Knowledge-gradient methods for ranking and selection

Peter Frazier, Warren Powell, Savas Dayanik

Department of Operations Research and Financial Engineering
Princeton University

Wednesday May 7, 2008
Graduate Student Research Conference
Ranking and Selection

- We have a collection of \( M \) alternatives, e.g. drug compounds.
- We have time for \( N \) experiments, to be performed sequentially.
- In each experiment, we select one alternative and obtain a noisy measurement of its value.
- Afterward, we select one alternative and receive its value as reward.

Begin with $n=0$, fixed $Y$, and $\mathcal{N}(\mu^0, \Sigma^0)$ belief on $Y$.

Choose alternative $x^n$ to measure.

Observe measurement $\hat{y}^{n+1} \sim \mathcal{N}(\hat{Y}_{x^n}, (\sigma^e)^2)$.

Receive reward $Y_{x^N}$.

Choose alternative $x^N$ to implement.

Increment $n$.

If $n < N$

If $n = N$
Model

- We have $M$ alternatives with a corresponding vector of unknown values $Y := (Y_1, \ldots, Y_M)$.
- At time 0, we have a $\mathcal{N}(\mu^0, \Sigma^0)$-distributed belief on $Y$.
- We are allowed $N$ samples $(x^0, x^1, \ldots, x^{N-1})$ where $x^n \in \{1, \ldots, M\}$ selects from which alternative to sample at time $n$.
- Associated with sampling decision $x^n$ is an observation $\hat{y}^{n+1}$. Conditioned on $Y$ and $x^n$, $\hat{y}^{n+1} \sim \mathcal{N}(Y_{x^n}, (\sigma^2)^2)$ and is independent of all other $x$ and $\hat{y}$. 
• Define $\mathcal{F}_n := \sigma \{ x^0, \hat{y}^1, \ldots, x^{n-1}, \hat{y}^n \}$ and require $x^n \in \mathcal{F}_n$.

• Define $\mu^n := \mathbb{E}_n Y$, $\Sigma^n := \text{Var}_n Y_i$.

• After the $N$ samples are observed, we choose an alternative $x^N$ and receive its underlying value $Y_{x^N}$ as reward. Note that $x^N \in \mathcal{F}_N$.

• Our objective is

$$\sup_{x^0, \ldots, x^{N-1}, x^N} \mathbb{E} [ Y_{x^N} ] = \sup_{x^0, \ldots, x^{N-1}} \mathbb{E} [ \max_i \mu_i^N ].$$
Example
Example
Example
Example
Example
Example
Example
Example
Example
Example
Example
Transition Function

The time $n$ posterior belief on $Y$ is $\mathcal{N}(\mu^n, \Sigma^n)$, where $\mu^n$ and $\Sigma^n$ can be computed recursively by

$$
\mu^{n+1} = \Sigma^{n+1} ( (\Sigma^n)^{-1} \mu^n + (\sigma^\varepsilon)^{-2} \hat{y}^{n+1} e_{x^n} ),
$$

$$
\Sigma^{n+1} = ( (\Sigma^n)^{-1} + (\sigma^\varepsilon)^{-2} e_{x^n} e_{x^n}^T )^{-1}.
$$

Define $\tilde{\sigma}(\Sigma, x) := \sqrt{\Sigma_{xx} + (\sigma^\varepsilon)^2} \left[ (\Sigma)^{-1} + (\sigma^\varepsilon)^{-2} e_x e_x^T \right]^{-1} e_x$. Then

$$
\mu^{n+1} = \mu^n + \tilde{\sigma}(\Sigma^n, x^n) Z
$$

for some standard normal random variable $Z$ independent of $\mathcal{F}_n$. 
Optimal Policy

Given a very large amount of computational time, we could compute the optimal policy through dynamic programming. Our state at time $n$ is $S^n := (\mu^n, \Sigma^n)$. Define the value function for any policy $\pi = (x^0, \ldots, x^{N-1})$ as

$$V^{\pi, n}(s) := \mathbb{E}^\pi_n \left[ \max_{i} \mu_i^n | S^n = s \right].$$

Define the value function as

$$V^n(s) := \sup_{\pi} V^{\pi, n}(s).$$

$V^n$ satisfies Bellman’s recursion,

$$V^N(S^n) = \max_i \mu_i^N, \quad V^n(S^n) = \max_{x^n} \mathbb{E}_n \left[ V^{n+1}(S^{n+1}) \right].$$
The **knowledge-gradient policy** is defined to be the policy that chooses $x^n$ to maximize

$$\mathbb{I}E_n \left[ (\max_x \mu^{n+1}_x) - \max_x \mu^n_x \right].$$
The knowledge-gradient policy can be computed as

\[ \arg \max_{x^n} \mathbb{E}_n \max_x \mu_{x}^{n+1} = \arg \max_{x^n} \mathbb{E}_n \max_x \mu_x^n + e_x^T \tilde{\sigma}(\Sigma^n, x^n)Z \]

\[ = \arg \max_{x^n} h(\mu^n, \tilde{\sigma}(\Sigma^n, x^n)) \]

where \( h : \mathbb{R}^M \times \mathbb{R}^M \rightarrow \mathbb{R} \) is defined by \( h(a, b) = \mathbb{E} \max_x a_x + b_x Z \).
Algorithm for Computing $h(a, b) = \mathbb{E}\max_x a_x + b_x Z$

