CHAPTER 19: MARKOV DECISION PROCESSES

19.2-1.

Bank One, one of the major credit card issuers in the United States has developed the portfolio control and optimization (PORTICO) system to manage APR and credit-line changes of its card holders. Customers prefer low APR and high credit lines, which can reduce the bank's profitability and increase the risk. Consequently, the bank faces the need to find a balance between revenue growth and risk. PORTICO formulates the problem as a Markov decision process. The state variables are chosen in a way to satisfy Markovian assumption as closely as possible while keeping the dimension of the state space at a tractable level. The resulting variables are (x, y), where x corresponds to the credit line and APR level and y represents the behavior variables. The transition probabilities are estimated from the available data. The objective is to maximize the expected net present value of the cash flows over a 36-month horizon. The dynamic programming equation for the decision periods of the problem is

$$V_t(x, y) = \max_{a \in A(x,y)} \left\{ r(x \pm a, y) + \beta \sum_{j \in \mathcal{S}} p(x \pm a, y; j) V_{t+1}(x \pm a, j) \right\},$$

where $r(\,\cdot\,)$ denotes the immediate net cash flow and β is the discount factor. The solution obtained is then adjusted to conform to business rules.

Benchmark tests are performed to evaluate the output policy. These tests suggest that the new policy improves profitability. By adopting this policy, Bank One is expected to increase its annual profit by more than $75 million.

19.2-2.

(a) Let the states $i = 0, 1, 2$ be the number of customers at the facility. There are two possible actions when the facility has one or two customers. Let decision 1 be to use the slow configuration and decision 2 be to use the fast configuration. Also let C_{ij} denote the expected net immediate cost of using decision j in state i. Then,

$$C_{11} = C_{21} = 3 - \tfrac{3}{5} \times 50 = -27$$
$$C_{12} = C_{22} = 9 - \tfrac{4}{5} \times 50 = -31$$
$$C_{01} = 3$$
$$C_{02} = 9$$

(b) In state 0, the configuration chosen does not affect the transition probabilities, so it is best to choose the slow configuration when there are no customers in line. Consequently, the number of stationary policies is four.

i	$d_i(R_1)$	$d_i(R_2)$	$d_i(R_3)$	$d_i(R_4)$
1	1	1	2	2
2	1	2	1	2

Policy	Transition Matrix	Expected Average Cost
R_1	$\begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{3}{10} & \frac{1}{2} & \frac{1}{5} \\ 0 & \frac{3}{5} & \frac{2}{5} \end{pmatrix}$	$C_1 = 3\pi_0 - 27\pi_1 - 27\pi_2$
R_2	$\begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{3}{10} & \frac{1}{2} & \frac{1}{5} \\ 0 & \frac{4}{5} & \frac{1}{5} \end{pmatrix}$	$C_2 = 3\pi_0 - 27\pi_1 - 31\pi_2$
R_3	$\begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{2}{5} & \frac{1}{2} & \frac{1}{10} \\ 0 & \frac{3}{5} & \frac{2}{5} \end{pmatrix}$	$C_3 = 3\pi_0 - 31\pi_1 - 27\pi_2$
R_4	$\begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{2}{5} & \frac{1}{2} & \frac{1}{10} \\ 0 & \frac{4}{5} & \frac{1}{5} \end{pmatrix}$	$C_4 = 3\pi_0 - 31\pi_1 - 31\pi_2$

(c)

Policy	π_0	π_1	π_2	Average Cost
R_1	0.3103	0.5172	0.1724	$C_1 = -17.69$
R_2	0.3243	0.5405	0.1351	$C_2 = -17.81$
R_3	0.4068	0.5085	0.0847	$C_3 = -16.83$
R_4	0.416	0.519	0.065	$C_4 = -16.87$

C_2 is the minimum, so the optimal policy is R_2, i.e., to use slow configuration when no customer or only one customer is present and fast configuration when there are two customers.

19.2-3.

(a) Let the states represent whether the student's car is dented, $i = 1$, or not, $i = 0$.

