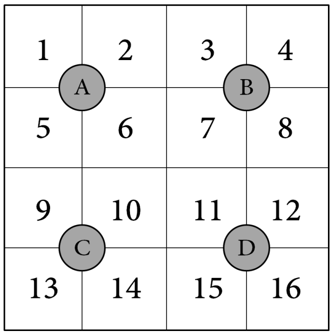
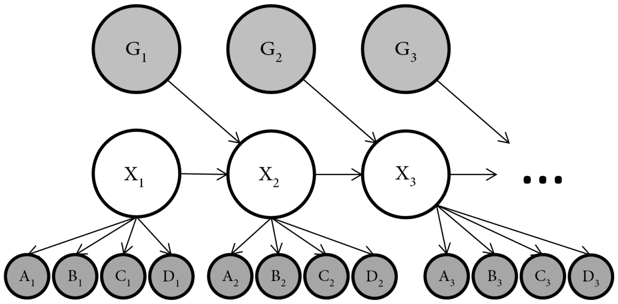
CS 6300 **HW10: Particle Filters and POMDPs** **Due April 26, 2022**

Please use the LaTeX template to produce your writeups. Hand in via gradescope.

1 Particle Filtering

A garbage-collecting robot lives in a 4x4 Manhattan grid city. The associated HMM includes the robot position X, the readings G from a garbage sensor, and (A,B,C,D) readings from the motion sensors.





The garbage sensor G takes on an integer value between 1 and 16, corresponding to the square with the most garbage at time t. The robot is programmed to move toward the square with the most garbage, but it will only take an optimal action with probability 0.9. In each time step, the robot can either stay in the same square, or move to an adjacent square. In case where multiple actions would move it equally close to the desired position, the robot has an equal probability of taking any of these actions. In case the robot fails to take an optimal action, it has an equal probability of taking any of the non-optimal actions. For example, if the robot is in square 2, the actions available are (EAST, SOUTH, WEST, STOP). If $G_t = 15$, the transition model will look like this:

X_{t+1}	$P(X_{t+1}\|X_t = 2, G_t = 15)$
1	0.05
2	0.05
3	0.45
6	0.45

The motion sensors, (A, B, C, D), take on a value of ON or OFF. At time t, the sensor adjacent to the square that the robot is on always outputs ON. Otherwise, the sensor will output ON or OFF with equal probability. For example, the sensor tables would look like this if $X = 6$:

A	$P(A\|X = 6)$
ON	1
OFF	0

B	$P(B\|X = 6)$
ON	0.5
OFF	0.5

C	$P(C\|X = 6)$
ON	0.5
OFF	0.5

D	$P(D\|X = 6)$
ON	0.5
OFF	0.5

1. Initially, at $t = 1$, there are particles $[X = 4, X = 2, X = 15]$. We observe that $G_1 = 6$. Use the following random numbers to apply the time update to each of the particles. Please assign square numbers to sample spaces in numerical order.

   [0.7349, 0.5324, 0.1670]

   The transition probabilities for each particle are:

X_2	$P(X_2 \mid X_1 = 4, G_1 = 6)$
3	0.45
4	0.10
8	0.45

X_2	$P(X_2 \mid X_1 = 2, G_1 = 6)$
1	0.033
2	0.033
3	0.033
6	0.900

X_2	$P(X_2 \mid X_1 = 15, G_1 = 6)$
11	0.45
14	0.45
15	0.05
16	0.05

   For particle $X_1 = 4$, the random number 0.7349 maps to $X_2 = 8$.

   For particle $X_1 = 2$, the random number 0.5324 maps to $X_2 = 6$.

   For particle $X_1 = 15$, the random number 0.1670 maps to $X_2 = 11$.

Particle at t=1	Particle after time update
$X = 4$	8
$X = 2$	6
$X = 15$	11

2. To decouple this question from the previous question, let's say the new particles after the time update are $[X = 8, X = 14, X = 11]$. The sensors read $[A = OFF, B = ON, C = ON, D = OFF]$.

   (a) What is the weight for each particle? Show your derivations.

