

University of Virginia
Department of Computer Science

**CS 4501: Introduction to Reinforcement Learning
Fall 2022**

Thursday 9:30am-9:45am EDT, October 27st, 2022

Name:
ComputingID:

- This is a **closed book** and **closed notes** quiz. No electronic aids or cheat sheets or discussing the questions with anyone else are allowed.
- You are expected to finish this quiz within 15 minutes.
- There are 2 pages, 3 parts of questions, and 20 total points in this quiz.
- The questions are printed on the back of this page!
- Please carefully read the instructions and questions before you answer them.
- If you need any clarification of the quiz questions, please raise your hand and discuss with the instructor within the quiz period.
- Try to keep your answers as concise as possible; our grading is *NOT* by keyword matching.

Total	/20
-------	-----

1 True/False Questions (2×3 pts)

Please choose either True or False for each of the following statements. For the statement you believe it is False, please give your brief explanation of it (you do **NOT** need to explain when you believe it is True). Three point for each question. *Note: the credit can only be granted if your explanation for the false statement is correct.*

1. An optimal policy must lead to optimal value for every state.
True
2. Because of the MDP assumption, the state-value evaluation under MC method depends on its evaluation of neighboring states. *False, and Explain: MC method does not depend on the MDP assumption, and its state-value estimations are independent from each other.*

2 Multiple Choice Questions (2×4 pts pts)

Please choose ALL the answers that you believe are correct for each question.

1. What's needed for off-policy MC control method? (c)
(a) complete knowledge of the environment; (b) MDP assumption;
(c) soft behavior policy; (d) soft target policy.
2. How can we accelerate policy iteration in practice: (a) (b)
(a) prioritized sweeping; (b) in-place policy evaluation;
(c) sample more episodes; (d) use a larger step size.

3 Short Answer Question (6 pts)

The question can be answered by one or two sentences; so please make your answer concise and to the point.

1. What is the main disadvantage of off-policy MC control method? Can you provide one method that can help alleviate the disadvantage?

High variance is the main concern of off-policy MC control method, especially when the target policy differs significantly from the target policy. Possible solutions to reduce the variance include per-decision importance sampling, discounting-aware importance sampling, clipping (but it is no longer unbiased).