Decision	Action	State	Immediate Cost
1	Park on street in one space	0	$C_{01} = 0$
2	Park on street in two spaces	0	$C_{02} = 4.5$
3	Park in lot	0	$C_{03} = 5$
4	Have it repaired	1	$C_{14} = 50$
5	Drive dented	1	$C_{15} = 9$

(b) Assuming the student's car has no dent initially, once she decides to park in lot, state 1 will never be entered. In that case, the decision chosen in state 1 does not affect the expected average cost. Hence, it is enough to consider five stationary deterministic policies.

i	$d_i(R_1)$	$d_i(R_2)$	$d_i(R_3)$	$d_i(R_4)$	$d_i(R_5)$
0	1	1	2	2	3
1	4	5	4	5	–

Policy	Transition Matrix	Expected Average Cost
R_1	$\begin{pmatrix} 0.9 & 0.1 \\ 1 & 0 \end{pmatrix}$	$C_1 = 0\pi_0 + 50\pi_1$
R_2	$\begin{pmatrix} 0.9 & 0.1 \\ 0 & 1 \end{pmatrix}$	$C_2 = 0\pi_0 + 9\pi_1$
R_3	$\begin{pmatrix} 0.98 & 0.02 \\ 1 & 0 \end{pmatrix}$	$C_3 = 4.5\pi_0 + 50\pi_1$
R_4	$\begin{pmatrix} 0.98 & 0.02 \\ 0 & 1 \end{pmatrix}$	$C_4 = 4.5\pi_0 + 9\pi_1$
R_5	$\begin{pmatrix} 1 & 0 \\ – & – \end{pmatrix}$	$C_5 = 5\pi_0$

(c)

Policy	π_0	π_1	Average Cost
R_1	0.909	0.091	4.55
R_2	0	1	9
R_3	0.98	0.02	5.41
R_4	0	1	9
R_5	1	0	5 (if initially not dented)

The policy R_1 has the minimum cost, so it is optimal to park on the street in one space if not dented and to have it repaired if dented.

19.2-4.

(a) Let states 0 and 1 denote the good and the bad mood respectively. The decision in each state is between providing refreshments or not.

Decision	Action	State	Immediate Cost
1	Provide refreshments	0	$C_{01} = 14$
2	Not provide refreshments	0	$C_{02} = 0$
1	Provide refreshments	1	$C_{11} = 14$
2	Not provide refreshments	1	$C_{12} = 75$

(b) There are four possible stationary policies.

i	$d_i(R_1)$	$d_i(R_2)$	$d_i(R_3)$	$d_i(R_4)$
0	1	1	2	2
1	1	2	1	2

Policy	Transition Matrix	Expected Average Cost
R_1	$\begin{pmatrix} 0.875 & 0.125 \\ 0.875 & 0.125 \end{pmatrix}$	$C_1 = 14\pi_0 + 14\pi_1$
R_2	$\begin{pmatrix} 0.875 & 0.125 \\ 0.125 & 0.875 \end{pmatrix}$	$C_2 = 14\pi_0 + 75\pi_1$
R_3	$\begin{pmatrix} 0.125 & 0.875 \\ 0.875 & 0.125 \end{pmatrix}$	$C_3 = 14\pi_1$
R_4	$\begin{pmatrix} 0.125 & 0.875 \\ 0.125 & 0.875 \end{pmatrix}$	$C_4 = 75\pi_1$

(c)

Policy	π_0	π_1	Average Cost
R_1	0.875	0.125	$C_1 = 14$
R_2	0.5	0.5	$C_2 = 44.5$
R_3	0.5	0.5	$C_3 = 7$
$R4$	0.125	0.875	$C_4 = 65.625$

The optimal policy is R_3, i.e., to provide refreshments only if the group begins the night in a bad mood.

19.2-5.

(a) Let state 0 denote point over, two serves to go on next point and state 1 denote one serve left. The decision in each state is to attempt an ace or a lob.

