Northeastern University

CS5100 Foundations of Artificial Intelligence

Assignment Problem Set 2 - Probabilistic Reasoning/Reinforcement Learning

Anuj Patel (NU ID: 002874710)

Q1. Given the above Markov model, explain whether the stationary distribution depends on the start state (without actually computing the stationary distribution)

to check given markov chain depend on the initial state I will check the irreducibility and aperiodicity.

1. Irreducibility

markov chain is **irreducible** if every state can reached from any other state in a finite number of steps.

- $S1$:
  - $S1 \to S2$ with probability 0.6
  - $S1 \to S3$ with probability 0.4
- $S2$:
  - $S2 \to S1$ with probability 0.7
  - $S2 \to S2$ with probability 0.3
- $S3$:
  - $S3 \to S1$ with probability 0.2
  - $S3 \to S2$ with probability 0.8

check connection between few of above states:

- $S1$ to $S2$ and $S3$: direct transitions.
- $S2$ to $S1$: direct transition .
- $S2$ to $S3$: $S2 \to S1 \to S3$ (has to take the two steps).
- $S3$ to $S1$ and $S2$: Direct transitions .
- $S3$ to $S2$: direct transition .

every state can reach every other state so markov is **irreducible**.

2. aperiodicity

markov is **aperiodic** if it can return to itself at irregular intervals not fixed multiples of some integer greater than 1.

- **s1:**
  - return $S1$ in 2 steps: $S1 \to S2 \to S1$.
  - return in 3 steps: $S1 \to S2 \to S2 \to S1$.
  - greatest common divisor of number of steps is 1.
- **s2:**

  - self-loop: $S2 \to S2$.
  - return itself in 1 step.

- **s3:**
  - return $S3$ in 2 steps: $S3 \to S1 \to S3$.
  - return in 3 steps: $S3 \to S2 \to S1 \to S3$.
  - gcd of the number of steps is 1.

All states have a period of 1 so markov chain is **aperiodic**.

Final ans:
Since the Markov chain is irreducible and aperiodic, the stationary distribution does not depend on the initial state.

Q2. Given the initial probability distribution $p_0 = [0.7, 0.2, 0.1]$ for states S_1, S_2, and S_3 respectively, compute the stationary distribution for the given Markov model. You may use a program to do the computations. Report only the final stationary distribution π.

Step 1: transition matrix

The transition matrix P based on given probabilities:

$$P = \begin{bmatrix} 0.0 & 0.6 & 0.4 \\ 0.7 & 0.3 & 0.0 \\ 0.2 & 0.8 & 0.0 \end{bmatrix}$$

Step 2: stationary dist

For stationary distribution $\pi = [\pi_1, \pi_2, \pi_3]$, we require:

- $\pi = \pi P$
- $\pi_1 + \pi_2 + \pi_3 = 1$

So we get:

$$\begin{aligned} &1) \quad \pi_1 = 0.7\pi_2 + 0.2\pi_3 \\ &2) \quad \pi_2 = 0.6\pi_1 + 0.3\pi_2 + 0.8\pi_3 \\ &3) \quad \pi_3 = 0.4\pi_1 \\ &4) \quad \pi_1 + \pi_2 + \pi_3 = 1 \end{aligned}$$

Step 3: calculation

1. From equation 3:

$$\pi_3 = 0.4\pi_1$$

$$\pi_1 = 0.7\pi_2 + 0.2 \times 0.4\pi_1$$
$$0.92\pi_1 = 0.7\pi_2$$
$$\pi_1 = \frac{0.7}{0.92}\pi_2 \approx 0.761\pi_2$$

$$\pi_2 = 0.484$$
$$\pi_3 = 0.147$$

Step 4: Verification

- Sum check: $0.368 + 0.484 + 0.147 = 0.999 \approx 1$
- $\pi P = \pi$ verification satisfied

Final Answer

$$\boxed{\text{stationary distribution is } \pi = [0.368, 0.484, 0.147]}$$

Q3. What is the probability of the following transition sequence: $S_2 \rightarrow S_2 \rightarrow S_1 \rightarrow S_3$?

1. use the abobe q2 stationary distribution:

- value is:

$$\pi = [0.3684, 0.4842, 0.1474]$$

2. calculate the probab of the sequence:

- probability of sequence is the product of:
  - probability of starting in S_2
  - transition probabilities along the sequence.
- **eq:**

$$P(\text{Sequence}) = \pi_{\text{start}} \times P(\text{Transition 1}) \times P(\text{Transition 2}) \times P(\text{Transition 3})$$

