Sequential Search Problem

Chris Gu


1 Motivating example

A research department tries to find a new and cheaper way to produce some product. Two substitute technologies are being considered, the benefits of which are uncertain and would not be known until development work is completed. Because they produce the same product, no more than one technology would actually be used even if both were developed.

<table>
<thead>
<tr>
<th>Project</th>
<th>$\alpha$</th>
<th>$\omega$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cost</td>
<td>15</td>
<td>20</td>
</tr>
<tr>
<td>Duration</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Reward</td>
<td>100</td>
<td>55</td>
</tr>
<tr>
<td>Probability</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>Interest Rate</td>
<td>10% per period</td>
<td></td>
</tr>
</tbody>
</table>

The problem is to find a sequential search strategy which maximizes the expected present discounted value, assuming that not developing any project has value 0.

It is easy to show that developing only $\alpha$ or only $\omega$ is better than not researching either project.

- The expected value of researching $\alpha$ is

  $$-15 + \frac{1}{1.1}[0.5 \times 100 + 0.5 \times 55] = 55.5$$

  (1)

- The expected value of researching $\omega$ is

  $$-20 + \left(\frac{1}{1.1}\right)^2[0.2 \times 240 + 0.8 \times 0] = 19.7$$

  (2)

The next question is to identify which technique should be researched first. By any of the standard economic criteria, $\alpha$ dominates $\omega$. $\alpha$ has lower research cost, shorter development lag, higher expected reward, greater minimum reward, less variance. However, there is a crucial difference between the value of a project and the order in which it should be researched. $\alpha$ is worth more in the sense that the expected value of an optimal program with only $\omega$ is lower than with only $\alpha$. Nevertheless, it turns out that the optimal sequential strategy is to develop $\omega$ first.
Suppose $\alpha$ is developed first. If the payoff turns out to be 55, it would then be worthwhile to
develop $\omega$, because the expected value of that strategy would be
\[
-20 + \left(\frac{1}{1.1}\right)^2 \left[0.2 \times 240 + 0.8 \times 55\right] = 56,
\]
which is greater than the value at that point of not developing $\omega$, 55. If the payoff turns out to
be 100, it would then not be worthwhile to develop $\omega$, because the expected value of that strategy
would be
\[
-20 + \left(\frac{1}{1.1}\right)^2 \left[0.2 \times 240 + 0.8 \times 100\right] = 85.8,
\]
which is less than the value at that point of not developing $\omega$, 100. Consequently, the expected
value of an optimal policy beginning with developing $\alpha$ is
\[
-15 + \frac{1}{1.1} \left[0.5 \times 100 + 0.5 \times \left(-20 + \left(\frac{1}{1.1}\right)^2 \left[0.2 \times 240 + 0.8 \times 55\right]\right)\right] = 55.9
\]
Suppose now $\omega$ is developed first. A similar calculation shows the expected value of an optimal
policy that starts by developing $\omega$ is
\[
-20 + \left(\frac{1}{1.1}\right)^2 \left[0.2 \times 240 + 0.8 \times \left(-15 + \frac{1}{1.1} \left[0.5 \times 100 + 0.5 \times 55\right]\right)\right] = 56.3
\]
Thus, the optimal policy for this example has the counter-intuitive property that $\omega$ is researched
first.

2 Sequential search problem

An abstract formulation of this sequential search problem is as follows. Assume that there are $n$
closed boxes at the beginning of search, labeled Box $i$, $1 \leq i \leq n$. We make the following assumptions.

- Each box contains a potential reward of $x_i$ with probability distribution function $F_i(x_i)$,
  independent of the other rewards.
- It costs $c_i$ to open box $i$ and learn its contents.
- It takes a time lag of $t_i$ to know its value, and hence each box has its own discount rate
  $\beta_i = e^{-r t_i}$, $r \in [0, 1]$.

Boxes can be sampled sequentially, in whatever order is desired. An initial amount $x_0$ is available,
representing a fallback reward that could always be collected if no sampling were undertaken or if
every sampled reward happened to be less than $x_0$. At each stage, the decision maker must decide
whether or not to open a box and proceeds as follows.

