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

Stephen E. Chick[1]
Peter I. Frazier[2]

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

Tuesday December 15, 2009
Winter Simulation Conference, Austin

Ranking and Selection for Discrete Event Simulation

- 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?

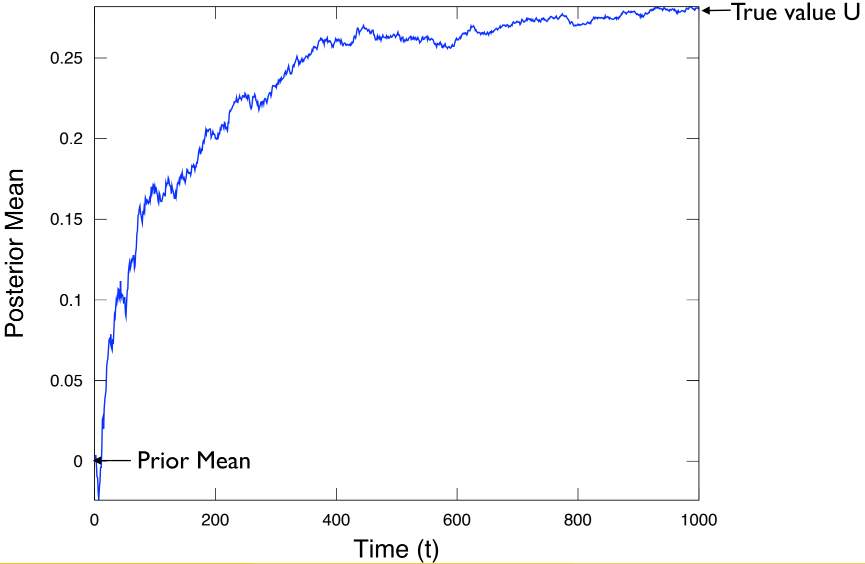
Ranking & Selection (R&S)

- 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,

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

  where we suppose the measurement variance σ_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.

Bayesian Posterior Probability Distribution



Cost and Benefit of Sampling

- A fully sequential policy π 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 π given the information in the posterior Θ is

$$V^{\pi}(\Theta) = \mathrm{E}_{\pi}\left[\sum_{t=0}^{T-1} -c_{i(t)} \ + \ U_{I(T)} \ | \vec{\Theta}_0 = \vec{\Theta}\right].$$

- Our goal is to find a policy with optimal value $V^{*}(\vec{\Theta}) = \sup_{\pi} V^{\pi}(\vec{\Theta})$.

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 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.

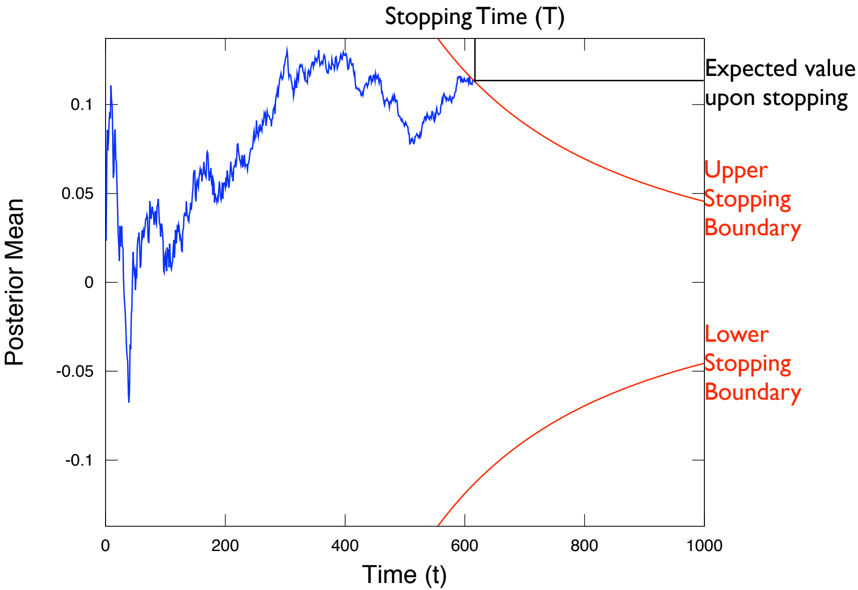
Special Case: $k = 1$

- Consider the special case of comparing a **single** alternative against a known standard. Then the problem becomes,

