*MAST30022 Decision Making*
*2021*
*Tutorial 11*

1. **PS9-8** (cf. lecture slides 27-30)
For a price of \$1/gallon, the Safeco Supermarket chain has purchased 6 gallons of milk from a local dairy. Each gallon of milk is sold in the chain's three stores for \$2/gallon. The dairy must buy back for \$0.5/gallon any milk that is left at the end of the day. Demand for each of the three stores is uncertain. Past data indicate that the daily demand for each store is as shown in the following table.

| Store | Daily demand (in gallons) | Probability |
|---|---|---|
| 1 | 1 | 0.6 |
| | 2 | 0 |
| | 3 | 0.4 |
| 2 | 1 | 0.5 |
| | 2 | 0.1 |
| | 3 | 0.4 |
| 3 | 1 | 0.4 |
| | 2 | 0.3 |
| | 3 | 0.3 |

How should Safeco allocate the 6 gallons of milk to the three stores in order to maximise the expected net daily profit (revenues less costs) earned from milk?

**Solution**

Since the daily purchase cost is a constant (\$6), we may concentrate on the problem of allocating the milk to maximise daily expected revenue.

In the lecture we have introduced the following notation:

Stage $i$: store $i$, $i = 1, 2, 3$.

State at stage $i$: amount $x_i$ of milk (gallons) left for stores $i$, $i+1$,…,3.

Action at stage $i$: allocate $y_i$ gallons of milk to store $i$, where $y_i \in \{0, 1, 2, \ldots, x_i\}$.

Obviously, we have

$$\begin{aligned} x_{i+1} &= x_i - y_i,\ \ i = 1, 2, 3\ (x_4 = 0) \\ x_1 &= 6 \\ y_3 &= \min\{x_3, 3\}. \end{aligned}$$

Let $\mathbb{E}_i(y_i)$ be the expected revenue earned from $y_i$ gallons allocated to store $i$ and $f_i(x_i)$ be the maximum expected revenue earned from $x_i$ gallons allocated to stores $i, i+1, \ldots, 3$.

The DP equations are given by

$$\begin{aligned}
f_i(x_i) &= \max_{y_i \in \{0,1,\ldots,x_i\}} \{\mathbb{E}_i(y_i) + f_{i+1}(x_{i+1})\} \\
&= \max_{y_i \in \{0,1,\ldots,x_i\}} \{\mathbb{E}_i(y_i) + f_{i+1}(x_i - y_i)\}, \quad i = 1, 2 \\
f_3(x_3) &= \mathbb{E}_3(x_3),\ x_3 = 0, 1, 2, 3 \\
f_3(x_3) &= \mathbb{E}_3(3) + 0.5 \times (x_3 - 3), \quad x_3 = 4, 5, 6 \\
&.
\end{aligned}$$

Note that if $y_i > 3$, then $\mathbb{E}_i(y_i) = \mathbb{E}_i(3)$, therefore it would be irrational to assign more than 3 gallons to any store. In the computations we leave out the computations of $\mathbb{E}_i(y_i)$ with $y_i > 3$, since an optimal allocation will never assign more than 3 gallons to any store.

Note that the total revenue is \$2 per gallon sold + \$0.5 per gallon unsold.

Stage 3 computations:

Note that the dairy only buys back any milk once the last store has milk allocated to it.

$$\begin{aligned}
\mathbb{E}_3(3) &= 0.4 \times (\text{revenue when demand is 1}) + 0.3 \times (\text{revenue when demand is 2}) \\
&\quad + 0.3 \times (\text{revenue when demand is 3}) \\
&= 0.4 \times (2 + 2 \times 0.5) + 0.3 \times (2 \times 2 + 0.5) + 0.3 \times (3 \times 2) \\
&= 0.4 \times 3 + 0.3 \times 4.5 + 0.3 \times 6 \\
&= 4.35 \\
\mathbb{E}_3(2) &= 0.4 \times (2 + 0.5) + 0.3 \times (2 \times 2) + 0.3 \times (2 \times 2) \\
&= 3.4 \\
\mathbb{E}_3(1) &= 0.4 \times 2 + 0.3 \times 2 + 0.3 \times 2 \\
&= 2 \\
\mathbb{E}_3(0) &= 0.
\end{aligned}$$