- If she chooses to stop searching, the decision maker collects at that time the maximum reward
  she has thus far uncovered.
• If the decision maker continues sampling, she must select the next box to be opened, pay at that time the cost for opening it, and wait for the outcome. Then will come the next decision stage.

The decision maker tries to maximize expected present discounted value. The decision maker wants a sequential decision rule that will tell her at each stage whether or not to continue searching, and if so, which box to open next.

2.1 Dynamic programming formulation

The decision maker’s problem can be formally posed in a dynamic programming format. Let the collection of \( n \) boxes, denoted \( I = \{1, 2, \ldots, n\} \), be partitioned into any set \( S \) of sampled boxes and its complement \( S^c \) of closed boxes. That is,

\[
S \cup S^c = I, S \cap S^c = \emptyset \tag{7}
\]

Let variable \( y \) represent the maximum sampled reward (from the opened boxes and the initial fall-back reward)

\[
y = \max_{i \in S^c} x_i \tag{8}
\]

The state of the system at any time is given by the statistic \((S, y)\). Let \( \Psi(S, y) \) be the expected present discounted value of following an optimal policy from this time on with the state \((S, y)\). At stage \((S, y)\) the decision maker could terminate search, collecting reward \( y \), or, she might open box \( i \), for each \( i \in S^c \), which results in expected discounted net gain

\[
-c_i + \beta_i \left[ \Psi(S - \{i\}, y) \int_{-\infty}^{y} dF_i(x_i) + \int_{y}^{\infty} \Psi(S - \{i\}, x_i) dF_i(x_i) \right] \tag{9}
\]

Therefore, for each subset \( S \) of \( I \) and every \( y \), the value functions \( \Psi \) must satisfy the recursive relation

\[
\Psi(S, y) = \max_{i \in S} \left\{ y, \max_{i \in S^c} \left[ -c_i + \beta_i \left[ \Psi(S - \{i\}, y) \int_{-\infty}^{y} dF_i(x_i) \right. \right. \right.
\]

\[
+ \left. \left. \left. \int_{y}^{\infty} \Psi(S - \{i\}, x_i) dF_i(x_i) \right] \right\} \right\} \tag{10}
\]

2.2 Optimal strategy

The main insight is that we can characterize each box using a simple value, called “reservation price,” to help construct the optimal strategy.

Suppose for the moment there are just two boxes. One is the closed box \( i \), the other is an already opened hypothetical box offering reward \( z_i \). If the searcher elects not to open box \( i \), she receives the sure reward

\[
z_i. \tag{11}
\]
If she opens box \( i \), the searcher can expect a net benefit

\[
-c_i + \beta_i \left[ z_i \int_{-\infty}^{z_i} dF_i(x_i) + \int_{z_i}^{\infty} x_i dF_i(x_i) \right].
\] (12)

The closed and opened boxes are equivalent if the searcher is indifferent between opening box \( i \) and not opening it. In other words, equating (12) and (11) and rearranging, we obtain

\[
c_i = \beta_i \int_{z_i}^{\infty} (x_i - z_i) dF_i(x_i) - (1 - \beta_i) z_i.
\] (13)

If we solve this equation in \( z_i \), we get a number called the reservation price of box \( i \). The policy function of this DP problem is presented as the following decision strategy:

- **Stopping Rule**: Terminate search whenever the maximum sampled reward exceeds the reservation price of every closed box.

- **Selection Rule**: If a box is to be opened, it should be the closed box with highest reservation price.

As an illustration, suppose you have 3 boxes with their reservation prices: \( a : 3 \), \( b : 4 \), \( c : 2 \) and suppose we have the following search history with \( x_0 = 0 \).

<table>
<thead>
<tr>
<th>((S, y))</th>
<th>Action</th>
<th>Realized ( x_i )</th>
</tr>
</thead>
<tbody>
<tr>
<td>((a, b, c), 0))</td>
<td>Sample ( b )</td>
<td>1.5</td>
</tr>
<tr>
<td>((a, c), 1.5)</td>
<td>Sample ( a )</td>
<td>2.5</td>
</tr>
<tr>
<td>((c), 2.5)</td>
<td>Stop with ( a, 2.5 )</td>
<td></td>
</tr>
</tbody>
</table>

There are a couple of remarkable facts about this rule:

- The entire structure of an optimal policy has been reduced to a simple statement about reservation prices.