Decision	Action	State	Immediate Cost
1	Attempt ace	0	$C_{01} = \frac{3}{8}\left(\frac{2}{3}(-1) + \frac{1}{3}(1)\right) = -\frac{1}{8}$
2	Attempt lob	0	$C_{02} = \frac{7}{8}\left(\frac{1}{3}(-1) + \frac{2}{3}(1)\right) = \frac{7}{24}$
1	Attempt ace	1	$C_{11} = \frac{3}{8}\left(\frac{2}{3}(-1) + \frac{1}{3}(1)\right) + \frac{5}{8}(1) = \frac{1}{2}$
2	Attempt lob	1	$C_{12} = \frac{7}{8}\left(\frac{1}{3}(-1) + \frac{2}{3}(1)\right) + \frac{1}{8}(1) = \frac{5}{12}$

(b) There are four possible stationary deterministic policies.

i	$d_i(R_1)$	$d_i(R_2)$	$d_i(R_3)$	$d_i(R_4)$
0	1	1	2	2
1	1	2	1	2

Policy	Transition Matrix	Expected Average Cost
R_1	$\begin{pmatrix} 3/8 & 5/8 \\ 1 & 0 \end{pmatrix}$	$C_1 = (-1/8)\pi_0 + (1/2)\pi_1$
R_2	$\begin{pmatrix} 3/8 & 5/8 \\ 1 & 0 \end{pmatrix}$	$C_2 = (-1/8)\pi_0 + (5/12)\pi_1$
R_3	$\begin{pmatrix} 7/8 & 1/8 \\ 1 & 0 \end{pmatrix}$	$C_3 = (7/24)\pi_0 + (1/2)\pi_1$
R_4	$\begin{pmatrix} 7/8 & 1/8 \\ 1 & 0 \end{pmatrix}$	$C_4 = (7/24)\pi_0 + (5/12)\pi_1$

(c)

Policy	π_0	π_1	Average Cost
R_1	0.615	0.385	$C_1 = 0.270$
R_2	0.615	0.385	$C_2 = 0.237$
R_3	0.889	0.111	$C_3 = 0.315$
$R4$	0.889	0.111	$C_4 = 0.306$

The optimal policy is R_3, i.e., to attempt lob in state 0 and ace in state 1.

19.2-6.

(a) Let states $i = 0, 1, 2$ represent the state of the market, $11,000$, $12,000$ and $13,000$ respectively. The decision is between two funds, namely the Go-Go Fund and the Go-Slow Mutual Fund. All the costs are expressed in thousand dollars.

Decision	Action	State	Immediate Cost
1	Invest in the Go-Go	0	$C_{01} = 0.5(-20) + 0.2(-50) = -20$
2	Invest in the Go-Slow	0	$C_{02} = 0.5(-10) + 0.2(-20) = -9$
1	Invest in the Go-Go	1	$C_{11} = 0.1(20) + 0.4(-20) = -6$
2	Invest in the Go-Slow	1	$C_{12} = 0.1(10) + 0.4(-10) = -3$
1	Invest in the Go-Go	2	$C_{21} = 0.2(50) + 0.4(20) = 18$
2	Invest in the Go-Slow	2	$C_{22} = 0.2(20) + 0.4(10) = 8$

(b) There are eight possible stationary policies.

i	$d_i(R_1)$	$d_i(R_2)$	$d_i(R_3)$	$d_i(R_4)$	$d_i(R_5)$	$d_i(R_6)$	$d_i(R_7)$	$d_i(R_8)$
0	1	1	1	1	2	2	2	2
1	1	1	2	2	2	1	1	2
2	1	2	2	1	1	2	1	2

All R_i's have the same transition matrix: $\begin{pmatrix} 0.3 & 0.5 & 0.2 \\ 0.1 & 0.5 & 0.4 \\ 0.2 & 0.4 & 0.1 \end{pmatrix}$.

