Mid Term Exam – I Solutions

1a) 2 identical, balanced coins

Let, A: Both coins show head
B : First shows head

$$P(A|B) = \frac{P(A\cap B)}{P(B)} \quad -(1)$$

$P(A\cap B) = P$ (Both coins are head and first is head)

$$= \frac{1}{4} \quad -(2)$$

Options: HH, HT, TT, TH

favourable

$P(B) = P(\text{First shows head}) = \frac{1}{2}$ — (3)

(2) & (3) in (1)

$$P(A/B) = \frac{1}{4} \div \frac{1}{2} = \boxed{\frac{1}{2}}$$

1b) Let A: Both are heads
B: Atleast one is head.

$$P(A/B) = \frac{P(A \cap B)}{P(B)} \quad \text{— (1)}$$

$P(A \cap B) = P(\text{Both are heads and atleast one is head})$

$$= \frac{1}{4} \quad \text{— (2)} \quad (\text{Refer 1a})$$

$P(B) = P(\text{Atleast one is head}) = 1 - P(\text{no head}) = 1 - \frac{1}{4} = \frac{3}{4}$

(3)

② & ③ in ①

$$P(A/B) = \frac{1}{4} \div \frac{3}{4} = \boxed{\frac{1}{3}}$$

2) 3-armed bandit

$$r_1 = \begin{cases} 4 & \text{w.p } 3/4 \\ 100 & \text{w.p } 1/4 \end{cases} \qquad r_2 = \begin{cases} 4 & \text{w.p } 1/2 \\ 100 & \text{w.p } 1/2 \end{cases}$$

$$r_3 = \begin{cases} 4 & \text{w.p } 1/4 \\ 100 & \text{w.p } 3/4 \end{cases}$$

a) $E(r_1) = \sum_r p(r) * r = \frac{3}{4} * 4 + \frac{1}{4} * 100$

$= 3 + 25 = 28$

$E(r_2) = 4 \times \frac{1}{2} + 100 \times \frac{1}{2} = 2 + 50 = 52$

$E(r_3) = 4 \times \frac{1}{4} + 100 \times \frac{3}{4} = 1 + 75 = 76$

b) Arm 3 is to be pulled often as average reward of this arm is highest.

c) Reward variance $= \Sigma(\text{Reward} - \text{mean})^2 / N$

$$\text{Var}(\text{arm}1) = [(4-28)^2 + (100-28)^2] / 2$$

$$= [576 + 5184] / 2 = 2880$$

c) Variance $= E(x^2) - (E(x))^2$

<u>Arm1</u>: $E(x^2) = \frac{3}{\cancel{4}} \times \cancel{4} \times 4 + \frac{1}{\cancel{4}} \times \cancel{100}^{25} \times 100$

$$= 12 + 2500 = 2512$$

$$[E(x)]^2 = 28 \times 28 = 784$$

$\therefore$ Variance $[\text{Arm}1] = 2512 - 784 = 1728$

Arm 2 : $E(x^2) = \frac{1}{\cancel{2}} \times \overset{2}{\cancel{4}} \times 4 + \frac{1}{\cancel{2}} \times \overset{50}{\cancel{100}} \times 100$

$= 8 + 5000 = 5008$

$[E(x)]^2 = 52 \times 52 = 2704$

Variance [Arm 2] $= 5008 - 2704 = 2304$

Arm 3 : $E(x^2) = \frac{1}{\cancel{4}} \times \cancel{4} \times 4 + \frac{3}{\cancel{4}} \times \overset{25}{\cancel{100}} \times 100$

$= 4 + 7500 = 7504$

$[E(x)]^2 = 76 \times 76 = 5776$

Variance [Arm 3] $= 7504 - 5776 = 1728$

d) Both Arms - Arm 1 and Arm 3 provides lower variance [~~A~~ of 1728] - Both are optimal to be pulled for lower variance

e) Variance : $(\text{Reward} - \text{mean})^2 / (N+1)$

Algorithm Based on Min. Reward Variance

① Initialise arms 1 to 3

$Q(a) \leftarrow 0$; $N(a) \leftarrow 0$; $RV(a) \leftarrow 0$

Loop forever:

$$A \leftarrow \begin{cases} \underset{a}{\operatorname{argmin}}\ RV(a) & \text{prob } 1-\epsilon \\ \text{A random action} & \epsilon \end{cases}$$