Since $f_3(x_3) = \mathbb{E}_3(x_3)$, we have

| $x_3$ | $f_3(x_3)$ | Optimum solution |
|---|---|---|
| 0 | 0 | $y_3^* = 0$ |
| 1 | 2 | $y_3^* = 1$ |
| 2 | 3.4 | $y_3^* = 2$ |
| 3 | 4.35 | $y_3^* = 3$ |
| 4 | 4.85 | $y_3^* = 3$ |
| 5 | 5.35 | $y_3^* = 3$ |
| 6 | 5.85 | $y_3^* = 3$ |

Stage 2 computations:

$$\begin{aligned}
\mathbb{E}_2(3) &= 0.5 \times (\text{revenue when demand is 1}) + 0.1 \times (\text{revenue when demand is 2}) \\
&\quad + 0.4 \times (\text{revenue when demand is 3}) \\
&= 0.5 \times 2 + 0.1 \times 4 + 0.4 \times 6 \\
&= 3.8 \\
\mathbb{E}_2(2) &= 0.5 \times 2 + 0.1 \times 4 + 0.4 \times 4 \\
&= 3 \\
\mathbb{E}_2(1) &= 0.5 \times 2 + 0.1 \times 2 + 0.4 \times 2 \\
&= 2 \\
\mathbb{E}_2(0) &= 0.
\end{aligned}$$

Using $f_2(x_2) = \max_{y_2 \in \{0,1,\ldots,x_2\}}\{\mathbb{E}_2(y_2) + \mathbb{E}_3(x_2 - y_2)\}$ we can work out the following table (note that we do not need to consider the case where $y_2 > 3$).

| | $\mathbb{E}_2(y_2) + f_3(x_2 - y_2)$ | | | | Optimum solution | |
|---|---|---|---|---|---|---|
| $x_2$ | $y_2 = 0$ | $y_2 = 1$ | $y_2 = 2$ | $y_2 = 3$ | $f_2(x_2)$ | $y_2^*(x_2)$ |
| 0 | 0 | – | – | – | 0 | 0 |
| 1 | 2 | 2 | – | – | 2 | 0 or 1 |
| 2 | 3.4 | 4 | 3 | – | 4 | 1 |
| 3 | 4.35 | 5.4 | 5 | 3.8 | 5.4 | 1 |
| 4 | 4.85 | 6.35 | 6.4 | 5.8 | 6.4 | 2 |
| 5 | 5.35 | 6.85 | 7.35 | 7.2 | 7.35 | 2 |
| 6 | 5.85 | 7.35 | 7.85 | 8.15 | 8.15 | 3 |

Stage 1 computations:

We have that $x_1 = 6$. Furthermore,

$$\begin{aligned}
\mathbb{E}_1(3) &= 0.6 \times (\text{revenue when demand is 1}) + 0 \times (\text{revenue when demand is 2}) \\
&\quad + 0.4 \times (\text{revenue when demand is 3}) \\
&= 0.6 \times 2 + 0 \times 4 + 0.4 \times 6 = 3.6 \\
\mathbb{E}_1(2) &= 0.6 \times 2 + 0 \times 4 + 0.4 \times 4 = 2.8 \\
\mathbb{E}_1(1) &= 0.6 \times 2 + 0 \times 2 + 0.4 \times 2 = 2 \\
\mathbb{E}_1(0) &= 0.
\end{aligned}$$

| | $\mathbb{E}_1(y_1) + f_2(6 - y_1)$ | | | | Optimum solution | |
|---|---|---|---|---|---|---|
| $x_1$ | $y_1 = 0$ | $y_1 = 1$ | $y_1 = 2$ | $y_1 = 3$ | $f_1(x_1)$ | $y_1^*(x_1)$ |
| 6 | 8.15 | 9.35 | 9.2 | 9.0 | 9.35 | 1 |

Optimal allocation:

$$y_1^* = 1 \Longrightarrow x_2 = x_1 - y_1^* = 6 - 1 = 5 \Longrightarrow y_2^* = 2 \Longrightarrow x_3 = x_2 - y_2^* = 5 - 2 = 3 \Longrightarrow y_3^* = 3,$$

hence, allocate 1 gallon to store 1, 2 gallons to store 2 and 3 gallons to store 3.

