The Conjunction of the Knowledge Gradient and the Economic Approach to Simulation Selection

Stephen E. Chick\textsuperscript{1}
Peter I. Frazier\textsuperscript{2}

\textsuperscript{1}Technology & Operations Management, INSEAD
\textsuperscript{2}Operations Research & Information Engineering, Cornell University

Tuesday December 15, 2009
Winter Simulation Conference, Austin
We have a discrete event simulator with which we can estimate the consequences of a future real-world decision. For example:

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

We want to find a real-world decision that will work well, according to the simulator.

Our simulator requires significant time to accurately characterize a decision, and we do not have enough time to do so for each one.

Question: which real-world decisions should we simulate and for how long?
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 $\tilde{\Theta}_0$ 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 $\tilde{\Theta}_t$ be this posterior.
Bayesian Posterior Probability Distribution

Posterior Mean vs Time (t)

Prior Mean vs True value U

Chick, Frazier (INSEAD, Cornell)
Cost and Benefit of Sampling

- A fully sequential policy $\pi$ is a rule for choosing at each time $t$ to either:
  - Sample an alternative $i(t)$ of our choosing, and pay a cost $c_{i(t)}$; We call the rule for choosing $i(t)$ the **allocation rule**.
  - Or, stop sampling ($T = t$), select an alternative $I(T)$ for implementation, and receive a reward $U_{I(T)}$. We call the rule for choosing $T$ the **stopping rule**.

- 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

\[
V^\pi(\Theta) = E_\pi \left[ \sum_{t=0}^{T-1} -c_{i(t)} + U_{I(T)} \mid \Theta_0 = \Theta \right].
\]

- Our goal is to find a policy with optimal value $V^*(\Theta) = \sup_\pi V^\pi(\Theta)$. 
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 novel stopping and allocation rules.

- **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. Then the problem becomes,

\[
V^*(\tilde{\Theta}) = \sup_{\pi} E_{\pi} \left[ -cT + \max\{U_0, \mu_T\} | \tilde{\Theta}_0 = \tilde{\Theta} \right]
\]

where \( \mu_T = E[U_1 | \tilde{\Theta}_T] \) is the posterior mean of the alternative.

- This is an optimal stopping problem.
- We relax our single-alternative 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.
Optimal Stopping Boundary

Stopping Time (T)

Expected value upon stopping

Upper Stopping Boundary

Lower Stopping Boundary

Time (t)

Posterior Mean
Ease of Use

- If we solve the free-boundary problem **once** for standard values $c = 1$, $\sigma = 1$, $U_0 = 0$, we may easily transform to solve for **any** values $c > 0$, $\sigma > 0$, $U_0$.
- For the standard problem, let $\pm b(1/t)$ be the optimal boundary.
- For general problems, the optimal boundary is

$$
m \pm \beta^{-1} b(1/(\gamma t)),
$$

where $\beta = c^{-1/3} \sigma^{-2/3}$ and $\gamma = c^{2/3}/\sigma^{-2/3}$.

- One can use the following approximation to the standard boundary $b$ with little loss in performance.

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

- **Practitioners need not solve the free-boundary problem.**
We now consider multiple alternatives, and derive or re-derive stopping and allocation rules using the idea of **value of information**.

- In general, the optimal stopping rule is

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

- \( \max_x \mu_{Tx} \) is the value obtained by taking no more samples.
- \( V^*(\tilde{\Theta}_t) \) is the maximal value that can be extracted given \( \tilde{\Theta}_t \).
- \( V^*(\tilde{\Theta}_t) - \max_x \mu_{Tx} \geq 0 \) is the **net value of continuing to sample** in an optimal way.

- \( V^*(\tilde{\Theta}_t) \) is hard to calculate for \( k > 1 \). Instead, we approximate the value of continuing to sample.
Stopping Rules

- Each of the following stopping rules approximate the net value of sampling with best achievable within a simple class of policies. Each stops when its approximation is 0.
  - KG$_1$: single alternative, single sample.
  - KG$_*$: single alternative, deterministic sample size.
  - PDE: single alternative, adaptive sample size.
  - EOC$_{c,k}$: multiple alternatives, deterministic sample size.
  - optimal: multiple alternatives, adaptive sample size and allocation. (the optimal policy is computationally intractable)

- KG$_*$, PDE, and EOC$_{c,k}$ require further approximations to find the best policy within their class.

