CS 747 (Spring 2025) Week 8 Test (Batch 1)

5.35 p.m. – 6.00 p.m., March 13, 2025, LA 001

Name: ______________________________ Roll number: ______________

Note. There is one question in this test. You can use the space on both pages for your answer. Draw a line (either vertical or horizontal) and do all your rough work on one side of it.

Question 1. Let $\mathcal{D}$ be the set of all *deterministic* MDPs having 2 or more states and 2 or more actions at each state. In a deterministic MDP, every state-action pair has a fixed next state: that is, all transition probabilities are either exactly 0 or exactly 1. Assume that the MDPs in $\mathcal{D}$ are all continuing tasks (that is, they have no terminal states).

Does there exist $M \in \mathcal{D}$ such that M is ergodic? If you claim yes, describe an such an MDP M. If you claim no, you must prove that every MDP $M \in \mathcal{D}$ is non-ergodic. [3 marks]

Answer 1.

Consider any MDP $M \in \mathcal{D}$; suppose that M has n states. Recall that for M to be ergodic, it must be aperiodic as well as irreducible under every policy. We show that on the other hand, M cannot be both aperiodic and irreducible under *any* policy. Indeed fix an arbitrary policy π for M. Consider the directed graph G_π obtained by fixing actions according to π. The vertices of G_π are the states of M. There is a directed edge from vertex s to vertex s' if and only if $T(s, \pi(s), s') = 1$, where T is the transition function of M. Observe that G_π has exactly one outgoing edge from each vertex. We furnish two separate arguments based on G_π to show that M cannot be simultaneously aperiodic and irreducible for π.

1. Since G_π has n vertices and n directed edges, with one outgoing edge from each vertex, it must contain a directed cycle. If even one directed cycle in G_π is of length 2 (edges) or more, then any state on that cycle repeats periodically with period greater than 1. On the other hand, if every cycle is of length 1 (edge), then there is no path between any two states in G_π, meaning it is not irreducible.

2. The only possible way for every pair of states s and s' to have a directed path from each to the other in G_π (which is the requirement for irreducibility) is for G_π to contain a cycle of length n. In turn, this would mean G_π is not aperiodic.

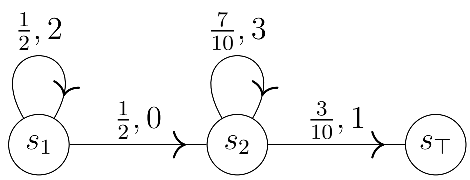
CS 747 (Spring 2025) Week 8 Test (Batch 2)

6.15 p.m. – 6.40 p.m., March 13, 2025, LA 001

Name: ______________________________ Roll number: _______________

Note. There is one question in this test. You can use the space on both pages for your answer. Draw a line (either vertical or horizontal) and do all your rough work on one side of it.

Question 1. On an MDP with non-terminal states s_1, s_2, and terminal state $s_\top$, taking actions according to a fixed policy π results in the transitions shown in the figure below, wherein arrows are annotated with "transition probability, reward".



There is no discounting. Suppose a single episode is generated by starting at s_1, and the First-visit Monter Carlo method is employed to estimate state values under π. Let $\hat{V}_1$ denote the estimate of the value of s_1 based on this one generated episode. Calculate the probability that $\hat{V}_1$ is less than half its expectation: that is, calculate

$$\mathbb{P}\left\{\hat{V}_1 < \frac{\mathbb{E}[\hat{V}_1]}{2}\right\}.$$

Write down the steps to arrive at your answer. [3 marks]

Answer 1.

We know that $\mathbb{E}[\hat{V}_1] = V^\pi(s_1)$. First we calculate this quantity by solving the Bellman equations.

$$\begin{aligned}
V^\pi(s_2) &= \frac{7}{10}(3 + V^\pi(s_2)) + \frac{3}{10}(1) \\
\implies \frac{3}{10}V^\pi(s_2) &= \frac{21}{10} + \frac{3}{10} \\
\implies V^\pi(s_2) &= 8.
\end{aligned}$$

$$\begin{aligned}
V^\pi(s_1) &= \frac{1}{2}(2 + V^\pi(s_1)) + \frac{1}{2}(V^\pi(s_2)) \\
\implies \frac{1}{2}V^\pi(s_1) &= \frac{2 + V^\pi(s_2)}{2} \\
\implies V^\pi(s_1) &= 10.
\end{aligned}$$

We are asked for the probability that $\hat{V}_1$ is smaller than $\frac{\mathbb{E}[\hat{V}_1]}{2} = \frac{V^\pi(s_1)}{2} = 5$. The only ways in which it can happen that $\hat{V}_1 < 5$ is when the generated episode is

- $s_1, s_2, s_\top$, in which case $\hat{V}_1 = 0 + 1 = 1$;
- $s_1, s_2, s_2, s_\top$, in which case $\hat{V}_1 = 0 + 3 + 1 = 4$; or
- $s_1, s_1, s_2, s_\top$, in which case $\hat{V}_1 = 2 + 0 + 1 = 3$.

The probabilities of these sequences being generated are $\frac{1}{2} \times \frac{3}{10}$, $\frac{1}{2} \times \frac{7}{10} \times \frac{3}{10}$, and $\frac{1}{2} \times \frac{1}{2} \times \frac{3}{10}$, respectively. The probability that any one of them is generated is their sum

$$\frac{3}{20} + \frac{21}{200} + \frac{3}{40} = \frac{33}{100},$$

which is our answer.