hw 1 Solution

chenyu wu

Sep 2023

1 Problem 1 (IE 5571: 20 pts/IE 8571: 15 pts)

Solution: To use dynamic programming to determine how to optimally allocate the five commercials over the four broadcasting regions to maximize the estimated number of votes won, we define the system as follows:
system stages: $i \in \{0, 1, \cdots, 4\}$ denotes the index of area, (with the initial stage 0)
system states: x_i - number of commercials left to determine for area $i+1, \cdots, 4$;
system controls: u_i - how many commercials to allocate on area $i+1$.

$$u_i \in \{0, 1, \cdots, x_i\} \quad \text{for } i \in \{0, 1, 2, 3\}$$

system dynamic: $x_{i+1} = f(x_i, u_i) = x_i - u_i$.
cost function: $g_i(x_i, u_i) = A_{u_i+1, i+1}$, where

$$A = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 4 & 6 & 5 & 3 \\ 7 & 8 & 9 & 7 \\ 9 & 10 & 11 & 12 \\ 12 & 11 & 10 & 14 \\ 15 & 12 & 9 & 16 \end{bmatrix}$$

dynamic programming problem:

$$V_i^*(x_i) = \max_{u_i} g_i(x_i, u_i) + V_{i+1}^*(f(x_i, u_i)) = A_{u_i+1, i+1} + V_{i+1}^*(x_i - u_i).$$

Then we solve this dynamic program problem by solving the tail problem gradually:

- Tail problem of length 1
$$V_4^*(x_4) = 0.$$
- Tail problem of length 2
$$V_3^*(x_3) = \max_{u_3} A_{u_3+1, 4} + 0, \quad u_3 \le x_3$$
Since the 4th column is monotonic increasing, we can see $u_3 = x_3$ is the optimal solution.
- Tail problem of length 3
$$V_2^*(x_2) = \max_{u_2} A_{u_2+1, 3} + A_{x_2 - u_2 + 1, 4}, \quad u_2 \le x_2.$$
Observe the matrix A, we can conclude:

- If $x_2 = 5, u_2 = 2, u_3 = 3, V_2^*(x_2) = 21$
- If $x_2 = 4, u_2 = 1, u_3 = 3, V_2^*(x_2) = 17$
- If $x_2 = 3, u_2 = 0/1/2, u_3 = 3/2/1, V_2^*(x_2) = 12$
- If $x_2 = 2, u_2 = 2, u_3 = 0, V_2^*(x_2) = 9$
- If $x_2 = 1, u_2 = 1, u_3 = 0, V_2^*(x_2) = 5$

- Tail problem of length 4

$$V_1^*(x_1) = \max_{u_1} A_{u_1+1,2} + V_2^*(x_1 - u_1), \quad u_1 \leq x_1.$$

Observe the matrix A, we can conclude by plugging in $V_2^*(x_2)$ we found in tail problem of length 3:

- If $x_1 = 5, u_1 = 1, u_2 = 1, u_3 = 3, V_1^*(x_1) = 23$
- If $x_1 = 4, u_1 = 1, u_2 = 0/1/2, u_3 = 3/2/1, V_1^*(x_1) = 18$
- If $x_1 = 3, u_1 = 1, u_2 = 2, u_3 = 0, V_1^*(x_1) = 15$
- If $x_1 = 2, u_1 = 1, u_2 = 1, u_3 = 0, V_1^*(x_1) = 11$
- If $x_1 = 1, u_1 = 1, u_2 = 0, u_3 = 0, V_1^*(x_1) = 6$

- Tail problem of length 5

$$V_0^*(x_0) = \max_{u_0} A_{u_0+1,2} + V_1^*(x_0 - u_0), \quad u_0 \leq x_0.$$

Observe the matrix A, we can conclude by plugging in $V_1^*(x_1)$ we found in tail problem of length 4: If $x_0 = 5, u_0 = 0, u_1 = 1, u_2 = 1, u_3 = 3, V_0^*(x_0) = 23$.