Policy	Expected Average Cost
R_1	$C_1 = -20\pi_0 - 6\pi_1 + 18\pi_2$
R_2	$C_2 = -20\pi_0 - 6\pi_1 + 8\pi_2$
R_3	$C_3 = -20\pi_0 - 3\pi_1 + 8\pi_2$
R_4	$C_4 = -20\pi_0 - 3\pi_1 + 18\pi_2$
R_5	$C_5 = -9\pi_0 - 3\pi_1 + 18\pi_2$
R_6	$C_6 = -9\pi_0 - 6\pi_1 + 8\pi_2$
R_7	$C_7 = -9\pi_0 - 6\pi_1 + 18\pi_2$
R_8	$C_8 = -9\pi_0 - 3\pi_1 + 8\pi_2$

(c) $\pi = (0.171, 0.463, 0.366)$

Policy	Average Cost
R_1	0.39
R_2	-3.27
R_3	-1.881
R_4	1.779
R_5	3.66
R_6	-1.389
R_7	2.271
R_8	0

The optimal policy is R_2, i.e. to invest in the Go-Go Fund in states 0 and 1, in the Go-Slow Fund in state 2.

19.2-7.

(a) Let states 0 and 1 represent whether the machine is broken down or is running respectively. The decision is between Buck and Bill.

Decision	Action	State	Immediate Cost
1	Buck	0	$C_{01} = 0$
2	Bill	0	$C_{02} = 0$
1	Buck	1	$C_{11} = -1200$
2	Bill	1	$C_{12} = -1200$

(b) There are four possible stationary deterministic policies.

i	$d_i(R_1)$	$d_i(R_2)$	$d_i(R_3)$	$d_i(R_4)$
0	1	1	2	2
1	1	2	1	2

Policy	Transition Matrix	Expected Average Cost
R_1	$\begin{pmatrix} 0.4 & 0.6 \\ 0.6 & 0.4 \end{pmatrix}$	$C_1 = -1200\pi_1$
R_2	$\begin{pmatrix} 0.4 & 0.6 \\ 0.4 & 0.6 \end{pmatrix}$	$C_2 = -1200\pi_1$
R_3	$\begin{pmatrix} 0.5 & 0.5 \\ 0.6 & 0.4 \end{pmatrix}$	$C_3 = -1200\pi_1$
R_4	$\begin{pmatrix} 0.5 & 0.5 \\ 0.4 & 0.6 \end{pmatrix}$	$C_4 = -1200\pi_1$

(c)

Policy	π_0	π_1	Average Cost
R_1	0.5	0.5	$C_1 = -600$
R_2	0.4	0.6	$C_2 = -720$
R_3	0.545	0.455	$C_3 = -546$
$R4$	0.444	0.556	$C_4 = -667.2$

The largest expected average profit is given by R_2.

19.2-8.

(a) Let the states be the number of items in inventory at the beginning of the period and the decision be the number of items ordered. To conform to the software package, one needs to relabel the decisions as $1, 2, 3$ respectively. The cost matrix is:

c_{ik}	1	2	3
0	$40/3$	$56/3$	24
1	4	19	–
2	4	–	–

Let R_3 denote the policy to order 2 items when the inventory level is initially 0 and not to order when the inventory level is initially either 0 or 1. In other words, $d_0(R_3) = 3$ and $d_1(R_3) = d_2(R_3) = 1$.

$$P(R_3) = \begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 2/3 & 1/3 & 0 \\ 1/3 & 1/3 & 1/3 \end{pmatrix} \Rightarrow \pi = (4/9, 3/9, 2/9)$$

Expected average cost: $(4/9)C_{03} + (3/9)C_{11} + (2/9)C_{21} = 116/9 \approx \$12.89$/period

(b) There are $3^3 = 27$ stationary policies, since one can order $0, 1$ or 2 items in each state. However, only six of these are feasible. The remaining 21 policies are infeasible and the decision at least in one of the states leads to over capacity.

i	$d_i(R_1)$	$d_i(R_2)$	$d_i(R_3)$	$d_i(R_4)$	$d_i(R_5)$	$d_i(R_6)$
0	1	2	3	1	2	3
1	1	1	1	2	2	2
2	1	1	1	1	1	1

19.3-1.

