National Taiwan University
Department of Electrical Engineering
Name: Ching-Hsiang, Wu
ID: R11921080

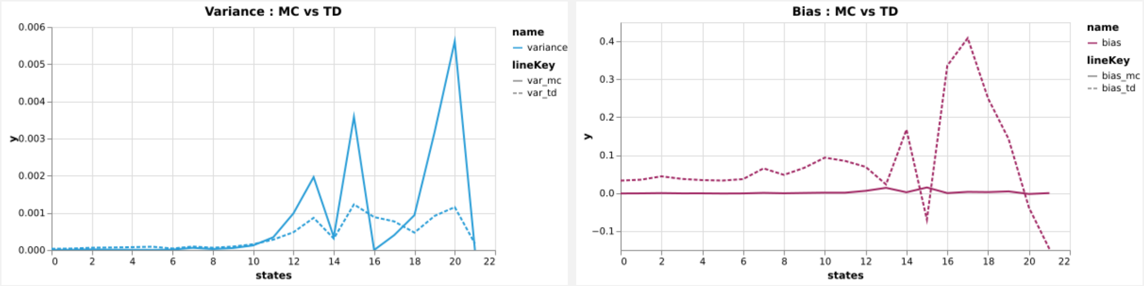
113-1 (Fall 2024)
Reinforcement Learning

Assignment 2 Report

Q1. Run Monte-Carlo prediction and TD(0) prediction for 50 seeds. Compare the resulting values with the GT values. Discuss the variance and bias.

Fig. 1 shows the result of variance and bias after running 50 seeds (0-49). The x-axis means states, y-axis is the magnitude of variance or bias, the solid line represents MC prediction, and the dotted line means TD(0) prediction.

As our expectation, in Fig. 1a, using TD(0) to predict the state value gains less variance compared to using MC prediction. On the other hand, Fig. 1b shows that TD(0) gains more bias than using MC for almost every state.



(a) MC vs. TD : Variance

(b) MC vs. TD : Bias

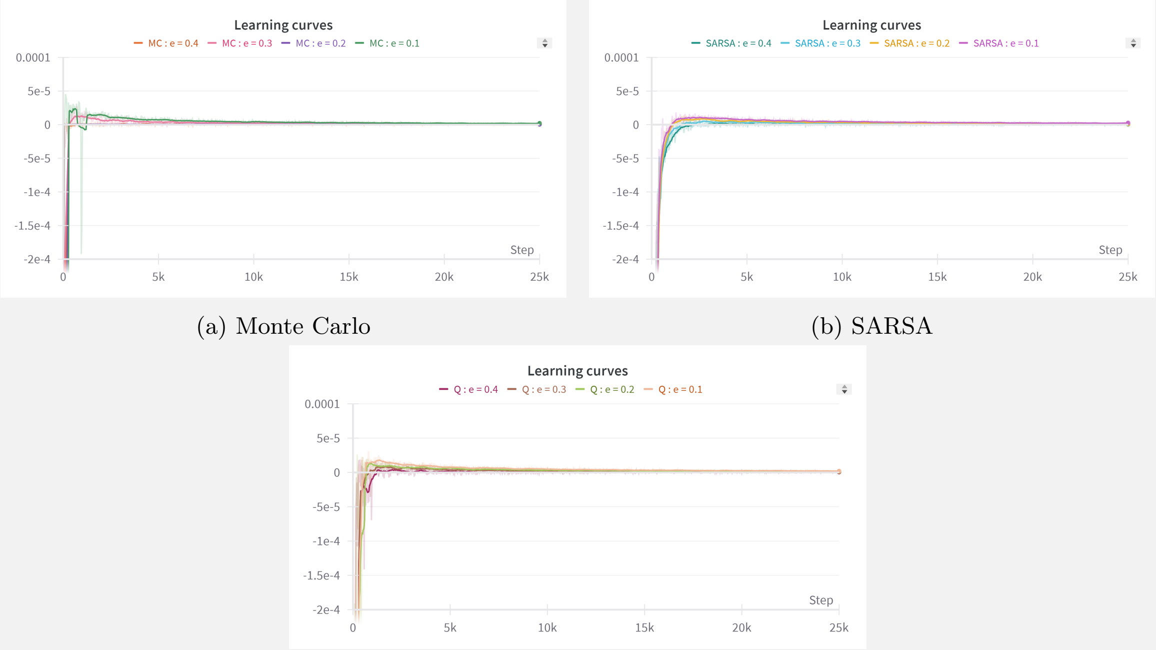
Figure 1: MC vs. TD : Variance and bias comparison

Q2. Discuss and plot learning curves under ϵ values of $(0.1, 0.2, 0.3, 0.4)$ on MC, SARSA, and Q-Learning

In this problem, I ran Monte Carlo(MC), SARSA control for 512000 episodes, and Q-Learning for 50000 episodes, however, only the first 25000 episodes were taken out for comparison. Each plot has been smoothed by a running average with 20 data lengths for a clearer view.

