Optimal Stopping

Peerapat Phatpanichot 1005780558

2022-11-10

Optimal stopping is a tool used for choosing the time to take an action that would give the best results. A stopping time is a formula in which given $X_1, X_2, ..., X_n$ tells us whether to stop at step n.

Pay off function is a reward that you get when you choose to stop at certain state.

A value function is a expected payoff you get from using optimal stopping method and can be represent as $v(x) = (\mathbb{E}(f(X_t) \mid X_0) = x$, and we want to maximize $f(X_t)$

The general formula for optimal stopping where we maximize our reward is $v(x) = \max(f(x), \sum_y p(x,y)v(y))$ where $f(x)$ is reward when you stop and $\sum_y p(x,y)v(y)$ is the expected reward if you decided to continue.

We apply this optimal stopping theory into finding the love of your life where you will settle down with. The rules goes like this ,there are N candidates for becoming your wife. Each woman has a rank to herself from 0 to 10. You won't know this number N and you will go on a date with each candidate randomly and sequentially. After each date, you must reject or choose to marry that date. So once you reject you can't go back and choose somebody else and once you pass somebody up you can't go back and choose them. Optimal stopping can be used to find the best strategy for you to choose the best candidate to be your future wife.

If we look at the problem intuitively we come up with the strategy that we need to let a certain number of k-1 of candidates go by and select a pivot where after the pivot we will choose the better candidates than the previous one we skipped. Noted that this parameter k is in the interval of 1 to n, if k = 1 we select the first candidate if k = n we select the last candidate

Letting Y_n be the number of best candidates in order and $A_{n,k}$ be the event we accepted the best candidate

We can clearly see that Y_n is uniformly distributed on interval $1, 2, ..., n$ because candidates are randomly ordered.

The conditional probability of the best candidate being accepted are the following

$$P(A_{n,k} \mid Y_n = j) = \begin{cases} 0, & j \in 1, 2, .., k-1 \\ \frac{k-1}{j-1}, & j \in k, k+1, ..., n \end{cases}$$ for the first case it is obvious that if best candidate is j < k then it is impossible to choose best candidate and second case if best candidate $j \geq k$ then we will have accepted the best candidate if first $k-1$ candidates is in $j-1$

The probability of success if k=1 we simply accept the first candidate where this candidate is best with probability $\frac{1}{n}$. Otherwise if $k \in (2, .., n)$ we have a probability of $P(A_{n,k}) = \sum_{j-1}^{n} P(Y_n = j) \, P(A_{n,k} \mid Y_n = j) = \sum_{j=k}^{n} \frac{1}{n} \frac{k-1}{j-1}$

therefore,

$$P(A_{n,k} = \begin{cases} \frac{1}{n}, & k = 1 \\ \sum_{j=k}^{n} \frac{1}{n} \frac{k-1}{j-1}, & k \in 2, 3, .., n \end{cases}$$