$$V^*(\vec{\Theta}) \;=\; \sup_{\pi} \mathrm{E}_{\pi}\left[-cT \;+\; \max\{U_0, \mu_T\} \middle| \vec{\Theta}_0 = \vec{\Theta}\right]$$

  where $\mu_T = \mathrm{E}[U_1 \mid \vec{\Theta}_T]$ 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 μ_T becomes a diffusion, and the optimal stopping problem becomes a free-boundary problem.

Optimal Stopping Boundary



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.**

Multiple Alternatives

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

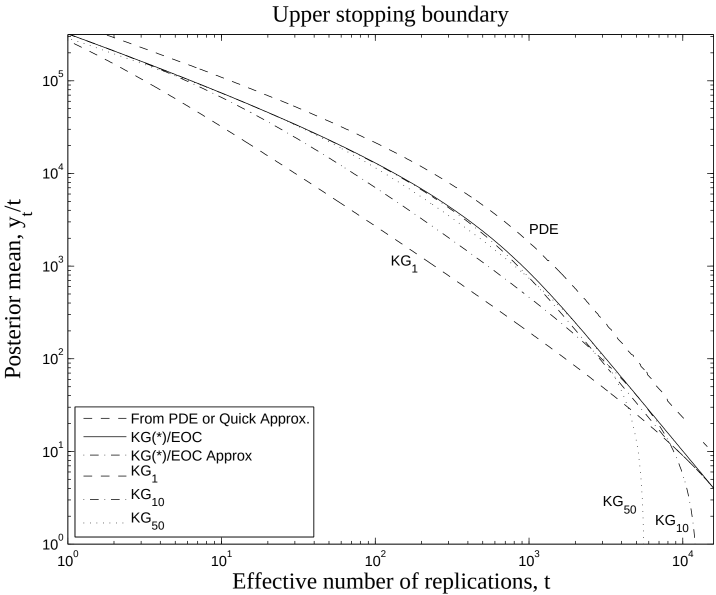
$$T = \inf\left\{ t \geq 0 : V^*(\vec{\Theta}_t) - \max_{x=0,\ldots,k} \mu_{Tx} = 0 \right\}.$$

  - $\max_x \mu_{Tx}$ is the value obtained by taking no more samples.
  - $V^*(\vec{\Theta}_t)$ is the maximal value that can be extracted given $\vec{\Theta}_t$.
  - $V^*(\vec{\Theta}_t) - \max_x \mu_{Tx} \geq 0$ is the **net value of continuing to sample** in an optimal way.
- $V^*(\vec{\Theta}_t)$ 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.
  - $\text{EOC}_{c,k}$: **multiple** alternatives, **deterministic** sample size.
  - optimal: **multiple** alternatives, **adaptive** sample size and allocation. (the optimal policy is compuationally intractable)
- KG_*, PDE, and $\text{EOC}_{c,k}$ require further approximations to find the best policy within their class.
- Of these rules, KG_* and PDE are new.

Stopping Boundary



Numerical Results ($k = 1$)

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 = $\max_i U_i - U_{I(T)}$

Stopping Rule	E[cT]	E[OC]	E[cT+OC]	% sub-optimality
PDE_b	321.85 ± 0.25	263 ± 1	585 ± 1	—
$\mathrm{EOC}_{c,k}$	142.53 ± 0.11	612 ± 2	755 ± 2	4.99%
KG_*	136.51 ± 0.11	634 ± 2	770 ± 2	5.45%
KG_1	10.53 ± 0.01	2505 ± 5	2515 ± 5	56.69%

Better approximations to the value of information result directly in better expected performance.

Allocation Rules

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 $LL, EOC_{c,1}$, which was best in the large empirical study [Branke et al., 2007].