The maximum expected revenue in both cases is \$9.35.

2. **(PS9-5)** The owner of a winery can advertise through one of three media: radio, TV, or newspaper. The weekly costs of an advertisement in the three media are 1, 2, and 1.5 thousand dollars, respectively. The winery classifies its sales volume during each week as (1) fair, (2) good, or (3) excellent. The transition probabilities associated with each choice of advertisement medium are as follows.

   Transition matrix for Radio:

| | fair | good | excellent |
|---|---|---|---|
| fair | 0.4 | 0.5 | 0.1 |
| good | 0.1 | 0.7 | 0.2 |
| excellent | 0.1 | 0.2 | 0.7 |

   (For example, if sales volume is good and the company chooses to advertise on radio, then with probability 0.2 sales will be excellent in the next period.)

   Transition matrix for TV:

| | fair | good | excellent |
|---|---|---|---|
| fair | 0.7 | 0.2 | 0.1 |
| good | 0.3 | 0.6 | 0.1 |
| excellent | 0.1 | 0.7 | 0.2 |

   Transition matrix for Newspaper:

| | fair | good | excellent |
|---|---|---|---|
| fair | 0.2 | 0.5 | 0.3 |
| good | 0 | 0.7 | 0.3 |
| excellent | 0 | 0.2 | 0.8 |

   A return value has been determined for each pair of states and each choice of media, as shown in the following tables.

   Weekly returns (in thousands) when advertising with Radio:

| | fair | good | excellent |
|---|---|---|---|
| fair | 4 | 5.2 | 6 |
| good | 3 | 4 | 7 |
| excellent | 2 | 2.5 | 5 |

   (For example, the entry in the 1st row and 2nd column of the return matrix is 5.2, so if in a fair state the company chooses to advertise on radio, and the subsequent state is good, the return to the company will be \$5,200.)

Weekly returns (in thousands) when advertising with TV:

| | fair | good | excellent |
|---|---|---|---|
| fair | 10 | 13 | 16 |
| good | 8 | 10 | 17 |
| excellent | 6 | 7 | 11 |

Weekly returns (in thousands) when advertising with Newspaper:

| | fair | good | excellent |
|---|---|---|---|
| fair | 4 | 5.3 | 7.1 |
| good | 3.5 | 4.5 | 8 |
| excellent | 2.5 | 4 | 6.5 |

The winery aims at the maximum expected net return over $N$ weeks.

(a) Formulate this problem as a Markov decision process.

(b) Find the optimal decision for the last week (i.e. for week $t = N$).

(c) Help the winery to make an optimal advertising policy so as to maximise the net return over 2 weeks. What is the maximum expected net return for two weeks?

### Solution

(a) For this problem there are 3 states, which we denote by 1 (fair), 2 (good), 3 (excellent). So the state space is

$$I = \{1 \text{ (fair)}, 2 \text{ (good)}, 3 \text{ (excellent)}\}.$$

In each state 3 decisions are available to the owner, which are 1 (advertising with radio), 2 (advertising with TV), 3 (advertising with newspaper). Thus the decision set $D(i)$ for each state $i = 1, 2, 3$ is

$$D(i) = \{1 \text{ (radio)}, 2 \text{ (TV)}, 3 \text{ (newspaper)}\}.$$

Transition matrix for action 1 (Radio):

$$P^{(1)} = (p_{ij}^{(1)}) = \begin{bmatrix} 0.4 & 0.5 & 0.1 \\ 0.1 & 0.7 & 0.2 \\ 0.1 & 0.2 & 0.7 \end{bmatrix}.$$