We can see from the graph that our optimal number of candidate skipped to maximize our chance of getting the best candidate is around 37.
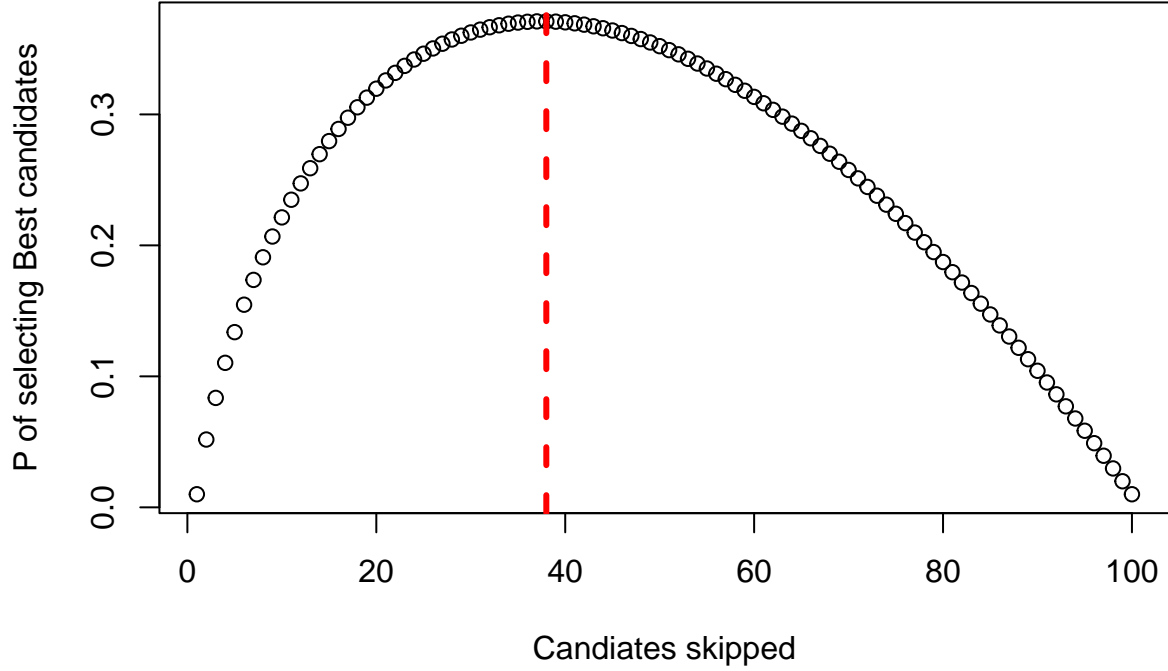
Another way to approach this problem is to use Bellman equation to find the general formula to choose the best candidate.

Here we are only interest is that the person we currently date is the best woman so far. At time t let we have states

$$x_t = \begin{cases} 1, & \text{if candidate is best so far} \\ 0, & \text{otherwise} \end{cases}$$

$$a_t = \begin{cases} 1, & \text{if candidate accepeted} \\ 0, & \text{if reject} \end{cases}$$

Knowing that $P(\text{t-th cadidate is best overall} \mid x_t = 1) = \frac{\frac{1}{N}}{\frac{1}{t}}$ gives us $\frac{t}{N}$

Letting our value function be $R_t = P(\text{the bext candidate} \mid \text{reject the first t-1})$

Now suppose at time $x_t = 0$ we got a candidate but not the best we've seen so far with probability $1 - \frac{1}{t}$

$$1 - \frac{1}{t} = \begin{cases} 1, & \text{reward} = 0 \\ 0, & \text{gives expected reward} R_{t+1} \end{cases}$$

On the other hand if we got the best candidate so far $x_t = 1$ with probability $\frac{1}{t}$

$$\frac{1}{t} = \begin{cases} 1, & \text{reward is } \frac{t}{N} \\ 0, & \text{gives expected reward} R_{t+1} \end{cases}$$

2

Using Bellman equation to get the long-termed reward taking account of current candidate and it's reward with expected reward from future candidate, we get the following equation

$R_t = \frac{1}{t}\max(\frac{t}{N}, R_{t+1}) + (1 - \frac{1}{t})\max(0, R_{t+1})$

$= \max(\frac{1}{N} + \frac{1-t}{t}R_{t+1}, R_{t+1})$ now we can see that maximum happens at $\begin{cases} R_{t+1}, & if\, R_{t+1} \geq \frac{t}{N} \\ \frac{t-1}{t}R_{t+1} + \frac{1}{N}, & if\, R_{t+1} < \frac{t}{N} \end{cases}$

Let t^* be the biggest t such that R^*_{t+1} is less than $\frac{t}{N}$ at the state where we reject all N candidate R_{N+1} we reward we get is 0 because we rejected everyone $R_{N+1} = 0$

from the case $\frac{t-1}{t}R_{t+1} + \frac{1}{N}$ i,f $R_{t+1} < \frac{t}{N}$

we get that $\frac{R_t}{t-1} = \frac{R_{t+1}}{t} + \frac{1}{N} \cdot \frac{1}{t-1}$