Particle	Weight
$X = 8$	P(A=OFF|X=8)P(B=ON|X=8)P(C=ON|X=8)P(D=OFF|X=8) = $(0.5)(1)(0.5)(0.5) = 0.125$
$X = 14$	P(A=OFF|X=14)P(B=ON|X=14)P(C=ON|X=14)P(D=OFF|X=14) = $(0.5)(0.5)(1)(0.5) = 0.125$
$X = 11$	P(A=OFF|X=11)P(B=ON|X=11)P(C=ON|X=11)P(D=OFF|X=11) = $(0.5)(0.5)(0.5)(0) = 0$

   (b) It seems sensor C is broken, and will always give a reading of ON. Recalculate the weights with this new knowledge, showing your derivations.

Particle	Weight
$X = 8$	P(A=OFF|X=8)P(B=ON|X=8)P(C=ON|X=8)P(D=OFF|X=8) = $(0.5)(1)(1)(0.5) = 0.25$
$X = 14$	P(A=OFF|X=14)P(B=ON|X=14)P(C=ON|X=14)P(D=OFF|X=14) = $(0.5)(0.5)(1)(0.5) = 0.125$
$X = 11$	P(A=OFF|X=11)P(B=ON|X=11)P(C=ON|X=11)P(D=OFF|X=11) = $(0.5)(0.5)(1)(0) = 0$

2 POMDPS

An agent is in one of the two cells s_1, s_2. There are two actions $a \in \{go, stay\}$: the agent can either stay in the cell, or attempt to go to the other cell. The transition probabilities $T(s_i, a, s_j)$ (take action a from state s_i and arrive in state s_j) are:

$$T(s_i, stay, s_j) = \begin{cases} 3/4 & \text{for} \quad i \neq j \\ 1/4 & \text{for} \quad i = j \end{cases}$$

$$T(s_i, go, s_j) = \begin{cases} 1/3 & \text{for} \quad i \neq j \\ 2/3 & \text{for} \quad i = j \end{cases}$$

The reward function has the simplified form $R(s_i, a, s_j) = R(s_j)$, i.e., it depends only on the state you end up in. There is a reward for transitioning to state s_2, but none to state s_1:

$$R(s_2) = 1, \quad R(s_1) = 0$$

The agent has an ultrasound sensor which helps to distinguish which cell it's in. There are two possible readings z_1 or z_2 corresponding to an estimation of being in cell s_1 or s_2 respectively, but the sensor is noisy and sometimes gives the wrong reading. Its conditional probability is given by:

$$P(z_i | s_j) = \begin{cases} 0.2 & \text{for} \quad i \neq j \\ 0.8 & \text{for} \quad i = j \end{cases}$$

The agent maintains and updates a belief function $b(s_i)$ based upon combinations of actions and associated sensor readings. For brevity, define $p_1 = b(s_1)$. Hence $b(s_2) = 1 - p_1$.

1. For the first action and without receiving any sensor readings yet, derive the one-time-step utilities $V^{stay}(s_i)$ and $V^{go}(s_i)$, $i = 1, 2$, for actions $stay$ and go.

   A one-step policy evaluation involves solving a simplified form of Bellman's equation:

   $$V^a(s_i) = r(s_i, a) = \sum_j T(s_i, a, s_j) R(s_i, a, s_j)$$

   For action $a = stay$,

   $$\begin{aligned} V^{stay}(s_1) &= T(s_1, stay, s_1) R(s_1, stay, s_1) + T(s_1, stay, s_2) R(s_1, stay, s_2) \\ &= 1/4 \cdot 0 + 3/4 \cdot 1 = 3/4 \\ V^{stay}(s_2) &= T(s_2, stay, s_1) R(s_2, stay, s_1) + T(s_2, stay, s_2) R(s_2, stay, s_2) \\ &= 3/4 \cdot 0 + 1/4 \cdot 1 = 1/4 \end{aligned}$$

   For action $a = go$,

$$\begin{aligned} V^{go}(s_1) &= T(s_1, go, s_1)R(s_1, go, s_1) + T(s_1, go, s_2)R(s_1, go, s_2) \\ &= 2/3 \cdot 0 + 1/3 \cdot 1 = 1/3 \\ V^{go}(s_2) &= T(s_2, go, s_1)R(s_2, go, s_1) + T(s_2, go, s_2)R(s_2, go, s_2) \\ &= 1/3 \cdot 0 + 2/3 \cdot 1 = 2/3 \end{aligned}$$

2. You don't actually know which state you're in, and you have to use your belief function $b(s_i)$ to combine the results above. Find the expected reward $V^{go}(b)$ for action go, and $V^{stay}(b)$ for action $stay$.

