


MAS Final Exam. Solution sheet

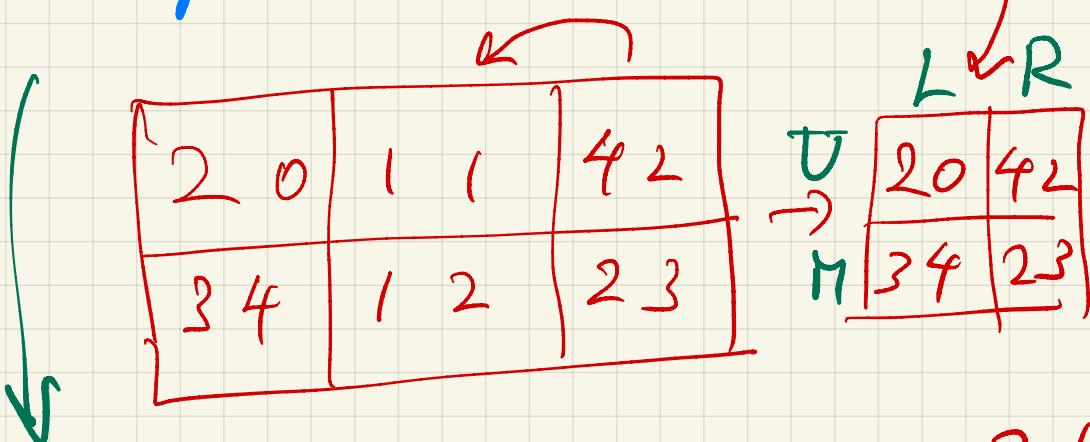
Question 1.



10/10

	L	C	R	
T	2 0	1 1	4, 2	
M	3 4	1 2	2 3	
D	1 3	0 2	3 0	

① Strategies that survive IESDS.



(T, M), (L, R)
Player 1 Player 2

2/2

1.2 Pure Strategy NE.

	L	R.
T	2, 0	4, 2
M	3, 4	2, 3

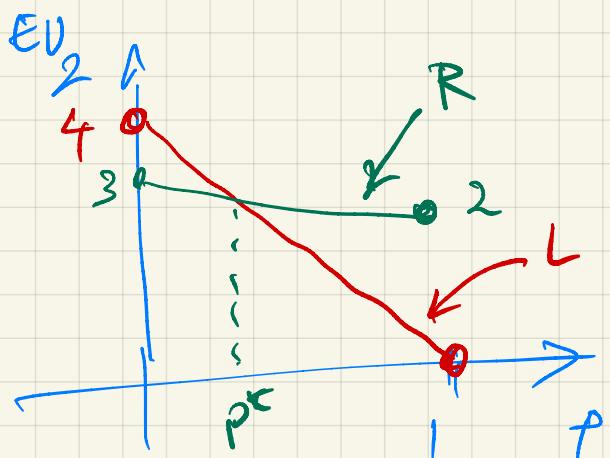
PNE: $(M, L) \rightarrow 3, 4$
 $(T, R) \rightarrow 4, 2.$