1. Order the index set $\{1, \ldots, M\}$ so that $b_x \leq b_{x+1}$.
2. If $b_x = b_{x+1}$ for any $x$ then remove the dimension $x$ or $x+1$ with smaller $a$ component.
3. Let $c_x = \frac{a_x - a_{x+1}}{b_{x+1} - b_x}$. This is the point where the line $a_x + b_x z$ intersects $a_{x+1} + b_{x+1} z$. If $c_x \leq c_{x-1}$ then remove dimension $x$. Let $c_0 = -\infty$, $c_M = \infty$.
4. Since $\max_x a_x + b_x Z = \sum_x (a_x + b_x Z) 1_{[c_{x-1}, c_x)}(Z)$,

$$h(a, b) = \sum_x a_x (\Phi(c_x) - \Phi(c_{x-1})) + b_x (\varphi(c_x) - \varphi(c_{x-1}))$$
General Optimality Results

The knowledge-gradient policy is optimal when either:

- $N = 1$;
- $N \to \infty$.

The policy’s suboptimality is bounded in the remaining cases by

$$V^n(S^n) - V^{n,KG}(S^n) \leq \frac{1}{\sqrt{2\pi}} \max_{x^n,\ldots,x^{N-2}} \sum_{k=n+1}^{N-1} ||\tilde{\sigma}(\Sigma^k, \cdot)||$$

where $||u|| := \max_i u_i - \min_j u_j$ and $||\tilde{\sigma}(\Sigma, \cdot)|| := \max_x ||\tilde{\sigma}(\Sigma, x)||$. 
If \( \Sigma^0 \) is diagonal, the knowledge-gradient computation simplifies to

\[
\arg \max_x \tilde{\sigma}_x^n f(\zeta_x^n),
\]

where

\[
\tilde{\sigma}_x^n := (\Sigma_{xx}^n)^2 / \sqrt{(\Sigma_{xx}^n)^2 + (\sigma^\epsilon)^2}
\]

\[
\zeta_x^n := -|\mu_x^n - \max_{x' \neq x} \mu_{x'}^n| / \tilde{\sigma}_x^n
\]

\[
f(z) := z\Phi(z) + \varphi(z),
\]

\( \Phi \) is the normal cdf, and \( \varphi \) is the normal pdf. This policy was first proposed in Gupta&Miescke 1996, and called the R1,\ldots,R1 policy. We call it the “independent KG” policy.
Suppose that $\Sigma^0$ is diagonal. Then:

1. If there are exactly 2 alternatives ($M=2$), the knowledge-gradient policy is optimal. In this case, the optimal policy reduces to

$$x^n = \arg \max_x \Sigma_{xx}^n.$$ 

2. If there is no measurement noise, and alternatives may be reordered so that

$$\mu_1^0 \geq \mu_2^0 \geq \ldots \geq \mu_M^0$$

$$\Sigma_{11}^0 \geq \Sigma_{22}^0 \geq \ldots \geq \Sigma_{MM}^0,$$

then the knowledge-gradient policy is optimal.
Numerical Results

Numerical results with $\Sigma^0$ diagonal (independent KG).

![Graphs showing comparisons of Value(KG) – Value(IE) and Value(KG) – Value(OCBA) with Value(KG) – Value(LL(S)).]
Correlated KG Applications

- Maximizing continuous functions on $\mathbb{R}^d$, where the correlation in our belief is proportional to distance:
  - exploring for oil; localizing the source of a radiation leak
  - calibrating stochastic models, e.g., for pricing financial instruments
  - maximizing the efficiency of a chemical process, e.g., etching of silicon wafers.
- Alternatives characterized by common discrete features:
  - finding the shortest path through a network – the features are the path-links
  - choosing the best R&D portfolio – the features are the technologies funded in the portfolio
- Alternatives related by other metrics:
  - finding drug compounds to treat disease – drug performance is correlated with chemical similarity
  - finding the best partition-based model to price credit derivatives – model performance is correlated with partition similarity.
For the normal ranking and selection problem illustrated here, the knowledge-gradient policy is easier to compute than the optimal policy, performs well in numerical experiments, and is provably optimal in certain special cases. It is also flexible enough to be extended to a much broader class of sequential information collection problems.
Future Work

• Improved computational efficiency
  • For continuous functions.
  • For functions of binary attribute vectors.
  • Methods for non-conjugate priors (empirical bayes, variational methods, importance sampling).

• Extensions
  • Calibration objective function.
  • Risk aversion.
  • Unknown sampling variance.