Exercise Sheet - POMDPs

This exercise concerns a POMDP where the underlying states form a chain on which the agent can walk left or right. To help conceptualize, the underlying Markov chain is shown in Figure 1:

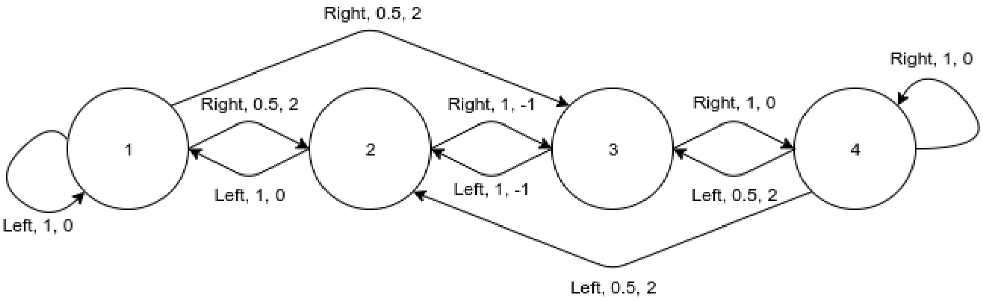

Figure 1: Underlying problem. The values along the arrow denote the action, the probability of the transition, and the immediate reward.

Formally, the POMDP can be described as folllows:

- $\mathcal{S} = \{1, 2, 3, 4\}$
- $\mathcal{A} = \{Left, Right\}$
- $\mathcal{O} = \{Green, Blue\}$
- Transitions:

```
%Transitions: P(s'|s,a) = T{a}(from, to)
T{Left} = [
[1,  0,  0,  0 ] %from S1
[1,  0,  0,  0 ] %from S2
[0,  1,  0,  0 ] %from S3
[0, .5, .5,  0 ] %from S4
];
T{Right} = [
[0, .5, .5,  0 ] %from S1
[0,  0,  1,  0 ] %from S2
[0,  0,  0,  1 ] %from S3
[0,  0,  0,  1 ] %from S4
];
```

- Observation probabilities:

```
%P(o|s') = O(s',o)
%i.e., the first row specifies the probabilities [ P(Green|S1), P(Blue|S1) ]
O = [
[ 1,  0] %to S1
[.5, .5],%to S2
[.5, .5],%to S3
[ 0,  1] %to S4
];
```

- Rewards:

```
%R(from, a)
%i.e., the first row specifies [ R(S1,Left), R(S1,Right) ]
R = [
[0,  2],
[0, -1],
[-1, 0],
[2,  0]
]
```

- $b_0 = (1, 0, 0, 0)$ is the initial belief. (I.e., we know we start in state 1)

Given this POMDP....

1. Compute the tree of all reachable beliefs by taking 2 actions.
2. For all these beliefs compute the expected immediate reward for taking action $Left$ or $Right$.
3. Now, perform backwards induction to compute $V^{\tau=2}(b_0)$ (the value of the initial belief for two timesteps to go).