Transition matrix for action 2 (TV):

$$P^{(2)} = (p_{ij}^{(2)}) = \begin{bmatrix} 0.7 & 0.2 & 0.1 \\ 0.3 & 0.6 & 0.1 \\ 0.1 & 0.7 & 0.2 \end{bmatrix}.$$

Transition matrix for action 3 (Newspaper):

$$P^{(3)} = (p_{ij}^{(3)}) = \begin{bmatrix} 0.2 & 0.5 & 0.3 \\ 0 & 0.7 & 0.3 \\ 0 & 0.2 & 0.8 \end{bmatrix}.$$

Reward matrix for action 1 (Radio):

$$R^{(1)} = (r_{ij}^{(1)}) = \begin{bmatrix} 4 & 5.2 & 6 \\ 3 & 4 & 7 \\ 2 & 2.5 & 5 \end{bmatrix}.$$

Reward matrix for action 2 (TV):

$$R^{(2)} = (r_{ij}^{(2)}) = \begin{bmatrix} 10 & 13 & 16 \\ 8 & 10 & 17 \\ 6 & 7 & 11 \end{bmatrix}.$$

Reward matrix for action 3 (Newspaper):

$$R^{(3)} = (r_{ij}^{(3)}) = \begin{bmatrix} 4 & 5.3 & 7.1 \\ 3.5 & 4.5 & 8 \\ 2.5 & 4 & 6.5 \end{bmatrix}.$$

If the state is $i$ and the action is 1 (advertising with radio), then the expected net profit over one week is

$$r_i^{(1)} = \sum_{j=1}^{3} p_{ij}^{(1)} r_{ij}^{(1)} - 1.$$

Here we need to take away 1 thousand, which is the cost for advertising with radio. Similarly,

$$r_i^{(2)} = \sum_{j=1}^{3} p_{ij}^{(2)} r_{ij}^{(2)} - 2.$$

$$r_i^{(3)} = \sum_{j=1}^{3} p_{ij}^{(3)} r_{ij}^{(3)} - 1.5.$$

From the given matrices $P^{(k)}$ and $R^{(k)}$ for $k = 1, 2, 3$ we get:

$$r_1^{(1)} = 3.8, \quad r_2^{(1)} = 3.5, \quad r_3^{(1)} = 3.2$$

$$r_1^{(2)} = 9.2, \quad r_2^{(2)} = 8.1, \quad r_3^{(2)} = 5.7$$

$$r_1^{(3)} = 4.08, \quad r_2^{(3)} = 4.05, \quad r_3^{(3)} = 4.5.$$

(b) For $t = 1, 2, \ldots, N$ and $i = 1, 2, 3$, define $f_t(i)$ to be the maximum expected net return for the period from week $t$ to week $N$, given that the state in week $t$ is $i$. For the last week we have, for each $i$,

$$f_N(i) = \max_{k \in \{1,2,3\}} \{r_i^{(k)}\} = \max\{r_i^{(1)}, r_i^{(2)}, r_i^{(3)}\}.$$

Thus

$$f_N(1) = \max\{r_1^{(1)}, r_1^{(2)}, r_1^{(3)}\} = 9.2, \quad k^* = 2$$

(where $k^* = 2$ indicates that the optimal decision at time $t = N$ when in state 1 (fair) is to advertise with TV.)

$$f_N(2) = \max\{r_2^{(1)}, r_2^{(2)}, r_2^{(3)}\} = 8.1, \quad k^* = 2$$

$$f_N(3) = \max\{r_3^{(1)}, r_3^{(2)}, r_3^{(3)}\} = 5.7, \quad k^* = 2.$$

(c) For $1 \le t \le N-1$,

$$f_t(i) = \max_{k \in \{1,2,3\}} \left\{ r_i^{(k)} + \sum_{j=1}^{3} p_{ij}^{(k)} f_{t+1}(j) \right\}.$$

Solving this equation recursively for $t = N-1, N-2, \ldots, 1$, we can get $f_1(i)$ for each initial state $i$.

