Sequential Bayesian Ranking and Selection with Knowledge Gradients

Peter Frazier
Advisor: Warren Powell

Department of Operations Research and Financial Engineering
Princeton University

Thursday May 3, 2007
ORFE General Exam
Example, Time 0
Example, Time 1

[Graph showing scattered data points with error bars labeled as prior and measurement.]

x=1 x=2 x=3 x=4 x=5
Example, Time 2
Example, Time 5
Example, Final Time 10
General Model

- $\theta$ is a random variable representing “the true state of the world.”
- We are allowed $N$ measurements $(x^0, x^1, \ldots, x^{N-1})$ where $x^n \in \mathcal{F}_n$ (defined below) is a random variable takes values in $\mathcal{X}$.
- We observe $(\hat{y}^1, \hat{y}^2, \ldots, \hat{y}^N)$. Conditioned on $\theta$ and $x^n$, $\hat{y}^{n+1}$ is independent and distributed according to the cdf $F_{x^n}(\cdot; \theta)$.
- $\mathcal{F}_n := \sigma \{ x^0, \hat{y}^1, \ldots, x^{n-1}, \hat{y}^n \}$.
- After all measurements, we choose an implemention decision, $i \in \mathcal{F}_N$, which is a random variable taking values in $\mathcal{I}$. We receive reward $C_i(\theta)$.
- Our objective is $\sup_{x^0, \ldots, x^{N-1}, i} \mathbb{E} C_i(\theta)$. 
Our state at time $n$ is the conditional distribution of $\theta$ given $\mathcal{F}_n$. Consider our “utility of information”,

$$U(S^n) := \sup_{i \in \mathcal{I}} \mathbb{E}_n C_i(\theta),$$

and consider the random change in utility due to a measurement at time $n$

$$\Delta U^{n+1} := U(S^{n+1}) - U(S^n)$$
Knowledge Gradient Definition

The knowledge-gradient policy chooses the measurement that maximizes this expected increase in utility,

\[ X^{KG}(S^n) \in \arg \max_{X^n} \mathbb{E}_n \Delta U^{n+1} = \arg \max_{X^n} \mathbb{E}_n \left[ \sup_{i \in \mathcal{I}} \mathbb{E}_{n+1} C_i(\theta) \right]. \]

Additional assumptions are required to ensure existence and uniqueness.
Knowledge Gradient Motivation

- Optimal by construction for $N = 1$.
- Promising results for special cases.
- Asymptotically optimal as $N \to \infty$ (pending verification).
Independent Normal Model

- Let $M$ be a finite integer.
- Let $Y = (Y_1, \ldots, Y_M)$ be a multivariate normal random variable with independent components.
- Define $\mu^n_x := \mathbb{E}_n Y_x$, $(\sigma_x^n)^2 := \text{Var}_n Y_x$.
- Let $(\varepsilon^1, \ldots, \varepsilon^N)$ be independent $N(0, (\sigma^\varepsilon)^2)$.
- $\hat{Y}^{n+1} = Y_x^n + \varepsilon^{n+1}$
- Objective is $\mathbb{E}[Y_i] = \mathbb{E}\left[\max_x \mu_x^N\right]$.
Independent Normal Transition Function

At time $n$ measure alternative $x^n$. We update our estimate of $Y_{x^n}$ based on the measurement $\hat{y}^{n+1}$. For $x = x^n$,

$$
\mu_{x}^{n+1} = \frac{(\sigma_x^n)^{-2} \mu_x^n + (\sigma^e)^{-2} \hat{y}^{n+1}}{(\sigma_x^n)^{-2} + (\sigma^e)^{-2}}
$$

$$(\sigma_{x}^{n+1})^{-2} = (\sigma_x^n)^{-2} + (\sigma^e)^{-2}$$

Estimates of other $Y_x$ do not change.

$$
\mu_{x}^{n+1} = \mu_x^n \text{ for } x \neq x^n \\
\sigma_{x}^{n+1} = \sigma_x^n \text{ for } x \neq x^n
$$

Conditioned on $\mathcal{F}_n$ and $x = x^n$, $\mu_{x}^{n+1} \sim N(\mu_x^n, (\tilde{\sigma}_x^n)^2)$, where

$$
(\sigma_x^{n+1})^2 = (\sigma_x^n)^2 - (\tilde{\sigma}_x^{n+1})^2
$$
Independent Normal Knowledge-Gradient Policy

The knowledge-gradient policy for the independent normal model is

\[
\mathbb{E}_n [\Delta U^{n+1}] := \mathbb{E}_n [U(\mu^{n+1}, \sigma^{n+1}) - U(\mu^n, \sigma^n)]
\]

\[
= \mathbb{E}_n \left[ \left( \max_x \mu_x^{n+1} \right) - \max_x \mu^n \right]
\]

\[
= \mathbb{E}_n \left[ \left( \mu_x^{n+1} \vee \max_{x \neq x} \mu_x^n \right) - \max_x \mu^n \right],
\]

which is the expectation of the maximum of a normal random variable and a constant.
Independent Normal Knowledge-Gradient Policy

The computation becomes

$$\arg \max_x \tilde{\sigma}_x^n f(\zeta_x^n)$$

where

$$\tilde{\sigma}_x^n := (\sigma_x^n)^2 / \sqrt{(\sigma_x^n)^2 + (\sigma^\epsilon)^2}$$

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

$$f(z) := z \Phi(z) + \varphi(z),$$

$\Phi$ is the normal cdf, and $\varphi$ is the normal pdf.
Knowledge Gradient n=2

Prior

$Y_{x}$

$y_{hat}$
Knowledge Gradient $n=4$

Prior

$Y_x$

$Y_{\hat{x}}$

$\hat{y}$
Knowledge Gradient $n=5$

The graph shows the Knowledge Gradient for $n=5$ with the following parameters:

- Prior
- $Y_x$
- $yhat$

The x-axis represents different indices, and the y-axis represents values ranging from -2 to 2.