$\frac{R_{t^*}}{t^*-1} = \sum_{t=t^*}^{N} \frac{R_t}{t-1} - \frac{R_{t+1}}{t} = \sum_{t=t^*}^{N} \frac{1}{N(t-1)}$

now we have a policy such that we will reject all the candidates until that t^* and accept the better candidate after t^* we represents this by $\sum_{t=t^*}^{N} \frac{1}{t-1} \geq 1$

now if we have tons of candidates we can assume N -> ∞ we get $\int_{t^*}^{N} \frac{1}{t}dt$ which yields $log(N) - log(t^*)$ which will equal to $\frac{t^*}{N} = e^{-1}$ this means that if we have N amount of candidates we will reject e^{-1} proportion of candidates before we start choosing our best candidate according to the information we gathered from the e^{-1} candidate we reject

Optimal Skipped Canddiate for Dating with N = 1000
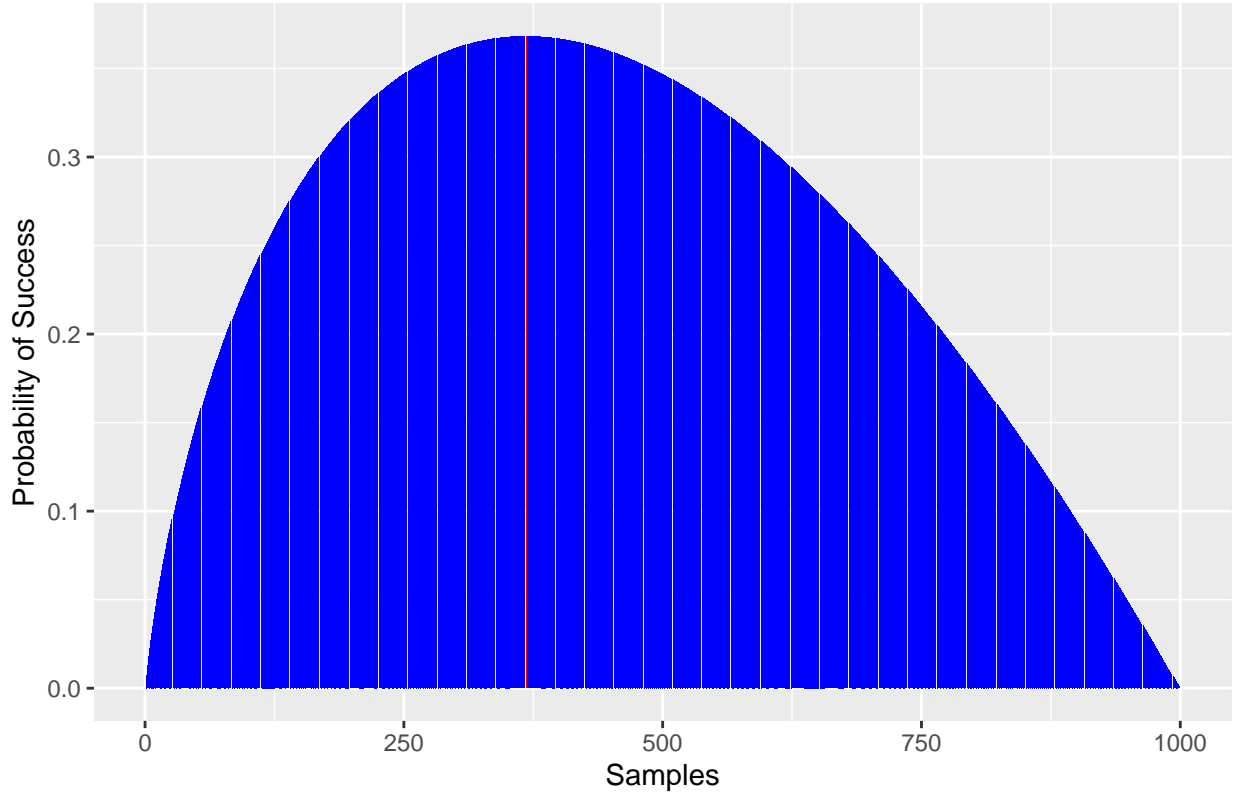
Moving forward let's now talk about Optimal stopping with discounting. The additional of discounting into optimal policy let us further examine our payoff function associated with optimal stopping time. In general, we define our