   The expected reward is the weighted sum of the two possible outcomes for each action.

$$V^a(b) = E_i[V^a(s_i)]$$

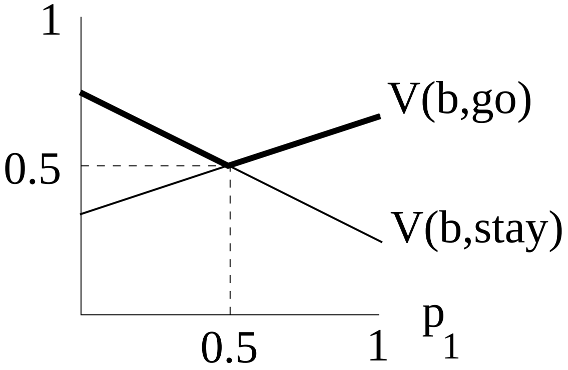
   The expectation operator means weighting by $b(s_i)$. Using the results above, for action $a = stay$,

$$\begin{aligned} V^{stay}(b) &= b(s_1)V^{stay}(s_1) + b(s_2)V^{stay}(s_2) \\ &= p_1 V^{stay}(s_1) + (1 - p_1)V^{stay}(s_2) \\ &= p_1 \cdot 3/4 + (1 - p_1) \cdot 1/4 \\ &= 1/4 + p_1/2 \end{aligned}$$

   For action $a = go$,

$$\begin{aligned} V^{go}(b) &= b(s_1)V^{go}(s_1) + b(s_2)V^{go}(s_2) \\ &= p_1 V^{go}(s_1) + (1 - p_1)V^{go}(s_2) \\ &= p_1 \cdot 1/3 + (1 - p_1) \cdot 2/3 \\ &= 2/3 - p_1/3 \end{aligned}$$

3. Plot both expected reward functions on the same plot with p_1 on the x-axis. Identify the optimal strategy based on your plot.



The optimal strategy is the outer envelope, where $stay$ is the optimal strategy from 0 to 0.5, and go thereafter.

4. Suppose you are able to get a sensor reading before taking an action, and you observe z_1. Update your belief to find $p(s_1|z_1)$ and $p(s_2|z_1)$.

We use Bayes' rule to update the belief function. First, the probability of the evidence is:

$$p(z_1) = p(z_1|s_1)p(s_1) + p(z_1|s_2)p(s_2) = 0.8p_1 + 0.2(1 - p_1) = 0.6p_1 + 0.2$$

Then

$$p(s_1|z_1) = \frac{p(z_1|s_1)p(s_1)}{p(z_1)} = \frac{0.8p_1}{0.6p_1 + 0.2}$$

$$p(s_2|z_1) = \frac{p(z_1|s_2)p(s_2)}{p(z_1)} = \frac{0.2(1 - p_1)}{0.6p_1 + 0.2}$$

5. Solve for the new value functions given b'.

The expectation operator now uses the new belief function b'.

$$\begin{aligned} V^{stay}(b') &= b'(s_1)V^{stay}(s_1) + b'(s_2)V^{stay}(s_2) \\ &= p(s_1|z_1)V^{stay}(s_1) + p(s_2|z_1)V^{stay}(s_2) \\ &= \frac{0.8p_1}{0.6p_1 + 0.2} \cdot 3/4 + \frac{0.2(1 - p_1)}{0.6p_1 + 0.2} \cdot 1/4 \\ &= \frac{0.1 + 1.1p_1}{0.4 + 1.2p_1} \end{aligned}$$

For action $a = go$,

$$\begin{aligned}
V^{go}(b') &= b'(s_1)V^{go}(s_1) + b'(s_2)V^{go}(s_2) \\
&= p(s_1|z_1)V^{go}(s_1) + p(s_2|z_1)V^{go}(s_2) \\
&= \frac{0.8p_1}{0.6p_1 + 0.2} \cdot 1/3 + \frac{0.2(1-p_1)}{0.6p_1 + 0.2} \cdot 2/3 \\
&= \frac{0.4}{3}\frac{1+p_1}{0.6p_1 + 0.2}
\end{aligned}$$