CS 747 (Autumn 2025) Week 7 Test (Batch 1)

5.35 p.m. – 6.00 p.m., September 30, 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. This question requires you to provide examples of MDPs satisfying certain properties. There are three parts. For each part the MDP you specify must be a continuing MDP with exactly **3 states** and **1 action** per state (hence with only a single policy).

Recall that an ergodic MDP is both irreducible and aperiodic.

1a. Give an example of an MDP that is irreducible but not aperiodic. Call this MDP M_1. [1 mark]

1b. Give an example of an MDP that is not irreducible but aperiodic. Call this MDP M_2. [1 mark]

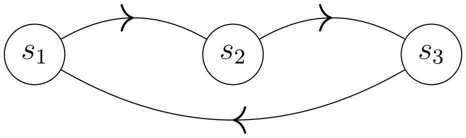
1c. Give an example of an MDP that is not irreducible and not aperiodic. Call this MDP M_3. [1 mark]

Provide complete specifications of M_1, M_2, and M_3. You are encouraged to show their transitions using state transition diagrams. You can also use textual/mathematical descriptions.

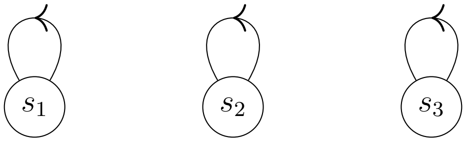
Answer 1. Irreducibility and aperiodicity do not depend on the rewards and discounting; for all our examples below they can be set arbitrarily—say all the rewards to 0 and the discount factor to 1/2.

Indeed we can give a deterministic MDP as an example for each of M_1, M_2, and M_3. One possible set of solutions is given below (several others exist). The transitions functions of the MDPs are shown through state-transition diagrams.

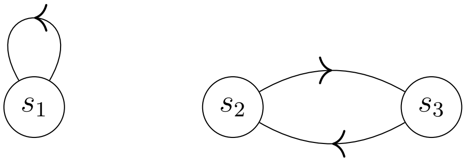
1a. M_1 (irreducible, not aperiodic).



1b. M_2 (not irreducible, aperiodic).



1c. M_3 (not irreducible, not aperiodic).



CS 747 (Autumn 2025) Week 7 Test (Batch 2)

6.15 p.m. – 6.40 p.m., September 30, 2025, LA 001

Name: ________________________________ Roll number: ________________

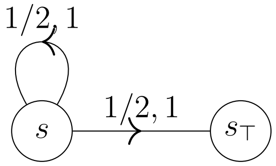
Note. There is one question in this test. You can use the space on both pages for your answer. Draw a line (either vertical or horizontal) and do all your rough work on one side of it.

Question 1. A policy π is followed on an episodic MDP, in which every reward is strictly positive, and no discounting is applied. A finite number of episodes are executed, each starting at the same state s. The data from these episodes is used to estimate the value of s under π.

- The estimate $\widehat{V}_{\text{FV}}(s)$ is obtained by using the First-Visit Monte Carlo technique.
- The estimate $\widehat{V}_{\text{EV}}(s)$ is obtained by using the Every-Visit Monte Carlo technique.

Let C_1 be the claim that $\widehat{V}_{\text{EV}}(s) \leq \widehat{V}_{\text{FV}}(s)$, and let C_2 be the claim that $\widehat{V}_{\text{EV}}(s) \geq \widehat{V}_{\text{FV}}(s)$. Is one of these claims necessarily correct? If your answer is yes, specify which claim, and prove its correctness. If your answer is no, provide a counterexample to each claim. A correct claim must hold for all qualifying MDPs and possible sequences of episodes generated. A counterexample need only specify a single MDP and one possible sequence of episodes generated on it. [3 marks]

Answer 1. Consider an MDP with a single non-terminal state s and single action a. The action transitions into s with probability 1/2 and terminates with the remaining probability, as shown in the figure below. Both rewards are 1.



An episode in this MDP is fully described by the number of steps it lasts. For ease of calculation, the table below shows the full sequence of rewards from each episode, on one particular run. Shown alongside are the first-visit and every-visit Monte Carlo estimates of the value of s after each episode.

Episode number	Sequence of rewards	$\widehat{V}_{\mathrm{FV}}(s)$	$\widehat{V}_{\mathrm{EV}}(s)$
1	1, 1, 1, 1, 1	5	3
2	1	3	8/3
3	1	7/3	17/7

Notice that after the first and second episodes, $\widehat{V}_{\mathrm{FV}}(s) > \widehat{V}_{\mathrm{EV}}(s)$. However, after the third episode, $\widehat{V}_{\mathrm{FV}}(s) < \widehat{V}_{\mathrm{EV}}(s)$. Clearly claims C_1 and C_2 are both incorrect.