2 Problem 2 (20 pts/15 pts)

Solution: To use dynamic programming to maximize the amount of cash farmer J has on hand at the end of 3 months, we define the following dynamic programming problem:
system stages: $i \in \{0, 1, 2, 3\}$
system states: x_i - the amount of money and wheat J has at the end of the month i, where $x_i^{(1)}$ denotes the money and $x_i^{(2)}$ denotes the amount of wheat, e.g. $x_0 = (5000, 10)$.
system controls: $u_i, i \in \{0, 1, 2\}$ - amount of change in wheat during month $i + 1$.
From the restriction, we conclude:

$$-x_i^{(2)} \leq u_i \leq \min(10 - x_i^{(2)}, x_i^{(1)}/p_{i+1})$$

system dynamic: $x_{i+1}^{(1)} = x_i^{(1)} - p_{i+1}u_i$, $x_{i+1}^{(2)} = x_i^{(2)} + u_i$.
Cost function: $g_3(x_3) = x_3^{(1)}$ and $g_i(x_i) = 0, i \in \{0, 1, 2\}$.

Now we can use dynamic programming to formulate the tail problem:

- Tail problem of length 1

$$V_3^*(x_3) = x_3^{(1)}.$$

- Tail problem of length 2

$$V_2^*(x_2) = \max_{-x_2^{(2)} \le u_2 \le \min(10-x_2^{(2)}, x_2^{(1)}/p_3)} x_2^{(1)} - p_3 u_2.$$

  It is easy to get that $u_2^* = -x_2^{(2)}$ is the optimal solution.

- Tail problem of length 3

$$V_1^*(x_1) = \max_{-x_1^{(2)} \le u_1 \le \min(10-x_1^{(2)}, x_1^{(1)}/p_2)} x_1^{(1)} - p_2 u_1 + p_3(x_1^{(2)} + u_1).$$

  It is easy to get that

$$u_1^* = \begin{cases} -x_1^{(2)} & p_3 \le p_2 \\ \min(10 - x_1^{(2)}, x_1^{(1)}/p_2) & p_2 < p_3. \end{cases}$$

- Tail problem of length 4

$$V_1^*(x_0) = \max_{-x_0^{(2)} \le u_0 \le \min(10-x_0^{(2)}, x_0^{(1)}/p_1)} 5000 - p_1 u_0 - p_2 u_1^* - p_3 u_2^*.$$

  Notice that, u_1^* depends on the value of p_2 and p_3. Based on the value of p_1, p_2, p_3 we have the optimal solution:

$$u_0^* = \begin{cases} -10 & p_2 \le p_1 \\ 0 & p_2 > p_1 \end{cases}$$

3 Problem 3 (20 pts/15 pts)

3.1 (a) Write down a dynamic programming equation to study this system.

The DPE of this problem is:

$$\begin{aligned} V_N(x) &:= 0 \\ V_n(x) &:= \inf_{u \ge 0} \left[\bar{c}(u; x) + \beta \int_{\mathbb{R}} V_{n+1}^*(x + u - y) f(y) dy \right], \end{aligned}$$

where

$$\begin{aligned} \bar{c}(u; x) &:= pu + h \int_{\mathbb{R}} (x + u - y)^+ f(y) dy + b \int_{\mathbb{R}} (x + u - y)^- f(y) dy \\ &:= pu + L(x + u). \end{aligned}$$

We observe that

$$\frac{d}{du} L(x + u) = (h + b) \int_{-\infty}^{x+u} f(y) dy - b = (h + b) F(x + u) - b,$$

where $F(\cdot)$ is the CDF of Y_n.
Now we want to show that $L(\cdot)$ is convex and here we provide two approaches:

- It is not hard to see that the first derivative of L is non-decreasing if $h + b \geq 0$ due to the non-decreasing property of CDF. Therefore $L(\cdot)$ is a convex function.
- In fact, one can conclude this from another angle. As we defined functions $f(x) = x^+ := \max\{x, 0\}$ and $g(x) = x^- := -\min\{x, 0\}$ are convex, we can see that

$$L(x) = h\mathbb{E}[f(x - Y)] + b\mathbb{E}[g(x - Y)].$$

Then it is easy to see that

$$\begin{aligned}
L(\lambda x_1 + (1-\lambda)x_2) &= h\mathbb{E}[f(\lambda x_1 + (1-\lambda)x_2 - Y)] + b\mathbb{E}[g(\lambda x_1 + (1-\lambda)x_2 - Y)] \\
&= h\mathbb{E}[f(\lambda(x_1 - Y) + (1-\lambda)(x_2 - Y)] + b\mathbb{E}[g(\lambda(x_1 - Y) + (1-\lambda)(x_2 - Y)] \\
&\leq h\mathbb{E}[\lambda f(x_1 - Y) + (1-\lambda)f(x_2 - Y)] + b\mathbb{E}[\lambda g(x_1 - Y) + (1-\lambda)g(x_2 - Y)] \\
&= \lambda\left\{h\mathbb{E}[f(x_1 - Y)] + b\mathbb{E}[x_1 - Y]\right\} + (1-\lambda)\left\{h\mathbb{E}[f(x_2 - Y)] + b\mathbb{E}[x_2 - Y]\right\} \\
&= \lambda L(x_1) + (1-\lambda)L(x_2)
\end{aligned}$$

3.2 (b) Specify the value function and optimal policy when on stage N-1.

On stage $N - 1$, we have

$$V_{N-1}(x) = \inf_{u \geq 0} \bar{c}(u; x).$$

It follows from the convexity of L that $\bar{c}(u; x)$ is also convex. Therefore, we can simply let $\frac{d}{du}\bar{c}(u; x) = 0$ to obtain

$$p + (h + b)F(x + u) - b = 0 \Rightarrow u = F^{-1}\left(\frac{b - p}{b + h}\right) - x.$$

Then we define that

$$u^*_{N-1}(x) = F^{-1}\left(\frac{b - p}{b + h}\right) - x,$$

where $\bar{c}(u; x)$ achieves its global minimum.
Since $u_{N-1} \geq 0$ and convexity of $\bar{c}(u; x)$, we can simply let $u^*_{N-1}(x) = 0$ if $F^{-1}\left(\frac{b-p}{b+h}\right) - x \leq 0$ to reach the optimal. We conclude the optimal policy at the stage $N - 1$ as:

$$u^*_{N-1}(x) = \begin{cases} 0 & \text{if } x \geq F^{-1}\left(\frac{b-p}{b+h}\right) \\ F^{-1}\left(\frac{b-p}{b+h}\right) - x & \text{if } x < F^{-1}\left(\frac{b-p}{b+h}\right) \end{cases}$$

Plug-in the optimal policy, we have the optimal value:

$$V^*_{N-1}(x) = \begin{cases} L(x) & \text{if } x \geq F^{-1}\left(\frac{b-p}{b+h}\right) \\ L(F^{-1}\left(\frac{b-p}{b+h}\right)) + p(u^*_{N-1}(x)) & \text{if } x < F^{-1}\left(\frac{b-p}{b+h}\right). \end{cases}$$

Clearly, $V_{N_1}(x)$ is convex (even though it is the minimum of convex functions). Easy to see this by check that when $x < F^{-1}\left(\frac{b-p}{b+h}\right)$, $V^*_{N-1}(x)$ is a linear function and when $x \geq F^{-1}\left(\frac{b-p}{b+h}\right)$, $V^*_{N-1}(x) = L(x)$ is also convex.

3.3 (c) Can you say anything about the general structure of the optimal policy?

Now consider $n = N-2$. The DPE implies

$$V_{N-2}(x) := \inf_{u\geq 0}\left[pu + L(x+u) + \beta\int_{\mathbb{R}} V^*_{N-1}(x+u-y)f(y)dy\right].$$

The argument will be exactly the same except that function $L(x+u)$ is replaced by

$$\bar{L}(x+u) := L(x+u) + \beta\int_{\mathbb{R}} V^*_{N-1}(x+u-y)f(y)dy,$$

observing that $\bar{L}$ is again a convex function. To see this, we have $V_{N-1}(x+u-y)$ as a convex function and use a similar idea as the third idea of showing convexity of $L(\cdot)$, i.e.