I. cost as C(i,a), C(i,a) $\geq$ 0 for all (i,a) where a is the action at time t

II. state we choose according to transition probability $P_{ij}(a)$

III. the discount factor α as 0 or 1

The goal is to minimize the expected total discounting cost for our stopping policy which is denoted by π where we can defined our function as follows

$V_\pi(i) = E_\pi[\sum_{t=0}^\infty \alpha^t C(X_t, a_t) \mid X_o = i]$ $i \geq 0$

$V_\pi(i)$ represent expected total discount cost when optimal policy π is used at initial state i. The concept of discounting factor is that the value decay overtime for example our reward starts at initial state with value of 1 over time at the next step it is worth α our discount factor and following step α^2 if our reward decay exponentially. Given our value discount cost at a rate of α per unit of time. The discount optimal is expected at expected discounted cost for all initial state therefore our optimal policy in regards of discounting can be represent as $V_\alpha(i)$

$V_\alpha(i) = \min(C(i,a) + \alpha \sum_j^\infty P_{ij}(a)(V_\alpha(j)))$

Given that we own a factory and want to optimal our management of machines, all equipment have an expire date and therefore need to be replace and we can find the optimal replacement with the optimal stopping, lets say given

I. if the number of defective items in produced more than upper threshold the machine will be replaced

II. if the machine produces defective items less than lower threshold the machine is not replaced.

III. now if the defective items produced is in the bound of upper threshold and lower threshold we keep the machine and possibly repair.

IV. c1 the lower threshold for number of defective items c2 the upper threshold for number of defective items

V. I represents the cost of inspecting, c represent the cost of producing defects, and R is cost of replacing machine.

Now we assumes that the machine produces items continuous and machines are inspected at every period.

We will be finding an optimal replacement policy by looking at the expected total cost of the system, ETC, with regards of expected total cost of not replacing the machine, EAC, and the expected total cost of replacing the machine, ERP, and expected total cost of inspecting and rearing machine, EI.In other words our expected total cost can be model as ETC = EAC + ERP + EI.

respectively the probability of each scenario happen is the following:

P of inspecting the machine = $\sum_{i=c_1+1}^{c_2} \binom{n}{i} * p^i * (1-p)^{n-i}$ = F(c_2)-F(c_1)

P of not replacing the machine = $\sum_{i=0}^{c_1} \binom{n}{i} * p^i * (1-p)^{n-i}$ = F(c_1)

P of replacing the machine = $\sum_{i=c_2+1}^{n} \binom{n}{i} * p^i * (1-p)$ = 1- F(c_2)

where F() represents the cumulative binomial distribution function with n and p as parameters

We can represents this process with a transition matrix of

$$\begin{bmatrix} F(c\_2) - F(c\_1) & F(c\_1) & 1 - F(c\_2) \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

from the transition matrix we can calculate the expected long-run number of times the machine is getting inspected before moving to stages where replace or not replacing the machine is in action. $S = \frac{1}{1-(F(c_2)-F(c_1))}$ and $S_1 = \frac{F(c_1)}{1-(F(c_2)-F(c_1))} = S_2 = \frac{1-F(c_2)}{1-(F(c_2)-F(c_2))}$ where S_1 and S_2 is the probability of machine being replace or not replaced

Therefore ETC can be represents as ETC $= (c * N * P * S_1) + (R * S_2) + I(S - 1)$

ETC $= \frac{1}{1-F(c_2)+F(c_1)} \left([cNp * F(c_1)] \right) + R[1 - F(c_2)] + I[F(c_2) - F(c_1)]$

