>
<td>$2666 \pm 25$</td>
</tr>
<tr>
<td>Equal,$EOC_{c,k}$</td>
<td>$433 \pm 2$</td>
<td>$1040 \pm 3$</td>
<td>$1815 \pm 3$</td>
<td>$4220 \pm 16$</td>
<td>$8425 \pm 29$</td>
</tr>
<tr>
<td>LL,$EOC_{c,k}$</td>
<td>$429 \pm 2$</td>
<td>$821 \pm 4$</td>
<td>$1095 \pm 4$</td>
<td>$1577 \pm 17$</td>
<td>$2168 \pm 22$</td>
</tr>
<tr>
<td>$KG_*$,$EOC_{c,k}$</td>
<td>$424 \pm 2$</td>
<td>$799 \pm 3$</td>
<td>$1057 \pm 4$</td>
<td>$1489 \pm 16$</td>
<td>$2027 \pm 19$</td>
</tr>
<tr>
<td>$KG_1$,EOC$_{c,k}$</td>
<td>$419 \pm 2$</td>
<td>$728 \pm 3$</td>
<td>$916 \pm 3$</td>
<td>$1223 \pm 11$</td>
<td>$1577 \pm 11$</td>
</tr>
<tr>
<td>$KG_1$,PDE</td>
<td>$348 \pm 2$</td>
<td>$694 \pm 3$</td>
<td>$875 \pm 3$</td>
<td>$1158 \pm 10$</td>
<td>$1516 \pm 10$</td>
</tr>
<tr>
<td>PDE,PDE</td>
<td>$344 \pm 2$</td>
<td>$700 \pm 3$</td>
<td>$856 \pm 3$</td>
<td>$1075 \pm 11$</td>
<td>$1308 \pm 12$</td>
</tr>
<tr>
<td>$KG_*$,PDE</td>
<td>$327 \pm 2$</td>
<td>$600 \pm 2$</td>
<td>$722 \pm 3$</td>
<td>$905 \pm 8$</td>
<td>$1111 \pm 9$</td>
</tr>
</tbody>
</table>
This approach balances the cost of sampling with the rewards of having information – financial criteria may be more appropriate than statistical criteria in most business decisions.

We can solve the $k = 1$ case “exactly” with a PDE, and that solution supports understanding of the $k > 1$ problem.

The resulting PDE stopping rule is empirically better than those from prior experiments.

This approach may have application in more complex simulation optimization problems (e.g. unknown variances, CRN, correlated beliefs, metamodels), not just independent variance-known ranking and selection.
If you are interested in these topics, please consider submitting a paper to an upcoming **special issue of IIE Transactions** devoted to simulation optimization and its applications.

**Due Date for Submission:** June 2011

**Special Issue Editors:** Loo Hay Lee; Ek Peng Chew; Samuel Qing-Shan Jia; Peter Frazier; Chun-Hung Chen


Expected loss of stopping rules for $k = 1$, $c = 1$, $\mu_0 = 0$, $t_0 = 100$, and $\sigma = 10^5$ calculated using Monte Carlo simulation with $10^6$ samples. OC = Opportunity Cost = $\max_i U_i - U_{I(T)}$

<table>
<thead>
<tr>
<th>Stopping Rule</th>
<th>$E[cT]$</th>
<th>$E[OC]$</th>
<th>$E[cT + OC]$</th>
<th>% sub-optimality</th>
</tr>
</thead>
<tbody>
<tr>
<td>PDE</td>
<td>321.85 ± 0.25</td>
<td>263 ± 1</td>
<td>585 ± 1</td>
<td>—</td>
</tr>
<tr>
<td>EOC_{c,k}</td>
<td>142.53 ± 0.11</td>
<td>612 ± 2</td>
<td>755 ± 2</td>
<td>4.99%</td>
</tr>
<tr>
<td>KG_{*}</td>
<td>136.51 ± 0.11</td>
<td>634 ± 2</td>
<td>770 ± 2</td>
<td>5.45%</td>
</tr>
<tr>
<td>KG_{1}</td>
<td>10.53 ± 0.01</td>
<td>2505 ± 5</td>
<td>2515 ± 5</td>
<td>56.69%</td>
</tr>
</tbody>
</table>

Better approximations to the value of information give better performance.