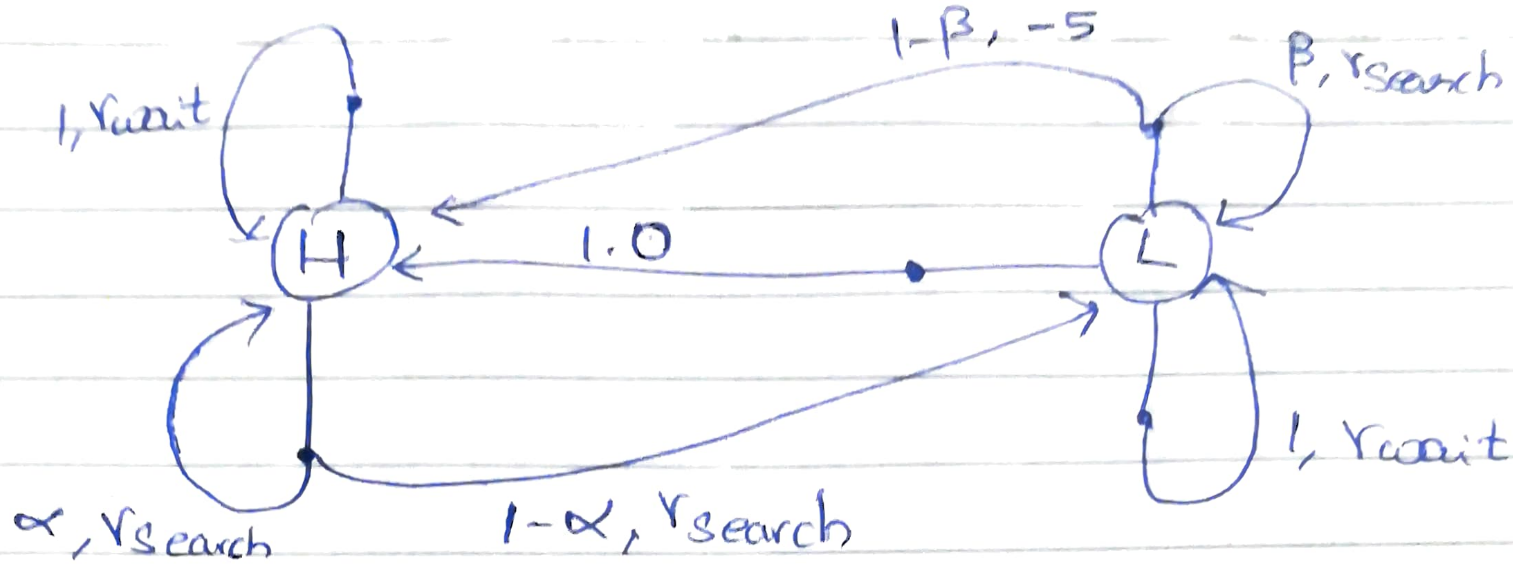
$R \leftarrow \text{bandit}(A)$

$N(A) \leftarrow N(A) + 1$

$$Q(A) \leftarrow Q(A) + \frac{1}{N(A)} \left(R - Q(A)\right)$$

$$RV(A) \leftarrow RV(A) + \frac{1}{N(A)+1} \left(R - Q(A)\right)^2$$

③



a) States: {High, Low}

Actions:

A(High) : {Wait, Search}

A(Low) : {Wait, Search, Recharge}

Rewards:

$r_{search} = 5$; $r_{wait} = 1$

Parameters:

$\gamma = 0.9$, $\alpha = 0.3$, $\beta = 0.5$

Bellman Equation

$$V(s) = \sum_{a} \pi(a|s) \sum_{r,s'} p(s',r|s,a)\,[r + \gamma V(s')]$$

$V(\text{High})$ = Value from action Wait + Value from action Search (Equi-probable)

$$= \frac{1}{2}\big[p(\text{high}, r_{wait} \mid \text{high}, \text{wait})\,[r_{wait} + \gamma V(\text{high})]\big]$$

$$+ \frac{1}{2}\big[p(\text{high}, r_{search} \mid \text{high}, \text{search})\,[r_{search} + \gamma V(\text{high})] +$$

$$p(\text{low}, r_{search} \mid \text{high}, \text{search})\,[r_{search} + \gamma V(\text{low})]\big]$$

$$V(High) = \frac{1}{2}\left[1[r_{wait} + \gamma V(High)]\right]$$
$$+ \frac{1}{2}\left[\alpha\,[r_{search} + \gamma V(High)] + \right.$$
$$\left.(1-\alpha)\,[r_{search} + \gamma V(low)]\right]$$

$$2V(High) = 1 + 0.9V(High)$$
$$+ 0.3[5 + 0.9V(high)] +$$
$$+ 0.7[5 + 0.9V(low)]$$

$$2V(High) = 1 + 0.9V(High)$$
$$+ 1.5 + 0.27V(High)$$
$$+ 3.5 + 0.63V(Low)$$

$$2V(High) = 6 + 1.17V(High) + 0.63V(Low)$$

$$\boxed{0.83V(High) - 0.63V(Low) = 6} \rightarrow (1)$$

$V(Low)$ = Value from action Wait + (Equi-probable)
Value from action Recharge +
Value from action Search

$$= \frac{1}{3}\left[p(low, r_{wait} \mid low, wait)\left[r_{wait} + \gamma V(Low)\right]\right]$$

$$+ \frac{1}{3}\left[p(high, r_{rc} \mid low, recharge)\left[r_{rc} + \gamma V(high)\right]\right]$$

$$+ \frac{1}{3}\left[p(high, r_{search} \mid low, search)\left[r_{search} + \gamma V(high)\right] + p(low, r_{search} \mid low, search)\left[r_{search} + \gamma V(low)\right]\right]$$

$$3V(Low) = 1(1 + 0.9V(Low)) +$$
$$r(0 + 0.9V(High)) +$$
$$(1-\beta)(-5 + 0.9V(high))) +$$
$$\beta(5 + 0.9V(low)))$$

$$3V(low) = 1 + 0.9V(low) + 0.9V(high)$$
$$- 2.5 + 0.45V(high)$$
$$+ 2.5 + 0.45V(low)$$

$$3V(low) = 1 + 1.35V(low) + 1.35V(high)$$

$$\boxed{1.65V(low) - 1.35V(high) = 1} \quad - \quad (2)$$

solving (1), (2)

$$V(high) = \overset{20.29}{\cancel{26.36}}; \quad V(low) = \overset{17.21}{\cancel{25.20}}$$

b) $v^*(\text{high}) = ?$ $\quad v^*(\text{low}) = ?$

Optimal Actions : $A^*(\text{high})$: Search
$A^*(\text{low})$ = Recharge

$$v^*(s) = \max_a \sum_{s',r} p(s',r|s,a)\left[r + \gamma v^*(s')\right]$$

$$\cancel{v^*(\text{high}) = \frac{1}{2}\left[r_{\text{search}} + \gamma\right.}$$

$$v^*(h) = p(\text{high}, r_{\text{search}} | \text{high}, \text{search})\left[r_{\text{search}} + \gamma v^*(\text{high})\right] + p(\text{low}, r_{\text{search}} | \text{high}, \text{search})\left[r_{\text{search}} + \gamma v^*(\text{low})\right]$$

$$v^*(\text{high}) = \alpha\left[r_{\text{search}} + \gamma v^*(\text{high})\right] +$$

$$(1-\alpha)\left[r_{search} + \gamma v^*(low)\right]$$

$$v^*(high) = 0.3\left[5 + 0.9 v^*(high)\right] + 0.7\left[5 + 0.9 v^*(low)\right]$$

$$v^*(high) = 1.5 + 0.27 v^*(high) + 3.5 + 0.63 v^*(low)$$

$$\boxed{0.73 v^*(high) - 0.63 v^*(low) = 5} \rightarrow (1)$$

$$v^*(low) = p(high, r_{recharge} \mid low, recharge)\left[r_{recharge} + \gamma v^*(high)\right]$$