]
2/2

1.3: Mixed NE

		L (q)	R (1-q)
		2, 0	4, 2
		3, 4	2, 3
ϕ	T		
$1-p$	M		

]
3/3



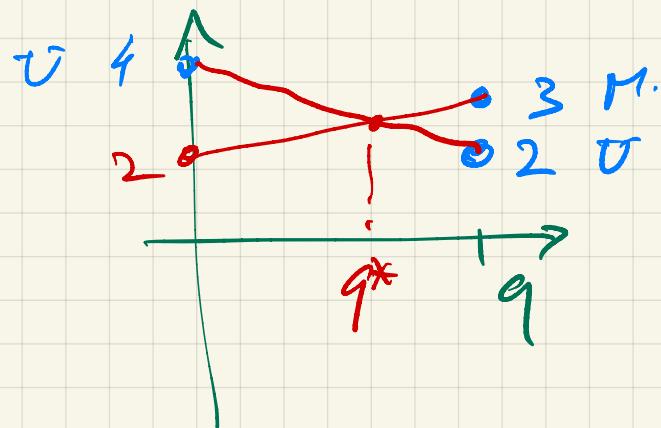
$$EU_2(p, L) = EU_2(p, R)$$

$$0.p + 4(1-p) = 2p + 3(1-p)$$

$$4 - 4p = 2p + 3 - 3p$$

$$3p = 1 \Rightarrow p^* = 1/3$$

$q^* = ?$



$$EU_1(U, q) = EO_1(M, q)$$

$$2q + 4(1-q) = 3q + 2(1-q)$$

$$2q + 4 - 4q = 3q + 2 - 2q = q + 2$$

$$-2q + 4 = q + 2$$

$$3q = 2 \Rightarrow q^* = \frac{2}{3}$$

4. Expected utilitites

$$MNE: p^* \rightarrow U, 1-p^* \rightarrow M$$

$$q^* \rightarrow L, 1-q^* \rightarrow R.$$

$$p^* = \frac{1}{3}, q^* = \frac{2}{3}$$

2/2

		q^*	$1-q^*$
p^*	2 0	4 2	
$1-p$	3 4	2 3	

		$2/3$	$1/3$
$p^*=1/3$	2/9	1/9	
$2/3$	4/9	2/9	

$$\begin{aligned}
 EU_1(p^*, q^*) &= \frac{1}{9} (2.2 + 1.4 + 3.4 + 2.2) \\
 &= \frac{1}{9} (24) = \frac{24}{9} = \frac{8}{3}
 \end{aligned}$$

$$\begin{aligned}
 EU_2(p^*, q^*) &= \frac{1}{9} (0.2 + 2.1 + 4.4 + 2.3) \\
 &= \frac{1}{9} (24) = \frac{8}{3}
 \end{aligned}$$

PNE: $(H, L) \rightarrow 3, 4.$ } utilities
 $(U, R) \rightarrow 4, 2$ } from
 table.

1.5: Prediction impossible
 because there are three NE.

(1/1)

2. Game Theory, Investment game \rightarrow 5/5

2.1 Majority situation:

2/2

Investors

Majority

Eg

$\begin{matrix} 20 \\ x \ x \ x \ x \end{matrix}$ 0
 $\underbrace{\quad\quad\quad}_{\text{invest}} \quad \underbrace{\quad\quad\quad}_{\text{no invest}}$

wants to change utility: $0 \rightarrow 20$.

$\underbrace{x \ x \ x \ x \ x}_{\text{still majority}}$

Stable if all invest

(NE)

$x \ x \ x \ x \ x \ x$
all invest.

Minority.

$\begin{matrix} -10 & 0 \\ x \ x \ x \ 0 \ 0 \ 0 \ 0 \end{matrix}$



wants to change utility: $-10 \rightarrow 0$

Only stable situation.