$$\mathbb{E}[V^*_{N-1}(\lambda x_1 + (1-\lambda)x_2 - Y)] \leq \lambda\mathbb{E}[V^*_{N-1}(x_1 - Y)] + (1-\lambda)\mathbb{E}[V^*_{N-1}(x_2 - Y)]$$

Then we conclude that there exists a x^*_{N-2} such that

$$V^*_{N-2}(x) = \begin{cases} \bar{L} & \text{if } x \geq x^*_{N-2} \\ \bar{L} + b(x^*_{N-2} - x) & \text{if } x < x^*_{N-2} \end{cases}$$

and

$$u^*_{N-2}(x) = \begin{cases} 0 & \text{if } x \geq x^*_{N-2} \\ x^*_{N-2} - x & \text{if } x < x^*_{N-2}. \end{cases}$$

Again V^*_{N-2} is convex.

It is not difficult to see that for any arbitrary $n = 0, 1, \cdots$, the function V_n is convex. In particular, $v_n = V_0$ is convex. The optimal policy is determined by

$$u^*_n(x) = \begin{cases} 0 & \text{if } x \geq x^*_n \\ x^*_n - x & \text{if } x < x^*_n \end{cases}$$

for a sequence of thresholds $\{x^*_n\}$. Sometimes it is called a *threshold-type* policy.

4 Problem 4 (20 pts/15 pts)

4.1 (a) Write down a dynamic programming equation for this system.

To solve the equation

$$v(x_0) = \max_{u_0,\cdots,u_{N-1}}\left\{g_N(x_n) + \sum_{n=0}^{N-1} g(u_n x_n)\right\},$$

we define a dynamic programming equation:

$$V_i(x_i) = \max_{u_i} g_i(x_i u_i) + V_{i+1}(Rx_i(1-u_i)), \quad i \in \{0, \cdots, N-1\}.$$

4.2 (b)

We have reward function $g_N(x) = x^\alpha, g(x) = x^\alpha$ for $\alpha \in (0,1)$. This form of utility function is often called the "power utility" or "constant relative risk aversion utility".

We can start with $i = N-1$:

$$\begin{aligned} V_{N-1}(x) &= \sup_{0\le u_{N-1}\le 1}[g(u_{N-1}x) + g_N(R(1-u_{N-1})x)] \\ &= x^\alpha \cdot \sup_{0\le u_{N-1}\le 1}[u_{N-1}^\alpha + R^\alpha(1-u_{N-1})^\alpha] \\ &:= c_{N-1}x^\alpha. \end{aligned}$$

Since we can factor out the x^α, we see that the optimal u_{N-1}^* is *independent* of x.

For $n = N-2$, we also have:

$$\begin{aligned} V_{N-2}(x) &= \sup_{0\le u_{N-2}\le 1}[g(u_{N-2}x) + V(R(1-u_{N-2})x)] \\ &= x^\alpha \cdot \sup_{0\le u_{N-2}\le 1}[u_{N-2}^\alpha + c_{N-1}R^\alpha(1-u_{N-2})^\alpha] \\ &:= c_{N-2}x^\alpha. \end{aligned}$$

We can see all the tail problems will end up with $c_i x^\alpha$ because u_i^* is *independent* of x, i.e.

$$V_n(x) = c_n x^\alpha,$$

where the constants $\{c_n\}$ are recursively determined by

$$\begin{aligned} c_N &:= 1 \\ c_n &:= \sup_{0\le u_n\le 1}[u_n^\alpha + c_{n+1}R^\alpha(1-u_n)^\alpha, \quad n = 0, \cdots, N-1]. \end{aligned}$$

The optimal consumption sequence $\{u_n^*\}$ is a sequence of fixed proportions.

We conclude that, if the utility functions are of power type, the maximum utility function is also of power type, and the optimal consumption proportion may depend on the current time but is independent of the current total wealth.