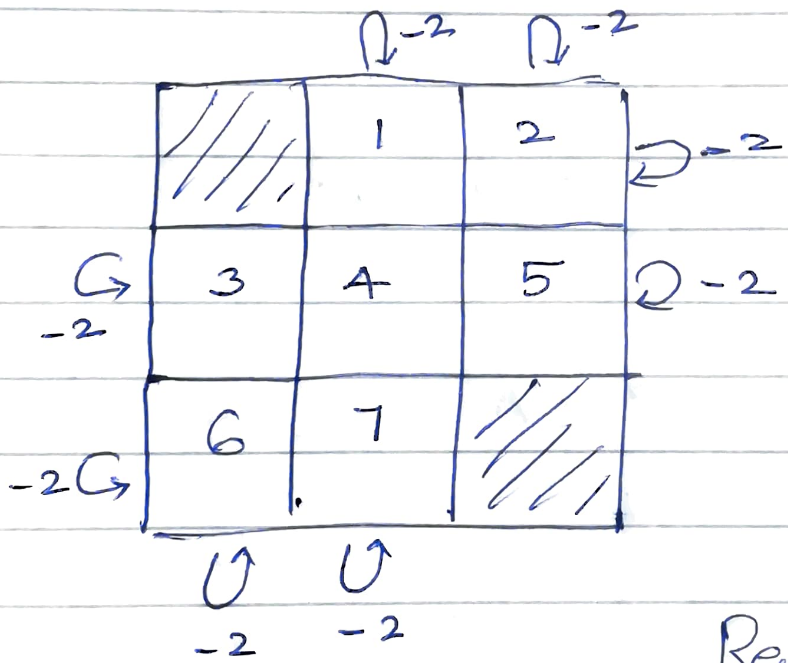
$$v^*(low) = 1\left[0 + 0.9 v^*(high)\right]$$

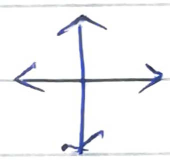
$$\boxed{v^*(low) = 0.9 v^*(high)} - (2)$$

Solving (1), (2)

$v^*(\text{high}) = 30.67$; $v^*(\text{low}) = 27.61$

(4)





Equiprobable policy

Reward (interstate) = -1

a) $V_0(s) = 0 \quad s \in (1, \ldots, 7)$; $\gamma = 1$

Step 1:

$$V_1(S_1) = \sum_a \pi(a|s_1) \sum_{r,s'} p(r,s'|s_1,a) \left[r + \gamma V(s') \right]$$

$$V_1(S_1) = \frac{1}{4}\left[1(-2+\gamma V(S_1)) + 1(-1+\gamma V(S_2))\right.$$
$$\left. + 1(-1+\gamma V(S_4)) + 1(-1+\gamma V(GS))\right]$$

$$V_1(S_1) = \frac{1}{4}\left[-2+0-1+0-1+0-1+0\right]$$

$$V_1(S_1) = \frac{-5}{4} = -1.25$$

$||^{ly}$ $V_1(S_3) = V_1(S_5) = V_1(S_7) = -1.25$

$$V_1(S_2) = \frac{1}{4}\left[-2-2-1-1\right] = \frac{-6}{4} = -1.5$$

$||^{ly}$ $V_1(S_6) = -1.5$

$$V_1(S_4) = \frac{1}{4}(-1-1-1-1) = \frac{-4}{4} = -1$$

$\therefore$ After step 1,

$$V_1(S_1) = V_1(S_3) = V_1(S_5) = V_1(S_7) = -1.25$$
$$V_1(S_2) = V_1(S_6) = -1.5$$
$$V_1(S_4) = -1$$

Step 2:

$$V_2(S_1) = \frac{1}{4}\left[-2+(-1.25) -1+(-1.5) -1+(-1) -1+(0)\right]$$

$$= \frac{1}{4}\left[-8.75\right] = -2.2$$

$$V_2(S_2) = \frac{1}{4}\left[-2+(-1.5) -2+(-1.5) -1+(-1.25) -1+(-1.25)\right]$$

$$= \frac{1}{4}\left[-11.5\right] = -2.875$$

$$V_2(S_4) = \frac{1}{4}[-1.25 -1 -1 -1.25 -1 -1.25 -1 -1.25]$$

$$= \frac{1}{4}[-9] = -2.25$$

$\therefore$ After step 2,

$V_2(S_1) = V_2(S_3) = V_2(S_5) = V_2(S_7) = -2.2$
$V_2(S_2) = V_2(S_6) = -2.875$
$V_2(S_4) = -2.25$

<u>Step 3</u>:

$$V_3(S_1) = \frac{1}{4}[-2 + (-2.2) -1 + (-2.875) -1 + (-2.25) -1 + (0)]$$

$$= \frac{1}{4}[-12.325] = -3.08$$

$$V_3(S_2) = \frac{1}{4}\Big[-2+(-2.875)-2+(-2.875)+ -1+(-2.2)-1+(-2.2)\Big]$$

$$= \frac{1}{4}\left[-16.15\right] = -4.04$$

$$V_3(S_4) = \frac{1}{4}\left[-1-2.2-1-2.2-1-2.2-1-2.2\right]$$

$$= \frac{1}{4}\left[-12.8\right] = -3.2$$

$\therefore$ After step 3,

$V_3(S_1) = V_3(S_3) = V_3(S_5) = V_3(S_7) = -3.08$

$V_3(S_2) = V_3(S_6) = -4.04$

$V_3(S_4) = -3.2$

b) Step-1

//////	-1.25	-1.5
-1.25	-1	-1.25
-1.5	-1.25	//////

//////	←	←↓
↑	←↑↓→	↓
↑→	→	//////

Step: 2

	-2.2	-2.875
-2.2	-2.25	-2.2
-2.875	-2.2	

	←	← ↓
↑	← ↑ ↓ →	↓
↑ →	→	

Step: 3

	-3.08	-4.04
-3.08	-3.2	-3.08
-4.04	-3.08	

	←	← ↓
↑	← ↑ ↓ →	↓
↑ →	→	

⑤ a) States : { Running, Broken }

Action (Running) : { Maintanance, No-maintanance }

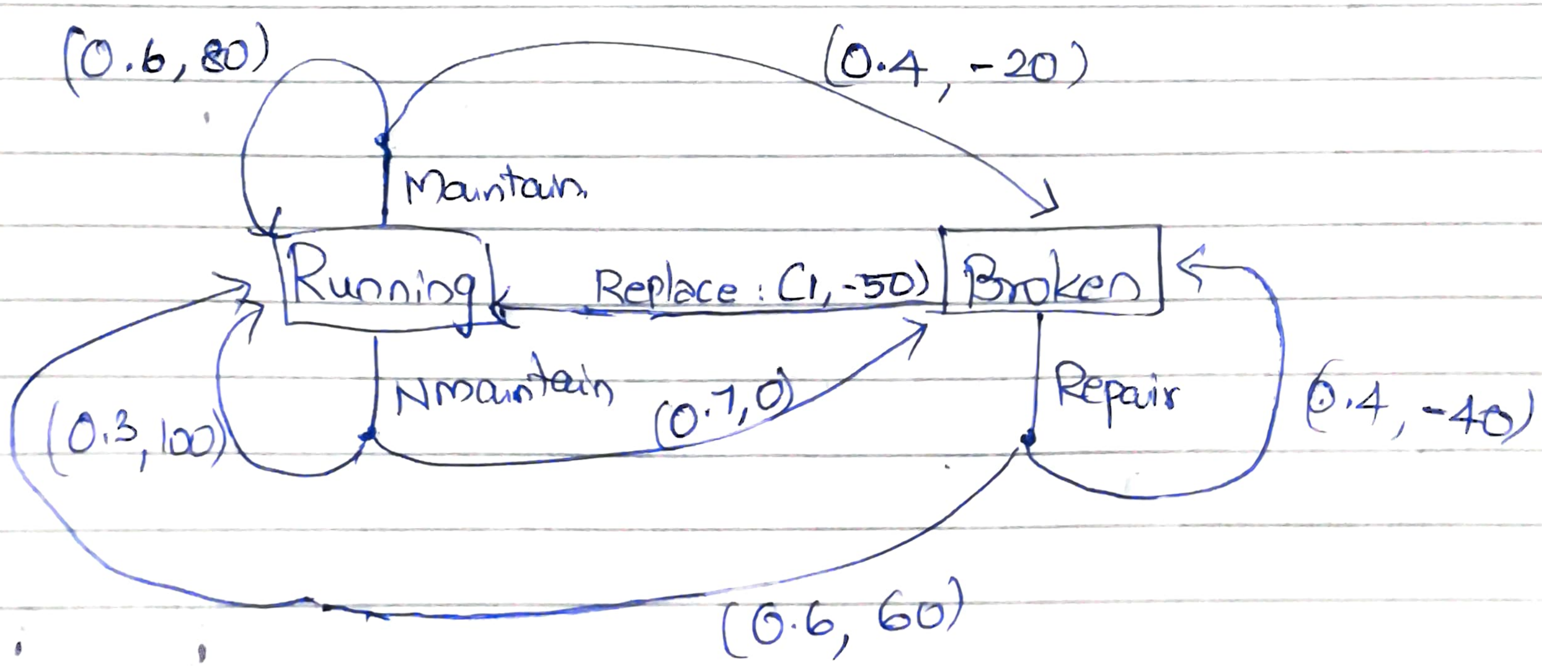
Action (Broken) : { Repair, Replace }

Costs (given) [- : Expenditure]

Running = \$100 ; Maintance = -\$20

Broken = \$0 ; Repair = -\$40 ; Replace = -\$150