In the following we do computations for $N = 2$. From the computation above we have:

$$f_2(1) = 9.2, \quad k^* = 2$$
$$f_2(2) = 8.1, \quad k^* = 2$$
$$f_2(3) = 5.7, \quad k^* = 2.$$

Based on these data we now work out $f_1(i)$ for each $i$.

$i = 1$:

| $k$ | $r_1^{(k)} + p_{11}^{(k)} f_2(1) + p_{12}^{(k)} f_2(2) + p_{13}^{(k)} f_2(3)$ | |
|---|---|---|
| 1 | $3.8 + 0.4 \times 9.2 + 0.5 \times 8.1 + 0.1 \times 5.7$ | 12.1 |
| 2 | $9.2 + 0.7 \times 9.2 + 0.2 \times 8.1 + 0.1 \times 5.7$ | 17.83 |
| 3 | $4.08 + 0.2 \times 9.2 + 0.5 \times 8.1 + 0.3 \times 5.7$ | 11.68 |

$$\begin{aligned} f_1(1) &= \max_{k \in \{1,2,3\}} \{r_1^{(k)} + \textstyle\sum_{j=1}^{3} p_{1j}^{(k)} f_2(j)\} \\ &= \max\{12.1, 17.83, 11.68\} \\ &= 17.83, \quad k^* = 2 \end{aligned}$$

$i = 2$:

| $k$ | $r_2^{(k)} + p_{21}^{(k)} f_2(1) + p_{22}^{(k)} f_2(2) + p_{23}^{(k)} f_2(3)$ | |
|---|---|---|
| 1 | $3.5 + 0.1 \times 9.2 + 0.7 \times 8.1 + 0.2 \times 5.7$ | 11.23 |
| 2 | $8.1 + 0.3 \times 9.2 + 0.6 \times 8.1 + 0.1 \times 5.7$ | 16.29 |
| 3 | $4.05 + 0 \times 9.2 + 0.7 \times 8.1 + 0.3 \times 5.7$ | 11.43 |

$$
\begin{aligned}
f_1(2) &= \max_{k\in\{1,2,3\}} \{r_2^{(k)} + \textstyle\sum_{j=1}^{3} p_{2j}^{(k)} f_2(j)\} \\
&= \max\{11.23, 16.29, 11.43\} \\
&= 16.29, \quad k^* = 2
\end{aligned}
$$

$i = 3$:

| $k$ | $r_3^{(k)} + p_{31}^{(k)} f_2(1) + p_{32}^{(k)} f_2(2) + p_{33}^{(k)} f_2(3)$ | |
|---|---|---|
| 1 | $3.2 + 0.1 \times 9.2 + 0.2 \times 8.1 + 0.7 \times 5.7$ | 9.73 |
| 2 | $5.7 + 0.1 \times 9.2 + 0.7 \times 8.1 + 0.2 \times 5.7$ | 13.43 |
| 3 | $4.5 + 0 \times 9.2 + 0.2 \times 8.1 + 0.8 \times 5.7$ | 10.68 |

$$
\begin{aligned}
f_1(3) &= \max_{k\in\{1,2,3\}} \{r_3^{(k)} + \textstyle\sum_{j=1}^{3} p_{3j}^{(k)} f_2(j)\} \\
&= \max\{9.73, 13.43, 10.68\} \\
&= 13.43, \quad k^* = 2
\end{aligned}
$$

**Optimal policy for two weeks ($N = 2$):** Advertise with TV in each week, regardless sales volumes (states). The maximum expected return for two weeks is 17.83 if the initial sales volume is fair, 16.29 if the initial sales volume is good, and 13.43 if the initial sales volume is excellent.