- Of these rules, KG$_*$ and PDE are new.
Stopping Boundary

Upper stopping boundary

- From PDE or Quick Approx.
- KG(*)/EOC
- KG(*)/EOC Approx
- KG
- KG_1
- KG_10
- KG_50

Effective number of replications, \( t \)

Posterior mean, \( y_t/\tau \)
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. \(\text{OC} = \text{Opportunity Cost} = \max_i U_i - U_i(T)\)

<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$_b$</td>
<td>$321.85 \pm 0.25$</td>
<td>$263 \pm 1$</td>
<td>$585 \pm 1$</td>
<td>—</td>
</tr>
<tr>
<td>EOC$_{c,k}$</td>
<td>$142.53 \pm 0.11$</td>
<td>$612 \pm 2$</td>
<td>$755 \pm 2$</td>
<td>$4.99%$</td>
</tr>
<tr>
<td>KG$_*$</td>
<td>$136.51 \pm 0.11$</td>
<td>$634 \pm 2$</td>
<td>$770 \pm 2$</td>
<td>$5.45%$</td>
</tr>
<tr>
<td>KG$_1$</td>
<td>$10.53 \pm 0.01$</td>
<td>$2505 \pm 5$</td>
<td>$2515 \pm 5$</td>
<td>$56.69%$</td>
</tr>
</tbody>
</table>

Better approximations to the value of information result directly in better expected performance.
These approximations to the value of information also imply allocation rules.

- **KG$_1$**: Sample the alternative with the largest expected value of information (EVI).
- **KG$_*$**: Sample the alternative with the largest average EVI per sample (over deterministic rules).
- **PDE**: Sample the alternative whose posterior mean is furthest from the $k=1$ stopping boundary.
- **Sequential LL (based on EOC)**: Sample the alternative to which the most samples are allocated by the allocation with the best net EVI.
Numerical Results ($k > 1$)

- Table shows expected loss $E[cT + OC]$ for pairs of stopping and allocation rules. Rows are sorted by loss at $k = 100$. The best (lowest loss) are at the bottom.
- PDE stopping with KG∗ allocation is the best policy.
- It is better than LL, EOCc,1, which was best in the large empirical study [Branke et al., 2007].

<table>
<thead>
<tr>
<th>Stop, Alloc</th>
<th>k=3</th>
<th>10</th>
<th>20</th>
<th>50</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>KG₁,KG₁</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∗,KG∗</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,EOCc,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,EOCc,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∗,EOCc,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>KG₁,EOCc,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>KG₁,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>

Chick, Frazier (INSEAD,Cornell)
Consider the effect of the stopping rule.

PDE is the best stopping rule, followed in order by $\text{EOC}_{c,k}$, $\text{KG*}$, and $\text{KG}_1$. $\text{KG}_1$ performed very badly because it dramatically 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>$\text{KG}_1,\text{KG}_1$</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>$\text{Equal,EOC}_{c,k}$</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>$\text{KG*},\text{KG*}$</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>$\text{LL,EOC}_{c,k}$</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>$\text{KG*},\text{EOC}_{c,k}$</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>$\text{KG}<em>1,\text{EOC}</em>{c,k}$</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>$\text{KG}_1,\text{PDE}$</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>$\text{PDE,PDE}$</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>$\text{KG*},\text{PDE}$</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>
Consider the effect of the allocation rule.

KG\_\ast\ast\ast 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 ± 12</td>
<td>7140 ± 18</td>
<td>8767 ± 19</td>
<td>10862 ± 67</td>
<td>12500 ± 72</td>
</tr>
<tr>
<td>KG_\ast\ast,KG_\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_\ast\ast,EOC_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_1,EOC_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\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>
Conclusion

- 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” when the variance is known with a PDE, and that solution supports understanding of the $k > 1$ problem.
- We extended both the KG and economics of simulation approaches with new stopping rules and allocation rules that are empirically better than those from prior experiments.
- We believe that 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.


Any questions?