Stop,Alloc	k=3	10	20	50	100
KG_1,KG_1	3508 ± 12	7140 ± 18	8767 ± 19	10862 ± 67	12500 ± 72
KG_*,KG_*	674 ± 4	1445 ± 6	1761 ± 6	2245 ± 23	2666 ± 25
Equal,$EOC_{c,k}$	433 ± 2	1040 ± 3	1815 ± 3	4220 ± 16	8425 ± 29
LL,$EOC_{c,k}$	429 ± 2	821 ± 4	1095 ± 4	1577 ± 17	2168 ± 22
KG_*,$EOC_{c,k}$	424 ± 2	799 ± 3	1057 ± 4	1489 ± 16	2027 ± 19
KG_1,$EOC_{c,k}$	419 ± 2	728 ± 3	916 ± 3	1223 ± 11	1577 ± 11
KG_1, PDE	348 ± 2	694 ± 3	875 ± 3	1158 ± 10	1516 ± 10
PDE, PDE	344 ± 2	700 ± 3	856 ± 3	1075 ± 11	1308 ± 12
KG_*, PDE	327 ± 2	600 ± 2	722 ± 3	905 ± 8	1111 ± 9

Numerical Results ($k > 1$): Stopping Rule

- 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 very badly because it dramatically underestimates the value of information.

	k=3	10	20	50	100
KG_1,KG_1	3508 ± 12	7140 ± 18	8767 ± 19	10862 ± 67	12500 ± 72
Equal,$EOC_{c,k}$	433±2	1040±3	1815±3	4220 ± 16	8425 ± 29
KG_*,KG_*	674 ± 4	1445 ± 6	1761 ± 6	2245 ± 23	2666 ± 25
LL,$EOC_{c,k}$	429 ± 2	821 ± 4	1095 ± 4	1577 ± 17	2168 ± 22
KG_*,$EOC_{c,k}$	424 ± 2	799 ± 3	1057 ± 4	1489 ± 16	2027 ± 19
KG_1,$EOC_{c,k}$	419 ± 2	728 ± 3	916 ± 3	1223 ± 11	1577 ± 11
KG_1,PDE	348 ± 2	694 ± 3	875 ± 3	1158 ± 10	1516 ± 10
PDE,PDE	344 ± 2	700 ± 3	856 ± 3	1075 ± 11	1308 ± 12
KG_*,PDE	327 ± 2	600 ± 2	722 ± 3	905 ± 8	1111 ± 9

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.

	k=3	10	20	50	100
KG_1, KG_1	3508 ± 12	7140 ± 18	8767 ± 19	10862 ± 67	12500 ± 72
KG_*,KG_*	674 ± 4	1445 ± 6	1761 ± 6	2245 ± 23	2666 ± 25
Equal,$EOC_{c,k}$	433±2	1040±3	1815±3	4220 ± 16	8425 ± 29
LL,$EOC_{c,k}$	429 ± 2	821 ± 4	1095 ± 4	1577 ± 17	2168 ± 22
KG_*,$EOC_{c,k}$	424 ± 2	799 ± 3	1057 ± 4	1489 ± 16	2027 ± 19
KG_1,$EOC_{c,k}$	419 ± 2	728 ± 3	916 ± 3	1223 ± 11	1577 ± 11
KG_1,PDE	348 ± 2	694 ± 3	875 ± 3	1158 ± 10	1516 ± 10
PDE,PDE	344 ± 2	700 ± 3	856 ± 3	1075 ± 11	1308 ± 12
KG_*,PDE	327 ± 2	600 ± 2	722 ± 3	905 ± 8	1111 ± 9

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.

Branke, J., Chick, S., and Schmidt, C. (2007).
Selecting a selection procedure.
Management Sci., 53(12):1916–1932.

Chick, S. and Gans, N. (2009).
Economic analysis of simulation selection problems.
Management Sci., to appear.

Frazier, P. and Powell, W. (2008).
The knowledge-gradient stopping rule for ranking and selection.
Winter Simul. Conf. Proc., 2008.

Frazier, P., Powell, W. B., and Dayanik, S. (2008).
A knowledge gradient policy for sequential information collection.
SIAM J. on Control and Optimization, 47(5):2410–2439.

Gupta, S. and Miescke, K. (1996).
Bayesian look ahead one-stage sampling allocations for selection of the best population.
Journal of statistical planning and inference, 54(2):229–244.

Thank You

Any questions?