- The reservation price of each box is calculated by equating a hypothetical gain of stopping (11) not with the full gain of opening the box and continuing on in an optimal manner, but rather with the myopic gain of opening the box and terminating (12).

In other words, the reservation price of a box depends only on the properties of that box and is independent of all other search opportunities.

One can derive a few properties of reservation prices based on Equation (13):

- Reservation price is insensitive to the probability distribution of rewards at the lower end of the tail. Any rearrangement of the probability mass located below \( z_i \) leaves \( z_i \) unaltered.

- As rewards become more dispersed at the upper end of the distribution, the reservation price increases. Similarly, moving the probability mass of rewards to the right increases the reservation price.

- Reservation price decreases with greater search cost, increased search time, or a higher interest rate.

- Reservation price has the interpretation of internal rate of return.
2.3 Proof of optimality of the strategy

First, we can show that reservation prices based on Equation (13) is well defined. Define

\[ H_i(z) = \beta_i \int_z^\infty (x_i - z) \, dF_i(x_i) - (1 - \beta_i) \, z. \]  

(14)

One can show that \( H_i(z) \) is continuous and monotonically decreasing in \( z \), and \( H_i(-\infty) = \infty \), \( H_i(\infty) = -\infty (= 0 \text{ if } \beta_i = 1) \). Rewriting Equation (13) as \( H_i(z_i) = c_i \), we note that if \( c_i > 0 \) or \( \beta_i < 1 \), there exists a solution \( z_i \) to the equation \( H_i(z_i) = c_i \), which is unique.

The proof of the optimality of the strategy is by induction on the number of closed boxes.

- **Case** \( m = 1 \). The optimality of the strategy is easily demonstrated by directly applying the definition of the reservation price.

- **Case** \( m + 1 \) assuming \( m \) true. Consider a set \( S \) with \( m + 1 \) closed boxes and any value of the maximum sampled reward \( y \). Let \( j \) be a box (not necessarily unique) with the highest reservation price in the collection of \( m + 1 \) closed boxes.

\[ j \in S \quad z_j = \max_{i \in S} z_i \]  

(15)

If \( y \geq z_j \), we show the optimality of not opening any boxes. If one box were opened, by the strategy applied to \( m \) closed boxes it would be optimal to stop. Hence the question is whether opening exactly one box is better than not opening any, which is easily answered in the negative using the base case argument. The stopping criterion of the strategy is thus proved for \( m + 1 \) closed boxes. If \( y < z_j \), we first show that searching further is sub-optimal. Just opening box \( j \) and then stopping would yield a higher expected present discounted value. Thus, at least one box should be opened. We now show (by contradiction) that opening Box \( j \) is optimal. Suppose that it is optimal to open Box \( k \) first, where \( k \) is any box in \( S \) having a lower reservation price than \( j \):

\[ k \in S \quad z_k < z_j. \]

If box \( k \) is opened first, by the induction assumption on the strategy for \( m \) closed boxes, there is an exact prescription of what to do in an optimal policy thereafter. Let the expected discounted present value of opening box \( k \) and following the strategy thereafter, which is alleged to constitute the best strategy, be \( B \).

We will show that this strategy \( B \) is dominated by another strategy, hence obtaining a contradiction showing that it is optimal to open \( j \) first. We call this competing strategy \( A \): Open box \( j \) first. Then let \( h \) be a box with second biggest reservation price in the collection of \( m + 1 \) closed boxes

\[ h \in S - \{ j \} \quad z_h = \max_{i \in S - \{ j \}} z_i \]

If \( x_j \geq z_h \), terminate. Otherwise, open box \( k \) next. From then on proceed by the strategy. The rest of the proof, technical in its details, essentially shows that \( A > B \). See paper for details.
3 Limitations of the model

Many of the underlying assumptions of the present formulation are unrealistic. There has been no provision made for:

- adaptive learning about correlated probability distributions
- pay-as-you-go research (with the possibility of backing out of a project if prospects start looking unfavorable)
- parallel search activity
- risk aversion
- incomplete or no recall
- collecting some reward before search is terminated
- randomly generated new opportunities
- a binding time horizon
- uncertain search costs or search time