3. **(PS9-7)** A machine in excellent condition earns \$100 profit per week; a machine in good condition earns \$70 per week; and a machine in bad condition earns \$20 per week. At the beginning of any week, a machine may be sent out for repairs at a cost of \$90. A machine that is sent out for repairs returns in excellent condition at the beginning of the next week. If a machine is not repaired, the condition of the machine evolves in accordance with the Markov chain shown in the following table. The company wants to maximize its expected discounted profit over an infinite horizon ($\alpha = 0.9$).

| | Next Week | | |
|---|---|---|---|
| This Week | Excellent | Good | Bad |
| Excellent | 0.7 | 0.2 | 0.1 |
| Good | 0 | 0.7 | 0.3 |
| Bad | 0 | 0.1 | 0.9 |

(a) Use the policy iteration method to determine an optimal stationary policy.

(b) Use linear programming to determine an optimal stationary policy.

(Adapted from "Operations Research: Appl. & Alg.", W. L. Winston, 4th ed., 2004)

**Solution**

(a) Omitted

(b)
- Time $t$ = week $t$
- State space: $\{1 \text{ (excellent)}, 2 \text{ (good)}, 3 \text{ (bad)}\}$

- Decision sets:
$$D(1) = \{1 \text{ (no repair)}\}$$
$$D(2) = D(3) = \{1 \text{ (no repair)}, 2 \text{ (repair)}\}$$

- $$p_{11}^{(1)} = 0.7, p_{12}^{(1)} = 0.2, p_{13}^{(1)} = 0.1$$
$$p_{21}^{(1)} = 0, p_{22}^{(1)} = 0.7, p_{23}^{(1)} = 0.3$$
$$p_{31}^{(1)} = 0, p_{32}^{(1)} = 0.1, p_{33}^{(1)} = 0.9$$

- $$p_{21}^{(2)} = 0.7, p_{22}^{(2)} = 0.2, p_{23}^{(2)} = 0.1$$
$$p_{31}^{(2)} = 0.7, p_{32}^{(2)} = 0.2, p_{33}^{(2)} = 0.1$$

- $$r_1^{(1)} = 100, r_2^{(1)} = 70, r_3^{(1)} = 20$$

- $$r_2^{(2)} = r_3^{(2)} = 100 - 90 = 10$$

Linear programming:
$$\min z = V_1 + V_2 + V_3$$
s.t.
$$V_1 - 0.9(0.7V_1 + 0.2V_2 + 0.1V_3) \geq 100 \quad (1 \in D(1))$$
$$V_2 - 0.9(0.7V_2 + 0.3V_3) \geq 70 \quad (1 \in D(2))$$
$$V_2 - 0.9(0.7V_1 + 0.2V_2 + 0.1V_3) \geq 10 \quad (2 \in D(2))$$
$$V_3 - 0.9(0.1V_2 + 0.9V_3) \geq 20 \quad (1 \in D(3))$$
$$V_3 - 0.9(0.7V_1 + 0.2V_2 + 0.1V_3) \geq 10 \quad (2 \in D(3))$$
$$V_1, V_2, V_3 \text{ are unrestricted in sign}$$
i.e.
$$\min z = V_1 + V_2 + V_3$$
s.t.
$$0.37V_1 - 0.18V_2 - 0.09V_3 \geq 100 \quad (1 \in D(1))$$
$$0.37V_2 - 0.27V_3 \geq 70 \quad (1 \in D(2))$$
$$-0.63V_1 + 0.82V_2 - 0.09V_3 \geq 10 \quad (2 \in D(2))$$
$$-0.09V_2 + 0.19V_3 \geq 20 \quad (1 \in D(3))$$
$$-0.63V_1 - 0.18V_2 + 0.91V_3 \geq 10 \quad (2 \in D(3))$$
$$V_1, V_2, V_3 \text{ are unrestricted in sign}$$
Using MATLAB, for example, we obtain the optimal solution to this LP problem,
$$(V_1, V_2, V_3) = (767.8, 683.8, 677.8)$$

and the 1st, 2nd and 5th constraints are binding.

**Optimal Policy:** Repair the machine if and only if it is in bad condition.

**Expected discounted profit:** If this policy is implemented, the maximum expected discounted profit over an infinite horizon is

$$V_1 = 767.8, V_2 = 683.8, \text{ or } V_3 = 677.8$$

if the machine's initial condition is excellent, good, or bad, respectively.