

1. Consider a weather monitoring station. Suppose that we cannot observe the weather directly, but instead rely on a sensor. The problem is that our sensor is noisy. Its measurements are governed by Figure 1.
- (a) Suppose Day-1 is sunny (this is already known from sensor data), and in the subsequent two days our sensor observes *cloudy*, *sunny*. What is the probability that Day-3 is indeed sunny as predicted by our sensor ? What is the probability of the most likely sequence of weather for Days 2 through 3 ? 4+1
- (b) Write mathematical expressions for – i) state transition prob and measurement prob without Markov assumption, ii) state transition prob and measurement prob with Markov assumption ? 1.5 + 1.5

		our sensor tells us...		
		sunny	cloudy	rainy
the actual weather is...	sunny	.6	.4	0
	cloudy	.3	.7	0
	rainy	0	0	1

Figure 1:

Solution:

Look into the next page for solution 1(a) and at the end of this document for solution 1(b).

Quiz-2

[Rubric]

1.(a) prior dist. based on Day-1 is actually sunny.

$$\left. \begin{array}{l} \text{bel}(x_1 = \text{sunny}) = 0.6 \\ \text{bel}(x_1 = \text{cloudy}) = 0.4 \end{array} \right\} \text{prior}$$

$$\text{bel}(x_1 = \text{rainy}) = 0$$

In this problem, the sensor only observes, and it does not act.

As per Bayes filter, in this problem,

① $\text{bel}(x_1) = \text{bel}(x_2)$ as there is no action. ①

So, $\text{bel}(x_2 = \text{sunny}) = 0.6$ $\text{bel}(x_2 = \text{rainy}) = 0.0$
 $\text{bel}(x_2 = \text{cloudy}) = 0.4$

Day-2

$$\text{bel}(x_2 = \text{sunny}) = \eta \cdot p(z_2 = \text{cloudy} | x_2 = \text{sunny}) \cdot \text{bel}(x_2 = \text{sunny})$$

$$= \eta \cdot 0.4 \times 0.6 = 0.24 \eta$$

$$\text{bel}(x_2 = \text{cloudy}) = 0.28 \eta$$

$$\text{bel}(x_2 = \text{rainy}) = 0 \times \eta = 0$$

$$0.28 \eta + 0.24 \eta + 0 = 1$$

$$\Rightarrow \eta = 1.9$$

$$\text{bel}(x_2 = \text{sunny}) = 0.46; \text{bel}(x_2 = \text{cloudy}) = 0.54$$

$$\text{bel}(x_2 = \text{rainy}) = 0.$$

Day-3

[Rubric]

Similarly,

$$\begin{aligned}\text{bel}(x_3 = \text{sunny}) &= \eta \cdot p(z_3 = \text{sunny} | x_3 = \text{sunny}) \cdot \bar{\text{bel}}(x_3 = \text{sunny}) \\ &= \eta \times 0.6 \times 0.46 = 0.28 \eta\end{aligned}$$

$$\begin{aligned}\text{bel}(x_3 = \text{cloudy}) &= \eta \cdot p(z_3 = \text{sunny} | x_3 = \text{cloudy}) \cdot \bar{\text{bel}}(x_3 = \text{cloudy}) \\ &= \eta \times 0.3 \times 0.54 = 0.16 \eta\end{aligned}$$

$$\text{bel}(x_3 = \text{rainy}) = 0$$

$$0.28 \eta + 0.16 \eta + 0 = 1$$

$$\Rightarrow \eta = 2.3$$

Therefore,

$$\text{bel}(x_3 = \text{sunny}) = 0.28 \times 2.3 = 0.64 \quad (\text{Answer})$$

$$\text{bel}(x_3 = \text{cloudy}) = 0.16 \times 2.3 = 0.36$$

$$\text{bel}(x_3 = \text{rainy}) = 0.$$

$$\begin{aligned}\text{prob of the sequence (cloudy, sunny)} &= 0.54 \times 0.64 \\ &= 0.38 \quad (\text{Answer}).\end{aligned}$$

2. Answer the following:

- (a) Write an algorithm for computing value iteration in an MDP (in a discrete setting). 3
- (b) Write an algorithm for computing the optimal policy in an MDP. 2

Solution:

(a) Refer to **Algorithm MDP-discrete-value-iteration**

- The formula for $\hat{V}(x)$ (line 7) carries 2 marks. It needs to be correct. Small error in the formula will attract a deduction of 1-2 marks.
- Check if the notion of *time* (line 5) and *all discrete points* (line 6) are written. 1 mark for this.

```

1:  Algorithm MDP_discrete_value_iteration( ):
2:      for  $i = 1$  to  $N$  do
3:           $\hat{V}(x_i) = r_{\min}$ 
4:      endfor
5:      repeat until convergence
6:          for  $i = 1$  to  $N$  do
7:              
$$\hat{V}(x_i) = \gamma \max_u \left[ r(x_i, u) + \sum_{j=1}^N \hat{V}(x_j) p(x_j | u, x_i) \right]$$

8:          endfor
9:      endrepeat
10:     return  $\hat{V}$ 

```

```

1:  Algorithm policy_MDP( $x, \hat{V}$ ):
2:      return  $\operatorname{argmax}_u \left[ r(x, u) + \sum_{j=1}^N \hat{V}(x_j) p(x_j | u, x_i) \right]$ 

```

Table 14.1 The value iteration algorithm for MDPs, stated here in its most general form and for MDPs with finite state and control spaces. The bottom algorithm computes the best control action.

- (b) 1.5 marks for the mathematical expression in line 2 of **Algorithm policy-MDP**. 0.5 marks for the parameter.

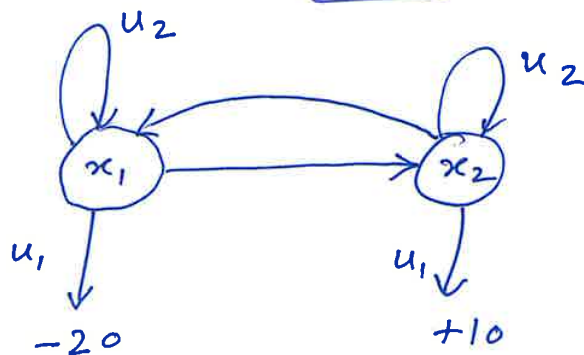
3. Consider the following POMDP. A person probes two doors. Behind one is a tiger, behind the other a reward of $+10$. The person can either *listen* or *open* one of the doors. When opening the door with a tiger, the person will be attacked, which has an associated cost of -20 . Listening costs -1 . When listening, the person will hear a roaring noise that indicates the presence of the tiger, but only with 0.85 probability will the person be able to localize the noise correctly. With 0.15 probability, the noise will appear as if it came from the door hiding the reward.
- (a) Provide the formal model of the POMDP, in which you define the state, action, and measurement spaces, the cost function, and the associated probability functions. Give a schematic diagram for better understanding. 5
- (b) For the above POMDP, write the mathematical expression for calculating the expected payoff for a given belief b and given action u . 2

Solution:

[Look into the next page.](#)

Answer 3 (a)

[Rubric]



* 3 marks for the correct state, action, measurement and cost functions.

* partial marks for the above.

* 2 marks for obs. model.
- No partial marking.

POMDP

States: $\{x_1 = \text{tiger}, x_2 = \text{reward}\}$

Actions: $\{u_1 = \text{open door}, u_2 = \text{listen}\}$

Measurements: $\{z_1 = \text{sense-roar}, z_2 = \text{no roar}\}$

cost function: $r(x_1, u_1) = -20$

$r(x_2, u_1) = +10$

$r(x_1, u_2) = r(x_2, u_2) = -1$

observation model:

$[z_1 = \text{sense-roar}, z_2 = \text{no-roar}]$

$p(z_1 | x_1) = 0.85$, Therefore $p(z_2 | x_1) = 0.15$

$p(z_1 | x_2) = 0.15$, Therefore $p(z_2 | x_2) = 0.85$

Transition model:

Give marks Don't deduct marks for arbitrarily chosen

$p(x'_1 | x_1, u_2)$, $p(x'_2 | x_1, u_2)$, $p(x'_1 | x_2, u_2)$

and $p(x'_2 | x_2, u_2)$.

Quiz-2 [Rubric]

3.(b)

Belief $b = (p_1, p_2)$ as there are only two states.

$$r(b, u) = E_x [r(x, u)]$$

$$= p_1 \cdot r(x_1, u) + p_2 \cdot r(x_2, u)$$

1.(b) (i) Without Markov assumption:

- State trans prob: $p(x_t | \underline{x_{0:t-1}}, z_{1:t-1}, u_{1:t})$

- Measurement prob: $p(z_t | \underline{x_{0:t}}, z_{1:t-1}, u_{1:t})$

* No marks if the state sequence is incorrect

(ii) With Markov assumption:

- State trans. prob: $p(\underline{x_t} | x_{t-1}, u_t)$

- Measurement prob: $p(\underline{z_t} | x_t)$

* No marks if the previous and current-state index (or observation and state index is incorrect).