0 0 0 0 0 0 0

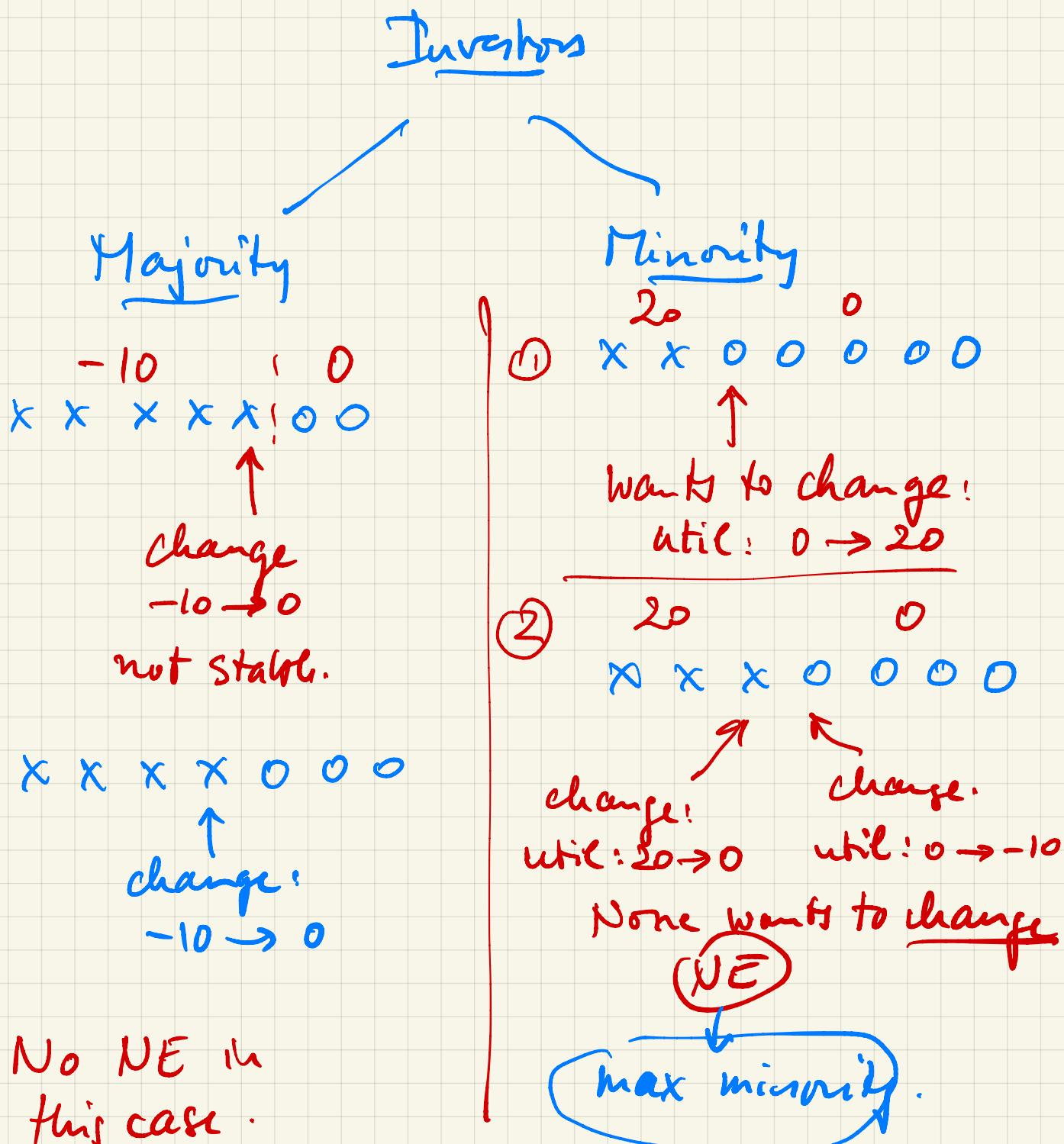
None invest

NE: no investors.

Conclusion: 2NE \rightarrow all none.

2.2 : Minority game $\rightarrow 3/3$

Inventors get pos. reward when in minority



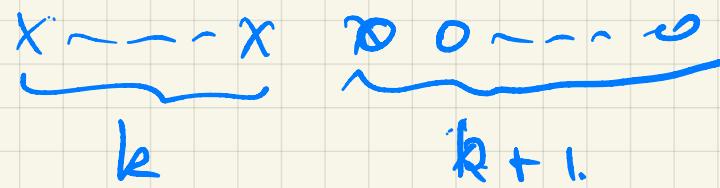
util: 0 → 20.

Conclusion

in minority game

NE: max minority

$$\text{I.e., } n = 2k+1 \rightarrow \text{max minority} \\ = k.$$



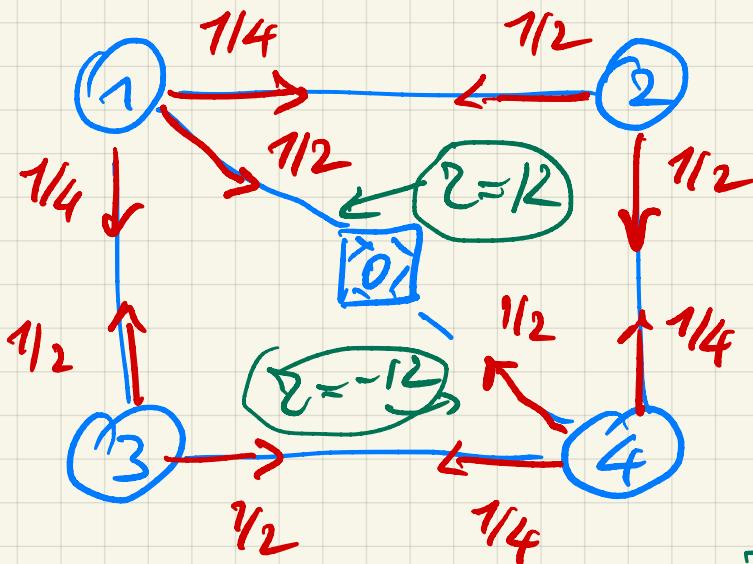
invest

not invest.