(a) minimize $\quad 3y_{01} + 9y_{02} + 3y_{11} + 9y_{12} + 28y_{21} + 34y_{22}$

subject to $\quad y_{01} + y_{02} + y_{11} + y_{12} + y_{21} + y_{22} = 1$

$$y_{01} + y_{02} - \left(\tfrac{1}{2}y_{01} + \tfrac{1}{2}y_{02} + \tfrac{3}{10}y_{11} + \tfrac{2}{5}y_{12}\right) = 0$$

$$y_{11} + y_{12} - \left(\tfrac{1}{2}y_{01} + \tfrac{1}{2}y_{02} + \tfrac{1}{2}y_{11} + \tfrac{1}{2}y_{12} + \tfrac{3}{5}y_{21} + \tfrac{4}{5}y_{22}\right) = 0$$

$$y_{21} + y_{22} - \left(\tfrac{2}{10}y_{11} + \tfrac{1}{10}y_{12} + \tfrac{2}{5}y_{21} + \tfrac{1}{5}y_{22}\right) = 0$$

$$y_{ik} \geq 0 \text{ for } i = 0, 1, 2 \text{ and } k = 1, 2$$

(b) Using the simplex method, we find $y_{01} = 0.32432, y_{11} = 0.54054, y_{22} = 0.13514$ and the remaining y_{ik}'s are zero. Hence, the optimal policy uses decision 1 in states 0 and 1, decision 2 in state 2.

19.3-2.

(a) minimize $4.5y_{02} + 5y_{03} + 50y_{14} + 9y_{15}$

subject to $y_{01} + y_{02} + y_{03} + y_{14} + y_{15} = 1$

$$y_{01} + y_{02} + y_{03} - \left(\tfrac{9}{10}y_{01} + \tfrac{49}{50}y_{02} + y_{03} + y_{14}\right) = 0$$

$$y_{14} + y_{15} - \left(\tfrac{1}{10}y_{01} + \tfrac{1}{50}y_{02} + y_{15}\right) = 0$$

$$y_{01}, y_{02}, y_{03}, y_{14}, y_{15} \geq 0$$

(b) Using the simplex method, all y_{ik}'s turn out to be zero except that $y_{01} = 0.90909$ and $y_{14} = 0.09091$, so the policy that uses decision 1 in state 0 and decision 4 in state 1 is optimal.

19.3-3.

(a) minimize $14y_{01} + 14y_{11} + 75y_{12}$

subject to $y_{01} + y_{02} + y_{11} + y_{12} = 1$

$$y_{01} + y_{02} - \left(\tfrac{7}{8}y_{01} + \tfrac{1}{8}y_{02} + \tfrac{7}{8}y_{11} + \tfrac{1}{8}y_{12}\right) = 0$$

$$y_{11} + y_{12} - \left(\tfrac{1}{8}y_{01} + \tfrac{7}{8}y_{02} + \tfrac{1}{8}y_{11} + \tfrac{7}{8}y_{12}\right) = 0$$

$y_{ik} \geq 0$ for $i = 0, 1$ and $k = 1, 2$

(b) Using the simplex method, we find $y_{02} = y_{11} = 0.5, y_{01} = y_{12} = 0$, so the optimal policy is to use decision 2 in state 0 and decision 1 in state 1.

19.3-4.

(a) minimize $-\tfrac{1}{8}y_{01} + \tfrac{7}{24}y_{02} + \tfrac{1}{2}y_{11} + \tfrac{5}{12}y_{12}$

subject to $y_{01} + y_{02} + y_{11} + y_{12} = 1$

$$y_{01} + y_{02} - \left(\tfrac{3}{8}y_{01} + \tfrac{7}{8}y_{02} + y_{11} + y_{12}\right) = 0$$

$$y_{11} + y_{12} - \left(\tfrac{5}{8}y_{01} + \tfrac{1}{8}y_{02}\right) = 0$$

$y_{ik} \geq 0$ for $i = 0, 1$ and $k = 1, 2$

(b) Using the simplex method, we find $y_{02} = 0.8889, y_{11} = 0.1111, y_{01} = y_{12} = 0$, so the optimal policy is to use decision 2 (lob) in state 0 and decision 1 (ace) in state 1.