5 Problem 5 (0 pts/15 pts)

We should $V_k = J_k^*$ by induction:

Consider $k = N-1$, we can obtain

$$V_{N-1}(x_{N-1}) = \min_{u_{N-1}\in\mathcal{U}(x_{N-1})}(g(x_{N-1}, u_{N-1}) + g_N(x_{N-1})) = J_{N-1}^*(x_{N-1})$$

for free.

Then suppose $k = N-n$, $V_k = J_k^*$ is true. We want to show that for $k = N-n-1$, $V_k = J_k^*$ is

true. We have:

$$\begin{aligned}
V_{N-n-1}(x_{N-n-1}) &= \min_{u_{N-n-1}} \{g(x_{N-n-1}, u_{N-n-1}) + V_{N-n}(f(x_{N-n-1}, u_{N-n-1}))\} \\
&= \min_{u_{N-n-1}} \left\{g(x_{N-n-1}, u_{N-n-1}) + J^*_{N-n}(x_{N-n})\right\} \\
&= \min_{u_{N-n-1}} \left\{g(x_{N-n-1}, u_{N-n-1}) + \min_{u_{N-n}, \cdots, u_{N-1}} \left[g_N(x_N) + \sum_{m=N-n}^{N-1} g(x_m, u_m)\right]\right\} \\
&= \min_{u_{N-n-1}, \cdots, u_{N-1}} \left\{g_N(x_N) + \sum_{m=N-n-1}^{N-1} g(x_m, u_m)\right\} \\
&= J^*_{N-n-1}(x_{N-n-1}).
\end{aligned}$$

Now we can state $V_k = J^*_k$ for all $k = 0, \cdots, N$ by induction. Note that $J^*_N(x_N) = g_N(x_N) = V_N(x_N)$ for free.

6 Problem 6 (20 pts/15 pts)

6.1 Exercise 2.4

Suppose we have weighted parameter α_k for step k, and we obtain:

$$\begin{aligned}
Q_{k+1} &= Q_k + \alpha_k \left[R_k - Q_k\right] \\
&= \alpha_k R_k + (1 - \alpha_k) Q_k \\
&= \alpha_k R_k + (1 - \alpha_k) \left[\alpha_{k-1} R_{k-1} + (1 - \alpha_{k-1}) Q_{k-1}\right] \\
&\cdots \\
&= \left[\prod_{\ell=1}^{k} (1 - \alpha_\ell)\right] Q_1 + \sum_{t=1}^{k} \left[\prod_{\ell=t+1}^{k} (1 - \alpha_\ell)\right] \alpha_t R_t
\end{aligned}$$

6.2 Exercise 2.5

The output should be similar to the following figure:

6.3 Exercise 2.7

We first show that Q_n is an exponential recency-weighted average. Observe that

$$\bar{o}_n = \alpha(1 + (1 - \alpha) + \cdots + (1 - \alpha)^{n-1}).$$

We denote $c := (1 - \alpha)$, and we can write

$$\beta_n = \frac{1}{1 + c + c^2 + \cdots + c^{n-1}}, \beta_1 = 1.$$

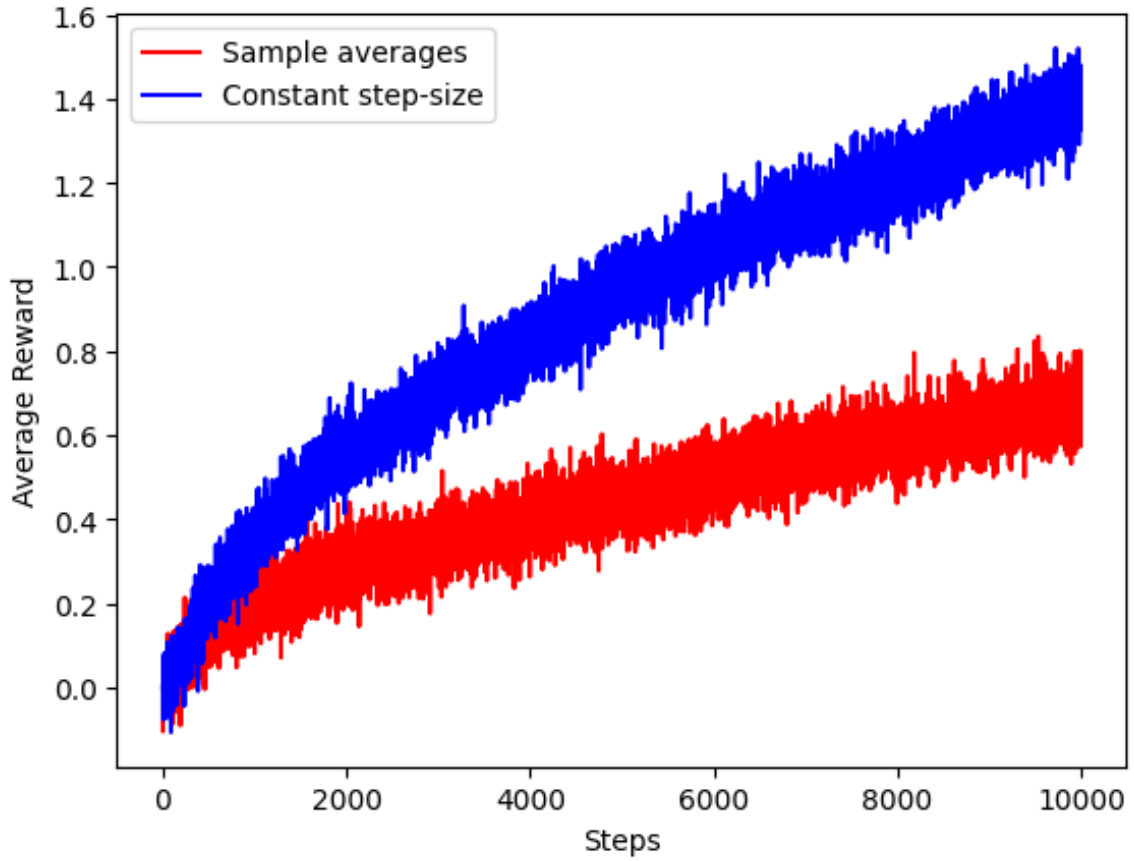

Figure 1: Exercise 2.5: Average reward from 500 experiments

Now we consider the formula we derive in exercise 2.4, we let α_n in exercise 2.4 be β_n here and derive:

$$\begin{aligned}\prod_{\ell=t+1}^{k}(1-\beta_\ell) &= \frac{c+\cdots+c^t}{1+c+\cdots+c^t}\frac{c(1+c+\cdots+c^t)}{1+c+\cdots+c^{t+1}}\cdots\frac{c(1+c+\cdots+c^{k-2})}{1+c+\cdots+c^{k-1}} \\ &= c^{k-t}\frac{1+c+\cdots+c^{t-1}}{1+c+\cdots+c^{k-1}}.\end{aligned}$$

Then we conclude

$$\begin{aligned}\sum_{t=1}^{k}\left[\prod_{\ell=t+1}^{k}(1-\beta_\ell)\right]\beta_t R_t &= \sum_{t=1}^{k} c^{k-t}\frac{1+c+\cdots+c^{t-1}}{1+c+\cdots+c^{k-1}}\frac{1}{1+\cdots+c^{t-1}}R_t \\ &= \sum_{t=1}^{k}\frac{1}{1+c+\cdots+c^{k-1}}c^{k-t}R_t \\ &:= \sum_{t=1}^{k}\hat{\alpha}c^{k-t}R_t,\end{aligned}$$

where $\hat{\alpha} = \frac{1}{1+c+\cdots+c^{k-1}}$ is a constant for a given k.

Now we consider the left term, which is the term related to the 'initial bias'. It is not hard to see that

$$\prod_{\ell=1}^{k}(1-\beta_\ell) = 0$$

due to $(1-\beta_1) = (1-1) = 0$. Therefore, we found that the term related to Q_1 disappeared.

7 Problem 7 (0 pts/10 pts)

This is an open question, so we will provide some brief ideas here:

If one does not have restriction $V_T \leq T/K$, it would be impossible to lower bound the regret. Considering a randomly large reward change, the value of μ_t^* of one arm other than the policy arm could be infinity, which is a trivial situation.