Current State (s)	Next state (s')	Action (a)	Reward (r)	p(s',r/s,a)
Running	Running	Maintain	100-20=80	0.6
Running	Broken	Mantain	0-20=-20	0.4
Running	Running	N.Mantain	100	0.3
Running	Broken	N.Mantain	0	0.7
Broken	Broken	Repair	0-40=-40	0.4
Broken	Running	Repair	100-40=60	0.6
Broken	Running	Replace	100-150=-50	1



b) $\gamma < 1$

$V(\text{Running}) =$ Value from action Maintanance + Value from action N. maint.

(Equiprobable)

$$V(Run) = \frac{1}{2}\Big[0.6[80 + \gamma V(Run)] + 0.4[-20 + \gamma V(Broken)]\Big]$$

$$+ \frac{1}{2}\Big[0.3[100 + \gamma V(Run)] + 0.7[0 + \gamma V(Broken)]\Big]$$

$$2V(Run) = 48 + 0.6\gamma V(Run) - 8 + 0.4\gamma V(Broken) + 30 + 0.3\gamma V(Run) + 0.7\gamma V(Broken)$$

$$2V(Run) = 70 + 0.9\gamma V(Run) + 1.1\gamma V(Broken)$$

$$\boxed{(2 - 0.9\gamma) V(Run) - 1.1\gamma V(Broken) = 70} \quad - (1)$$

V(Broken) = Value from action Repair + Value from action Replace

(Equiprobable)

$$V(\text{Broken}) = \frac{1}{2}\left[0.6\left[60 + \gamma V(\text{Run})\right] + 0.4\left[-40 + \gamma V(\text{Broken})\right]\right]$$

$$+ \frac{1}{2}\left[1\left[-50 + \gamma V(\text{Run})\right]\right]$$

$$2V(\text{Broken}) = 36 + 0.6\gamma V(\text{Run}) - 16 + 0.4\gamma V(\text{Broken}) - 50 + \gamma V(\text{Run})$$

$$2V(\text{Broken}) = -30 + 1.6\gamma V(\text{Run}) + 0.4\gamma V(\text{Broken})$$

$$\boxed{(2 - 0.4\gamma)V(\text{Broken}) - 1.6\gamma V(\text{Run}) = -30} \quad \text{—} \ (2)$$

If $\gamma = 0.1$, Solving (1) & (2), $V(\text{Run}) = 35.94$

$V(\text{Broken}) = -12.37$