In Fig. 2, SARSA control has a more stable (less variance) training process than MC control, which is aligned with the property of MC and TD. It is worth noting that when $\epsilon = 0.2$ in MC, the learning curve almost converges in 10 episodes, outperforming the other three cases.

On the other hand, we can see that the ϵ value is smaller, the performance (convergence speed in Fig. 2, the final average reward value in Fig. 3) is better. This is because ϵ value determines how often this agent explores. If this process has converged, more explorations are unnecessary.

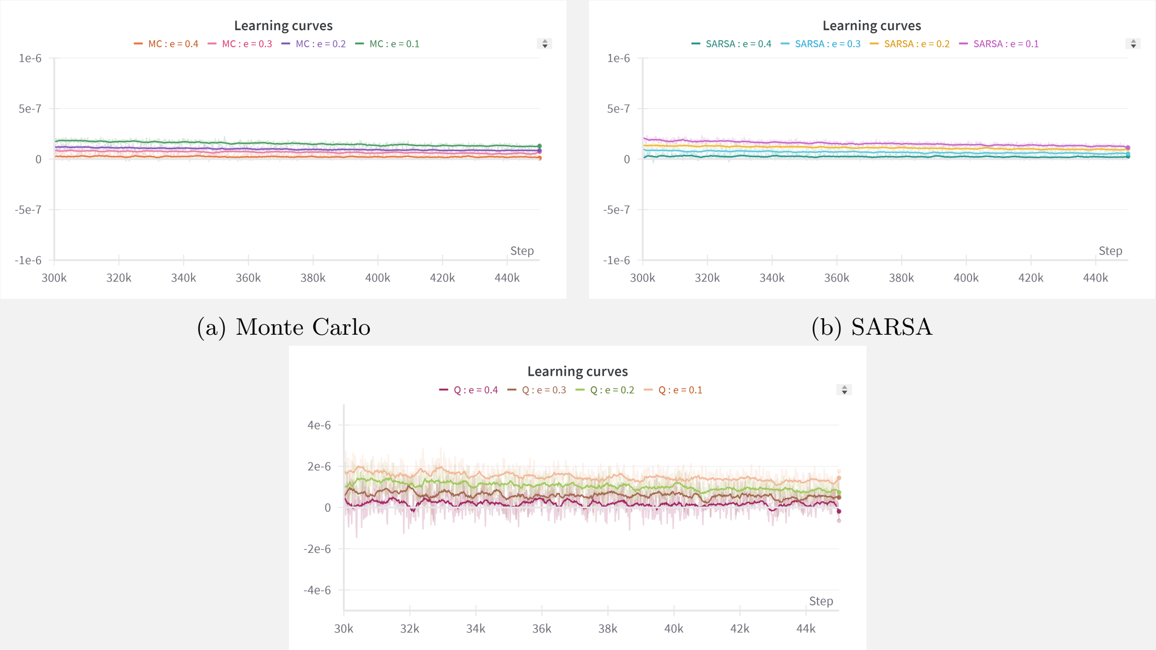


(a) Monte Carlo

(b) SARSA

(c) Q-Learning

Figure 2: Learning curves : MC vs. SARSA vs. Q-Learning



(a) Monte Carlo

