20.00

1) $D \in \{0,1\}$ detection
$B \in \{0,1\}$ balloon

a). want $P(D=0 \mid B=1)$

- know that $\sum_{d \in \{0,1\}} P(D=d \mid B=1) = 1$

$\therefore P(D=0 \mid B=1) = 1 - P(D=1 \mid B=1) = 1 - 0.63 = \boxed{0.37}$

b) - want $P(B=1 \mid D=0)$

- know $P(D \mid B)$
- Bayes rule

$$P(B \mid D) = \frac{P(D \mid B)\, P(B)}{P(D)}$$

- If we know $P(B)$, we can calculate $P(D)$ with $P(D) = \sum_{b} P(D \mid B=b) P(B=b)$.

$\therefore$ We need to know the marginal distribution of a balloon passing over, $P(B)$ to calculate $P(B \mid D)$

c) redefine $B \in \{0, 1, 2\}$ ← surveillance; 1 ↑ weather; 0 ↗ no balloon

B	P(B)
0	0.34
1	0.47
2	0.19

want $P(B=2 \mid D=1)$

$$P(B|D) = \frac{P(D|B)\,P(B)}{P(D)}$$

$$P(D=1) = \sum_b P(D=1 \mid B=b)\,P(B=b)$$

$$= 0.07 \cdot 0.34 + 0.63 \cdot 0.47 + 0.63 \cdot 0.19$$

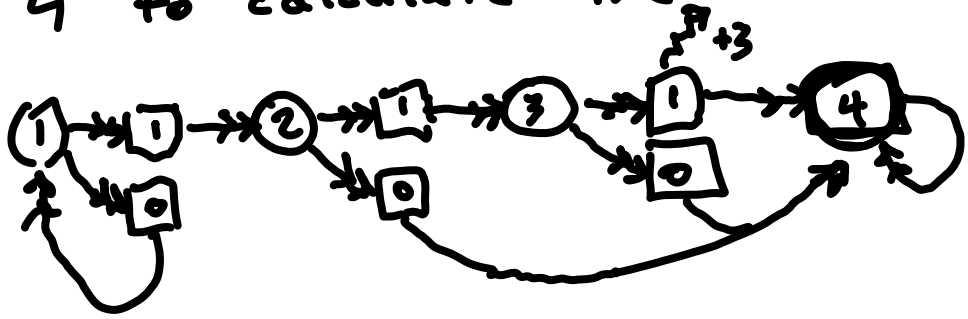
$$= 0.4396$$

$$P(B=2|D=1) = \frac{P(D=1|B=2)\,P(B=2)}{P(D=1)}$$

$$= \frac{0.63 \cdot 0.19}{0.4397} = \boxed{0.272}$$

12:00

2) State 4 is a terminal state since there will be no transitions away from it and it has no reward. We can work backwards from 4 to calculate the value.



s	$V^*(s)$	$\pi^*(s)$	a	$Q^*(s,a)$
4	0	0	0	$0 + \gamma V^*(4)$
3	3	1	1	$3 + \gamma V^*(4) = 3$
2	2.7	1	0	$0 + \gamma V^*(4) = 0$
			1	$0 + \gamma V^*(3) = 0.9 \cdot 3 = 2.7$
1	2.43	1	0	$0 + \gamma V^{\pi_1}(1) = 0$
			1	$0 + \gamma V^*(2) = 0.9 \cdot 2.7 = 2.43$

optimal value

optimal policy

To prove that this is 0, let π_1 be a policy where $\pi_1(1) = 1$ and $\pi_1(s)$ is the previously-calculated policy for all other states. Then we have $V^{\pi_1}(1) = 0 + \gamma V^{\pi_1}(1) \therefore V^{\pi_1}(1) = 0$.

3] (there are many correct answers) 10:00

$S = \{1, 2, g, e\}$ — 1: player 1 has ball; 2: player 2 has ball; g: goal; e: end: defender has ball or goal has been scored

$A = \{\text{pass}, \text{shoot}\}$

$$R(s,a) = \begin{cases} 1 & \text{if } s = g \\ 0 & \text{o.w.} \end{cases}$$

$$T^{\text{pass}} = \begin{array}{c} \\ 1 \\ 2 \\ g \\ e \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & g & e \end{array} \\ \begin{bmatrix} 0 & 0.85 & 0 & 0.15 \\ 0.8 & 0 & 0 & 0.2 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \end{array}$$

$$T^{\text{shoot}} = \begin{array}{c} \\ 1 \\ 2 \\ g \\ e \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & g & e \end{array} \\ \begin{bmatrix} 0 & 0 & 0.35 & 0 \\ 0 & 0 & 0.45 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \end{array}$$

$\gamma = 0.9$ (because there is a 10% chance of the game ending on every step)

5:00

4) Steps:

1. Create transition matrix, T^π, and reward vector R^π for the policy.

2. Evaluate the policy to find the policy value function:

$$V^\pi = (I - \gamma T^\pi)^{-1} R^\pi$$

3. Check if this value function satisfies Bellman's optimality equation

$$V^\pi(s) \stackrel{?}{=} \max_a \left(R(s,a) + \gamma \sum_{s'} T(s'|s,a) V^\pi(s') \right)$$

If this is satisfied for all s, V^π is the optimal value function and therefore π is an optimal policy.

5) $R(s,a) = -s^2 - a^2$ will have the desired optimal policy. This is because the problem fits the form of a Linear Quadratic Regulator (LQR) problem. Specifically, the reward has the form

$R(s,a) = s^T R_s s + a^T R_a a$ where R_s is negative semi-definite and R_a is negative definite.