Now given the probability of producing defective item is $p = 0.1$ the cost of producing defective item is $c = 7$, the cost of replacing machine, $R = 700$, the cost of inspection and repair, $I = 400$ the number of items we want to inspect is $n = 40$, $N = 1000$ we want to find the optimal value for control thresholds for machine replacement and binomial distribution formula is $nc_x p^x (1 - p)^{n-x}$

```
##   x Binomial_D Cumulative_binomial_D
## 1 1 0.06569281             0.0804737
## 2 2 0.14233443             0.2228081
## 3 3 0.20032253             0.4231307
## 4 4 0.20588704             0.6290177
## 5 5 0.16470963             0.7937273
## 6 6 0.10675624             0.9004836
```

```
##   c Cumulative_Binomial_D
## 1 1             0.0804737
## 2 2             0.2228081
## 3 3             0.4231307
## 4 4             0.6290177
## 5 5             0.7937273
## 6 6             0.9004836
```

```
##   C1 C2       ANS    C minimum
## 1  1  5 1694.9610  1,5   FALSE
## 2  4  9  930.8345  4,9   FALSE
## 3  9  6  665.4793  9,6   FALSE
## 4  5  3  591.8437  5,3   FALSE
## 5  8 10  705.6900 8,10   FALSE
## 6  6  1  519.7790  6,1   FALSE
```

5

Best Combination for Minimum Total Cost
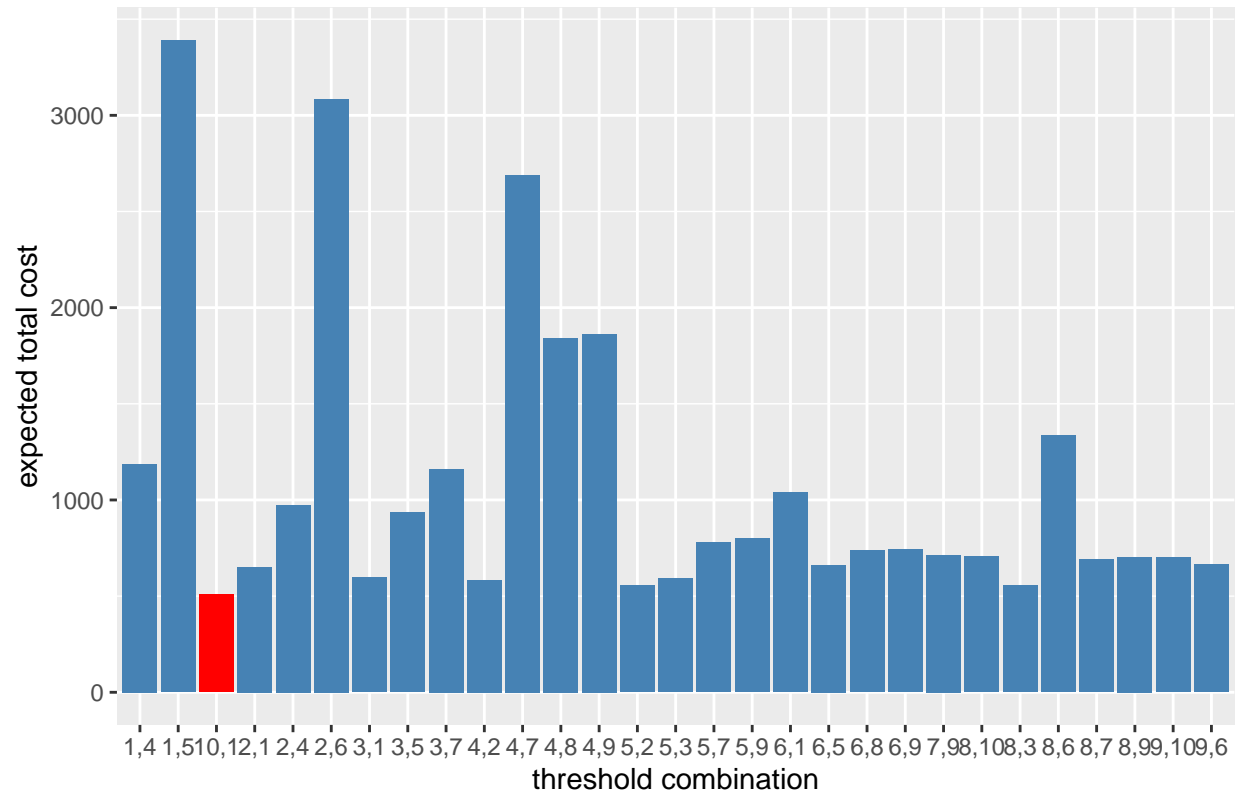
Here we can conclude that the best combination for our threshold value are when $c_1 = 1$ and $c_2 = 10$ with the minimum value of the expected total cost of 508.54$.