3. MDP

$(10/10)$

①



π : policy

T_0

$$P_{\pi} = \begin{bmatrix} & 0 & 1 & 2 & 3 & 4 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1/2 & 0 & 1/4 & 1/4 & 0 \\ 2 & 0 & 1/2 & 0 & 0 & 1/2 \\ 3 & 0 & 1/2 & 0 & 0 & 1/2 \\ 4 & 1/2 & 0 & 1/4 & 1/4 & 0 \end{bmatrix} \quad (2/2)$$

FROM \rightarrow

$$\mathcal{T}_{\pi} = \left(0, \frac{1}{2}(12-1), -1, -1, \frac{1}{2}(-12-1) \right) \quad (2/2)$$

0 1 2 3 4

2) Optimal v^*

$$\gamma = 2/3$$

2/2

optimal policy:
go to 0 via 1 asap.

Bellman optimality eq:

$$v^*(s) = \max_a (\tau(s, a, s') + \gamma v^*(s'))$$

$$v^*(0) = 0.$$

$$v^*(1) = 12$$

$$v^*(2) = -1 + \frac{2}{3} \cdot 12 = -1 + 8 = 7.$$

$$v^*(3) = 7.$$

$$v^*(4) = -1 + \frac{2}{3} \cdot 7 = \frac{11}{3}$$

Policy not unique since
we can choose in ④.

(2/1) ③ $q^*(1, a)$ where $1 \xrightarrow{a} 2$
 (assume: $\gamma = \frac{1}{3}$)

$$q^*(s, a) = r(s, a, s') + \gamma v^*(s')$$

$$= 2(s, a, s') + \gamma \max_{a'} q(s', a')$$

$$q^*(1, a) = r(1, a, 2) + \gamma v^*(2)$$

$$= -1 + \frac{2}{3} 7 = \frac{11}{3}$$

(4) Non-deterministic transitions

(3/3) $P_\pi(1, :)$ = $\left[\begin{array}{ccccc} \frac{3}{8} & \frac{1}{4} & \frac{3}{16} & \frac{3}{16} & 0 \end{array} \right]$

1 $\left[\begin{array}{ccccc} \frac{3}{4} & \frac{1}{2} & \frac{1}{4} & \frac{3}{4} \cdot \frac{1}{4} & \frac{3}{4} \cdot \frac{1}{4} & 0 \end{array} \right]$

0 1 2 3 4

$$z_{\pi}(1) = \frac{1}{4}(-2) + \frac{3}{4}\left(\frac{1}{2} \cdot 12 - \frac{1}{2} \cdot 1\right)$$

$$= -\frac{1}{2} + \frac{3}{4}\left(\frac{11}{2}\right)$$

$$= \frac{-4 + 33}{8} = \frac{29}{8}$$

$$z_{\pi}(2) = \frac{1}{4}(-2) + \frac{3}{4}(-1) = -\frac{5}{4}$$

1

4. RL & Exploration vs. Exploitation

4.1 Q-learning.

$s = \text{State}$	a	s'	r	$q(s,a)$	$v(s)$	$\pi(a s)$
2	R	3	-1	8	5	7/4
2	L	1	0	4	5	3/4
3	R	4	1	6	7	2/3
3	L	2	-2	9	7	1/3

Computing $v(s)$:

$$v(s) = \sum_a q(s,a) \pi(a|s)$$

① $5 = \frac{1}{4} q(2,R) + \frac{3}{4} \cdot 4 = \frac{1}{4} q(2,R) + 3$

$$q(2,R) = 8$$

$$11/3 = 3.67 \leftarrow \text{alternative (see next page)}$$

② $7 = 6 \cdot \frac{2}{3} + q(3,L) \cdot \frac{1}{3} \Rightarrow \underbrace{(7-4)}_3 \cdot 3 = q(3,L)$

$$q(3,L) = 9$$

$$1.33 = 4/3$$

Alternative computation of $q(2, R)$
 $q(3, L)$

Deterministic transitions:

$$s \xrightarrow{a} s'$$

$$q(s,a) = r(s,a,s') + \gamma v(s')$$

Hence

$$2 \xrightarrow{R} 3$$

$$\textcircled{1} \quad q(2, R) = -1 + \frac{2}{3} v(3) = -1 + \frac{2}{3} \cdot 7 = \frac{11}{3} \\ = 3.67.$$

$$3 \xrightarrow{L} 2$$

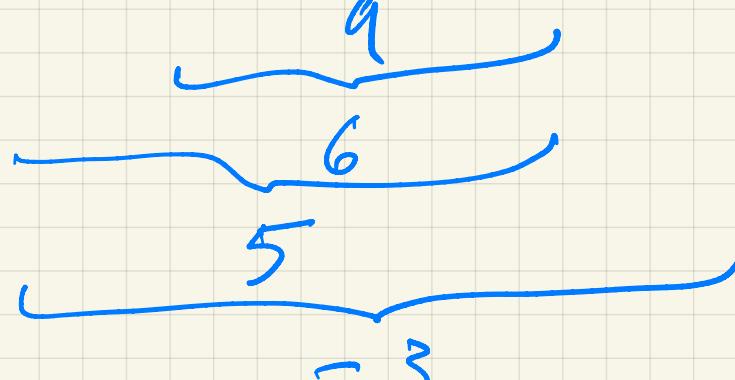
$$\textcircled{2} \quad q(3, L) = -2 + \frac{2}{3} v(2) = -2 + \frac{2}{3} \cdot 5 \\ = \frac{4}{3} = 1.33$$

$2 \rightarrow 3$

$$\textcircled{1} \quad q_{\pi}(2, R) \leftarrow q_{\pi}(2, R) + \alpha [r + \gamma \max_{a'} q(3, a') - q(2, R)]$$

$$8 + 0.9 \left[-1 + \frac{2}{3} \max(6, 9) - 8 \right]$$

$\textcircled{3}/3$



Q-values
+ update

$$= 8 - 3 \cdot (0.9) = 8 - 2.7 = 5.3$$

$$\text{Alternative} = 3.67 + 0.9 \left[-1 + \frac{2}{3} \max(6, 1.33) - 3.67 \right] = 2.7 + 0.37 = \underline{\underline{3.07}}$$

(2) SARSA

2/2

$s \xrightarrow{a} s' \xrightarrow{a'} \dots$

need to know a' !

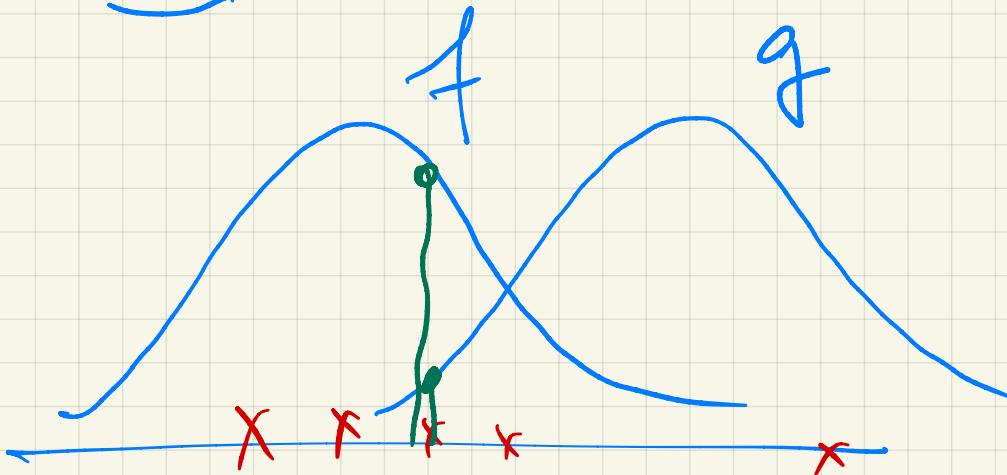
(3) see theory.

2/2

4.2

3/3

KL.



$X_i \sim f \rightarrow$ more sample points
near high f -values

for those, $\frac{f(x_i)}{g(x_i)} > 1$

$$\Rightarrow \log \frac{f(x_i)}{g(x_i)} > 0$$