4:00
with lookup

Q1

a) $\boxed{Q(s,a) \leftarrow Q(s,a) + \alpha\left(r + \gamma \max_{a'} Q(s',a') - Q(s,a)\right)}$

b) $\boxed{Q_1(s,a) \leftarrow Q(s,a) + \alpha\left(r + \gamma Q'\left(s, \operatorname{argmax}_{a'} Q(s',a')\right) - Q(s,a)\right)}$

Q' is the second Q table

Q2

9:00
with some
other
things

a) $Q(1,1) = \dfrac{R(1,1)}{1-\gamma} = \boxed{10}$

$Q(1,2) = R(1,2) + \gamma Q(1,1) = \boxed{9}$

b) $\pi_\lambda(2|1) = \dfrac{e^{(9\lambda)}}{e^{(9\lambda)} + e^{10\lambda}} = \boxed{26\% \text{ of the time}}$

c) Since $a=2$ is not a greedy action

$\pi_\varepsilon(2|1) = \dfrac{\varepsilon}{|A|} = \boxed{5\% \text{ of the time}}$

~25:00

Q3 a)

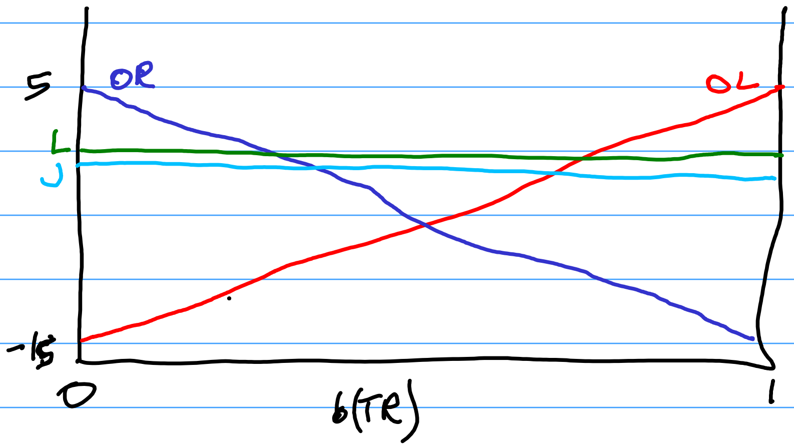
$\alpha_{OL} = [-15, 5]$

$\alpha_{OR} = [5, -15]$

$\alpha_{L} = [0, 0]$

$\alpha_{J} = [-1, -1]$

b)



c) $b = [0.4, 0.6]$

$b \cdot \alpha_{OL} \doteq -3$

$b \cdot \alpha_{OR} \doteq -7$

$b \cdot \alpha_{L} = 0$ ← best $b \cdot \alpha$ so $\boxed{\text{take } a = L}$

$b \cdot \alpha_{J} = -1$

d) $b'(s') \propto Z(O \mid a, s') \sum_s T(s' \mid s, a)\, b(s)$

$a = L, \quad o = TR$

$b'(TL) \propto Z(TR \mid L, TL)\left(T(TL \mid TL, L)\, b(TL) + T(TL \mid TR, L)\, b(TR)\right)$

$= (0.15)\ (1 \cdot 0.4 + 0 \cdot 0.6)$

$= 0.06$

$b'(TR) \propto Z(TR \mid L, TR)\left(T(TR \mid TL, L)\, b(TL) + T(TR \mid TR, L)\, b(TR)\right)$

$0.85 \quad (0 \cdot 0.4 \quad 1 \cdot 0.6)$

$= 0.51$

$$\boxed{b'(TR) = \frac{0.51}{0.51 + 0.06} = 0.895}$$

e) $b' = [0.105, 0.895]$

on the right half, α_{OL}, α_L dominate so we only need to consider them

$b' \cdot \alpha_{OL} = 2.89$ ← $\boxed{\text{take } a = OL}$

$b' \cdot \alpha_L = 0$

f) You would never take action $a = J$ because it is dominated over the entire belief space.

15:00

Q4 a)

path	d-separated by E?
$F \leftarrow A \rightarrow B$	No
$F \leftarrow A \rightarrow E \leftarrow B$	No

F and B are not d-separated by E

$\therefore F \perp B \mid E$ is $\boxed{\text{False}}$

b)

path	d-separated by A
$B \leftarrow A \leftarrow D$	Yes
$B \leftarrow A \leftarrow C \leftarrow D$	Yes
$B \rightarrow E \leftarrow A \leftarrow D$	Yes
$B \rightarrow E \leftarrow A \leftarrow C \leftarrow D$	Yes

B and D are d-separated by A

$\therefore B \perp D \mid A$ is $\boxed{\text{True}}$

c) From the first part of the statement, we have

$P(B=3 \mid D=1, A=2) = 1$

From part (b) we know $P(B \mid D, A) = P(B \mid A)$,

so $P(B=3 \mid A=2) = 1$

From the second part of the statement we have $P(B=1 \mid D=4, A=2) = 1$

Since $P(B \mid D, A) = P(B \mid A)$, we have $P(B=1 \mid A=2) = 1$ and therefore $P(B=3 \mid A=2) = 0$

This is a contradiction, so the statement is disproved.

Q5) a) 5:00

(1,1)	(1,1)
(1,1)	(1,1)

b)

(1,1)	(0,0)	(0,0)
(0,0)	(1,1)	(0,0)
(0,0)	(0,0)	(1,1)

c) There are 2 pure-strategy Nash Equilibria corresponding to the (3,3) and (10,10) entries

NE 1

$a^1 = b$
$a^2 = b$

NE 2

$a^1 = c$
$a^2 = c$