Now let's talk about Optimal stopping with cost, similarly to the optimal stoppign with discount we can denote our expected total cost for our stopping policy as

$V_\pi(i) = E_\pi[\sum_{t=0}^{\infty} C(X_t, a_t) \mid X_0 = i]$ where $i \geq 0$ notice that with discount there is a discount factor α but here our only concern is the non negative cost.

Similar to the optimal stopping with discount we can write our expected cost when it is minimal for all initial state as

$V(i) = min(C(i, a) + \sum_{j=0}^{\infty} P_{ij}(a)V(j)$, $i \geq 0$

When cost is incremented our goal is to minimize the long run expected average cost per unit of time.

Example.

Going back to the dating problem in our normal optimal stopping without cost and discount scenario. The situation happens that we are going on a marathon date in which we would like to find our soul mate on the day which limited amount of time and resources. The arrival of our date has a rate of λ and has a Poisson distribution. Each of our potential successive date is independent and can be represent with a (i, F_i) , with probability of $(i,F_i) $ occurring of $\sum_{i=1}^{N} P_i = 1$ where i is the rating of our date and F_i is the amount of time we spend together. Keep in mind that if we accept to go on this date the potential date during this timeline will be null. Here we want to maximize our chance of finding our soul mate (best rank) with limited amount of time.

Now we let $t_i = \int_0^\infty F_i(x)dx$ be the average time spent with i-date and 1 be accept and 2 reject and T as expected time util reject or accept occur (transition time) then we can come up with the followings

I. $P_i(1) = P_j$, $P_{ij}(2) = P_j$

II. T(i,1) = $t_i + \frac{1}{\lambda}$, T(i,2) = $\frac{1}{\lambda}$

III. C(i,1) = -i , C(i,2) = 0

then the optimal average cost policy in this scenario can be denoted as

$m(i) = \min[-i + \sum_{j=1}^{N} P_j m(j) - f(t_i + \frac{1}{\lambda}) ; \sum_{j=1}^{N} P_j h(j) - \frac{f}{\lambda}$ we can conclude that the optimal date (i, F_i) happens if and only if $\frac{i}{t_i} \geq -f$ where $-f$ is the optimal mean return time

References:

1) Practice. . ., T. M. (2016, June 12). Optimal Choice – Mathematical Advice for Real Life. R-Bloggers. https://www.r-bloggers.com/2016/06/optimal-choice-mathematical-advice-for-real-life/

2) The Secretary Problem. (n.d.). https://www.randomservices.org/random/urn/Secretary.html

3) The Secretary Problem and Its Extensions: A Review Author(s): P. R. Freeman Source: International Statistical Review / Revue Internationale de Statistique, Vol. 51, No. 2 (Aug., 1983), pp. 189-206 Published by: International

4) WordPress.com. (n.d.). https://appliedprobability.files.wordpress.com/2019/01/stochastic_control_2019

5) Ross, S. M. (1992). Applied Probability Models with Optimization Applications (Dover Books on Mathematics) (Reprint). Dover Publications. Chapter 6 - 7

6) A Machine Replacement Policy Based on M Arkov Chains - Acadpubl.eu. https://acadpubl.eu/hub/2018-119-15/3/541.pdf.

7) Kuhne R, Ruschendorf L, Optimal stopping with discount and observation costs, Journal of Applied Probability, 2000.

8) Stochastic Control with Applications to Finance, Neil Walton,January 26, 2019 https://appliedprobability.files.wordpress.com/2019/01/stochastic_control_2019.pdf