19.3-5.

(a) minimize $-20y_{01} - 9y_{02} - 6y_{11} - 3y_{12} + 18y_{21} + 8y_{22}$

subject to $y_{01} + y_{02} + y_{11} + y_{12} + y_{21} + y_{22} = 1$

$$y_{01} + y_{02} - \left(\tfrac{3}{10}y_{01} + \tfrac{3}{10}y_{02} + \tfrac{1}{10}y_{11} + \tfrac{1}{10}y_{12} + \tfrac{2}{10}y_{21} + \tfrac{2}{10}y_{22}\right) = 0$$

$$y_{11} + y_{12} - \left(\tfrac{5}{10}y_{01} + \tfrac{5}{10}y_{02} + \tfrac{5}{10}y_{11} + \tfrac{5}{10}y_{12} + \tfrac{4}{10}y_{21} + \tfrac{4}{10}y_{22}\right) = 0$$

$$y_{21} + y_{22} - \left(\tfrac{2}{10}y_{01} + \tfrac{2}{10}y_{02} + \tfrac{4}{10}y_{11} + \tfrac{4}{10}y_{12} + \tfrac{4}{10}y_{21} + \tfrac{4}{10}y_{22}\right) = 0$$

$y_{ik} \geq 0$ for $i = 0, 1, 2$ and $k = 1, 2$

(b) Using the simplex method, we find $y_{01} = 0.171, y_{11} = 0.463, y_{22} = 0.366$ and the remaining y_{ik}'s are zero. Hence, the optimal policy uses decision 1 (the Go-Go Fund) in states 0 and 1, decision 2 in state 2 (the Go-Slow Fund).

19.3-6.

(a) minimize $-1200y_{11} - 1200y_{12}$

subject to $y_{01} + y_{02} + y_{11} + y_{12} = 1$

$$y_{01} + y_{02} - \Big(0.4y_{01} + 0.5y_{02} + 0.6y_{11} + 0.4y_{12}\Big) = 0$$

$$y_{11} + y_{12} - \Big(0.6y_{01} + 0.5y_{02} + 0.4y_{11} + 0.6y_{12}\Big) = 0$$

$y_{ik} \geq 0$ for $i = 0, 1$ and $k = 1, 2$

(b) Using the simplex method, we find $y_{01} = 0.4, y_{12} = 0.6, y_{02} = y_{11} = 0$, so the optimal policy is to use decision 1 (Buck) in state 0 and decision 2 (Bill) in state 1.

19.3-7.

(a) minimize $\frac{40}{3}y_{01} + \frac{56}{3}y_{02} + 24y_{03} + 4y_{11} + 19y_{12} + 4y_{21}$

subject to $y_{01} + y_{02} + y_{03} + y_{11} + y_{12} + y_{21} = 1$

$$y_{01} + y_{02} - \left(y_{01} + \tfrac{2}{3}y_{02} + \tfrac{1}{3}y_{03} + \tfrac{2}{3}y_{11} + \tfrac{1}{3}y_{12} + \tfrac{1}{3}y_{21}\right) = 0$$

$$y_{11} + y_{12} - \left(\tfrac{1}{3}y_{02} + \tfrac{1}{3}y_{03} + \tfrac{1}{3}y_{11} + \tfrac{1}{3}y_{12} + \tfrac{1}{3}y_{21}\right) = 0$$

$$y_{21} - \left(\tfrac{1}{3}y_{03} + \tfrac{1}{3}y_{11} + \tfrac{1}{3}y_{12} + \tfrac{1}{3}y_{21}\right) = 0$$

$y_{ik} \geq 0$ for $i = 0, 1, 2$ and $k = 1, 2, 3$

(b) Using the simplex method, we find $y_{03} = 0.4444, y_{11} = 0.3333, y_{21} = 0.2222$ and the remaining y_{ik}'s are zero. Hence, the optimal policy is to order 2 items in state 0 and not to order in states 1 and 2.