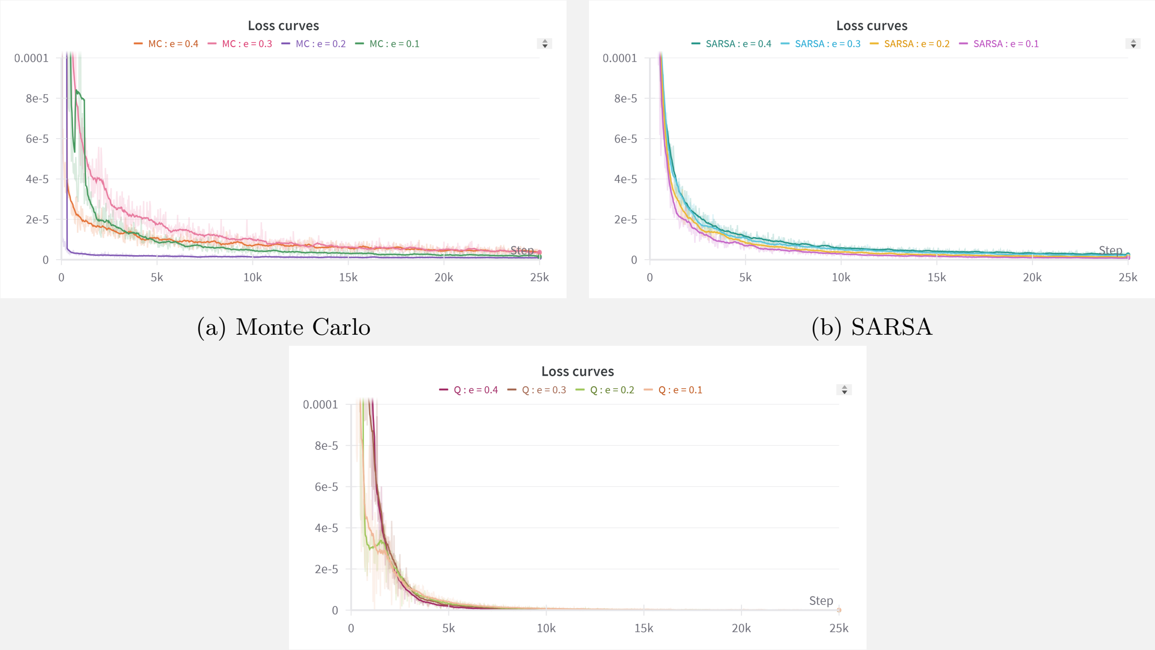
(b) SARSA

(c) Q-Learning

Figure 3: Learning curves final reward value: MC vs. SARSA vs. Q-Learning

Q3. Discuss and plot loss curves under ϵ values of $(0.1, 0.2, 0.3, 0.4)$ on MC, SARSA, and Q-Learning

Basically, the loss curves show the same properties as learning curves. The ϵ value is smaller, the loss curves are faster to converge, and the final loss value is smaller. However, Q-Learning in Fig. 5c shows that $\epsilon = 0.4$ has the smaller final loss value instead.

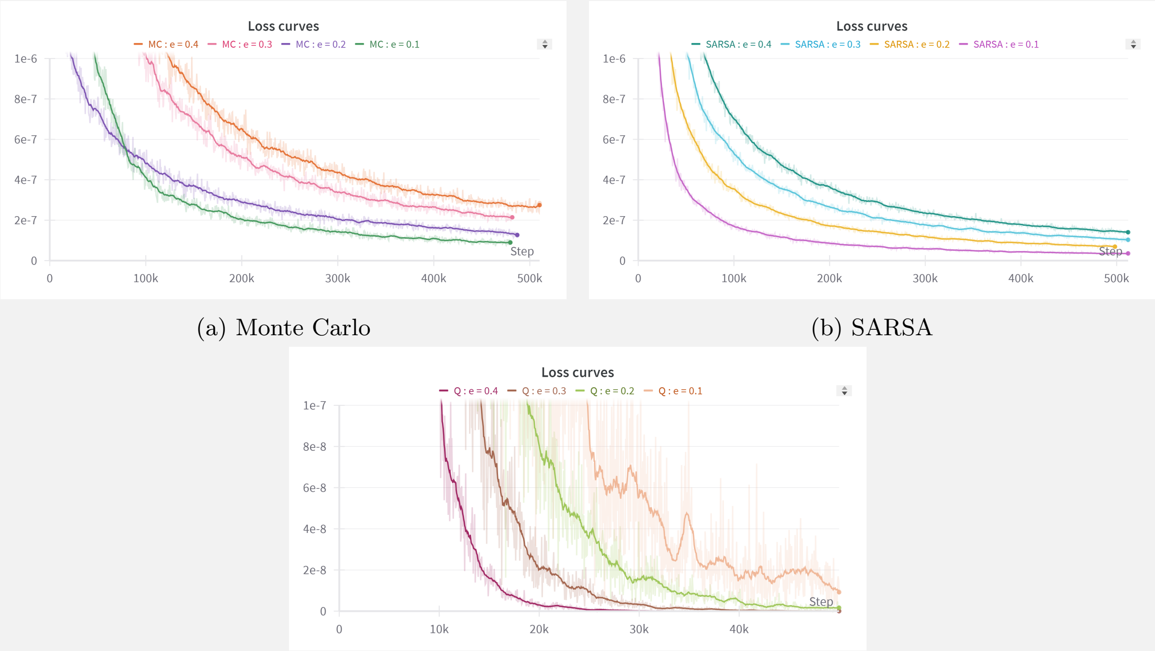


(a) Monte Carlo

(b) SARSA

(c) Q-Learning

Figure 4: Loss curves: MC vs. SARSA vs. Q-Learning



(a) Monte Carlo

(b) SARSA

(c) Q-Learning

Figure 5: Final loss comparison: MC vs. SARSA vs. Q-Learning