The data points are illustrated with red and green markers, indicating different measurements or conditions.
KnowledgeGradient n=7

prior
Y_x
yhat

22 / 67
Knowledge Gradient n=8

- Prior
- Y_x
- yhat

KnowledgeGradient n=8
Knowledge Gradient n=10

prior
Y_x
yhat

Graph showing the Knowledge Gradient with n=10 prior and Y_x and yhat data points.
KnowledgeGradient n=11

prior
Y_x
yhat
KnowledgeGradient $n=14$

prior

$Y_{X}$

$yhat$
KnowledgeGradient n=15

prior  Y_x  yhat
KnowledgeGradient n=18

prior
Y_x
yhat
Knowledge Gradient $n=20$

- Prior
- $Y_{xy}$
- $y_{hat}$
Independent Normal Optimality Results

- The knowledge-gradient policy is optimal when $N = 1$.
- The knowledge-gradient policy is asymptotically optimal as $N \to \infty$.
- For other $N$, the knowledge-gradient policy’s suboptimality is bounded by

$$V^{KG,n}(S^n) \geq V^n(S^n) - \frac{N - n - 1}{\sqrt{2\pi}} \max_x \tilde{\sigma}_x^n,$$

where $V^{KG,n}$ gives the value of the knowledge-gradient policy and $V^n$ the value of the optimal policy, both with $N - n$ measurements remaining.
Independent Normal Optimality Results

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

\[ \hat{x}^n = \arg \max_x \sigma_x^n. \]
Independent Normal Optimality Results

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

\[ \mu_1^0 \geq \mu_2^0 \geq \ldots \geq \mu_M^0 \]
\[ \sigma_1^0 \geq \sigma_2^0 \geq \ldots \geq \sigma_M^0, \]

then the knowledge-gradient policy is optimal.
Numerical Experiments

- 100 randomly generated problems
- $M$ Uniform $\{1, \ldots, 100\}$, $N$ Uniform $\{M, 3M, 10M\}$
- $\sigma^\varepsilon = 1$
- $\mu^0_x$ Uniform $[-1, 1]$
- $\mathbb{P}_0\{(\sigma^0_x)^2 = 1\} = .9$, $\mathbb{P}_0\{(\sigma^0_x)^2 = 10^{-3}\} = .1$, 


Numerical Experiments
Correlated Normal Model

- $\mathcal{X} = \mathcal{I} = \{1, \ldots, M\}$ for a finite integer $M$.
- Define $Y_x = F_x(\theta)$. Let $F$ and the distribution of $\theta$ be such that $Y = (Y_1, \ldots, Y_M)$ is a multivariate normal random variable.
- Define $\mu^n := \mathbb{E}_n Y$, $\Sigma^n := \text{Var}_n Y$.
- Let $(\epsilon^1, \ldots, \epsilon^N)$ be independent $N(0, (\sigma^\epsilon)^2)$.
- $\hat{y}^{n+1} = Y_{x^n} + \epsilon^{n+1}$
- Objective is $\mathbb{E}[Y_i] = \mathbb{E}[\max_x \mu_x^N]$.
Correlated Normal Transition Function

The transition function is given by

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

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

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

$$\mu^{n+1} = \mu^n + \tilde{\sigma}_{x^n}^n Z$$

for some standard normal random variable $Z$ independent of $\mathcal{F}_n$. 
The knowledge-gradient policy is

$$\begin{align*}
\text{arg max } & \boldsymbol{\mu}^{n+1}_x

= \text{arg max } \boldsymbol{\mu}^n_x + e^T_x \tilde{\sigma}^n_x Z

= \text{arg max } f(\mu^n_x, \tilde{\sigma}^n_x)

\text{where } f : \mathbb{R}^M \times \mathbb{R}^M \rightarrow \mathbb{R} \text{ is defined by } f(a, b) = \text{IE max}_x a_x + b_x Z.
\end{align*}$$
Algorithm for Computing $f(a, b) = \mathbb{E}\max_x a_x + b_x Z$

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 $\max_x a_x + b_x Z = \sum_x (a_x + b_x Z)1_{[c_{x-1}, c_x]}(Z)$,
   $$f(a, b) = \sum_x a_x (\Phi(c_x) - \Phi(c_{x-1})) + b_x (\phi(c_x) - \phi(c_{x-1}))$$
Correlated Normal Optimality Results

The knowledge-gradient policy is optimal for the correlated problem when

- $N = 1$
- $M = 2$
- $N \to \infty$ (pending verification)
- Can we bound suboptimality in general as in the independent normal case?
Future Work

- Asymptotic optimality for general case.
- Develop correlated normal case.
- Apply to approximate dynamic programming.
- Find counterexample demonstrating strict sub-optimality for independent normal case, or prove optimality.
- Improve and extend sub-optimality bound.
- Problem extensions: heterogeneous measurement costs; switching costs; continuous time; risk aversion; non-stationary measurement errors or $\theta$.
- Applications: neuroscience, maximum likelihood estimation, homeland security, stochastic algorithms (e.g. policy search), optimization of simulation, sensor management.