

題號 13.2-4

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	$C_3 = 7$
R_4	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.

題號 13.2-5

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	$C_3 = 0.315$
R_4	0.889	0.111	$C_4 = 0.306$

0.115

0.083

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

~~R_2~~ ~~ace~~ ~~lob~~