- **values:**
  - **Starting prob:**

$$\pi_{\text{start}} = \pi_2 = 0.4842$$

  - **transition prob:**
    - $P(S_2 \rightarrow S_2) = 0.3$
    - $P(S_2 \rightarrow S_1) = 0.7$
    - $P(S_1 \rightarrow S_3) = 0.4$
- **put in above eq:**

$$P(\text{Sequence}) = \pi_2 \times P(S_2 \rightarrow S_2) \times P(S_2 \rightarrow S_1) \times P(S_1 \rightarrow S_3)$$

$$P(\text{Sequence}) = 0.4842 \times 0.3 \times 0.7 \times 0.4$$

P(Sequence) = 0.0406728

Final ans:
probability of the sequence $S_2 \rightarrow S_2 \rightarrow S_1 \rightarrow S_3$ is approx **0.0407**

Q4. Given that we start in the state S_2, what is the probability of returning to S_2 after two transitions? (Show your calculations.)

find compute the total probability of all possible paths that begin at S_2, take two transitions and end at S_2.

Step 1: possible Paths
from S_2 need to identify all possible two-step paths that end at S_2. The possible paths are:

- **path A:** $S_2 \to S_2 \to S_2$
- **path B:** $S_2 \to S_1 \to S_2$

Step 2: collect transition probab

- **From S_2:**
  - $P(S_2 \to S_2) = 0.3$
  - $P(S_2 \to S_1) = 0.7$
- **From S_1:**
  - $P(S_1 \to S_2) = 0.6$
  - (Other transitions from S_1 are irrelevant for these paths)

Step 3: calculate the probab for each path
Path A: $S_2 \to S_2 \to S_2$

- **Path Probability:**

$$P(\text{Path A}) = P(S_2 \to S_2) \times P(S_2 \to S_2) = 0.3 \times 0.3 = 0.09$$

Path B: $S_2 \to S_1 \to S_2$

- **Path Probability:**

$$P(\text{Path B}) = P(S_2 \to S_1) \times P(S_1 \to S_2) = 0.7 \times 0.6 = 0.42$$

Step 4: sum the prob of all paths

- **total prob:**

$$P(\text{Return to } S_2 \text{ after two transitions}) = 0.09 + 0.42 = 0.51$$

Final ans:
prob of returning to S_2 after two transitions starting from S_2 is **0.51**.

Q5. Given the starting probability distribution $p_0 = [0.7, 0.2, 0.1]$, what is the probability of ending up in S_2 after exactly two transitions? (Show your calculations.)

Step 1: transition matrix P

Transition Matrix P:

$$P = \begin{bmatrix} 0 & 0.6 & 0.4 \\ 0.7 & 0.3 & 0 \\ 0.2 & 0.8 & 0 \end{bmatrix}$$

Step 2: calculate P^2 (transition after 2 step)

calculate $P^2 = P \times P$.

calculate for each elements of P^2:

- $P^2_{11} = (0)(0) + (0.6)(0.7) + (0.4)(0.2) = 0.5$
- $P^2_{12} = (0)(0.6) + (0.6)(0.3) + (0.4)(0.8) = 0.5$
- $P^2_{13} = (0)(0.4) + (0.6)(0) + (0.4)(0) = 0$
- $P^2_{21} = (0.7)(0) + (0.3)(0.7) + (0)(0.2) = 0.21$
- $P^2_{22} = (0.7)(0.6) + (0.3)(0.3) + (0)(0.8) = 0.51$
- $P^2_{23} = (0.7)(0.4) + (0.3)(0) + (0)(0) = 0.28$
- $P^2_{31} = (0.2)(0) + (0.8)(0.7) + (0)(0.2) = 0.56$
- $P^2_{32} = (0.2)(0.6) + (0.8)(0.3) + (0)(0.8) = 0.36$
- $P^2_{33} = (0.2)(0.4) + (0.8)(0) + (0)(0) = 0.08$

updated matrix P^2:

$$P^2 = \begin{bmatrix} 0.5 & 0.5 & 0 \\ 0.21 & 0.51 & 0.28 \\ 0.56 & 0.36 & 0.08 \end{bmatrix}$$

Step 3: calculate $p_2 = p_0 P^2$

calculate the prob distribution after two transitions by multiplying the initial distribution p_0 with P^2:

$$p_2 = p_0 P^2$$

initial probab distribution p_0:

$$p_0 = [0.7, 0.2, 0.1]$$

calculate each component of p_2:

- $p_2(S_1) = (0.7)(0.5) + (0.2)(0.21) + (0.1)(0.56) = 0.448$
- $p_2(S_2) = (0.7)(0.5) + (0.2)(0.51) + (0.1)(0.36) = 0.488$

- $p_2(S_3) = (0.7)(0) + (0.2)(0.28) + (0.1)(0.08) = 0.064$

prob dist after two transitions p_2

$$p_2 = [0.448, 0.488, 0.064]$$

$$\text{prob of being in } S_2 \text{ after two transitions} = p_2(S_2) = 0.488$$

Final ans:
prob of ending up in state S_2 after two transitions is: 0.488

Q6.Hidden Markov Model of the Coin-Flipping Game

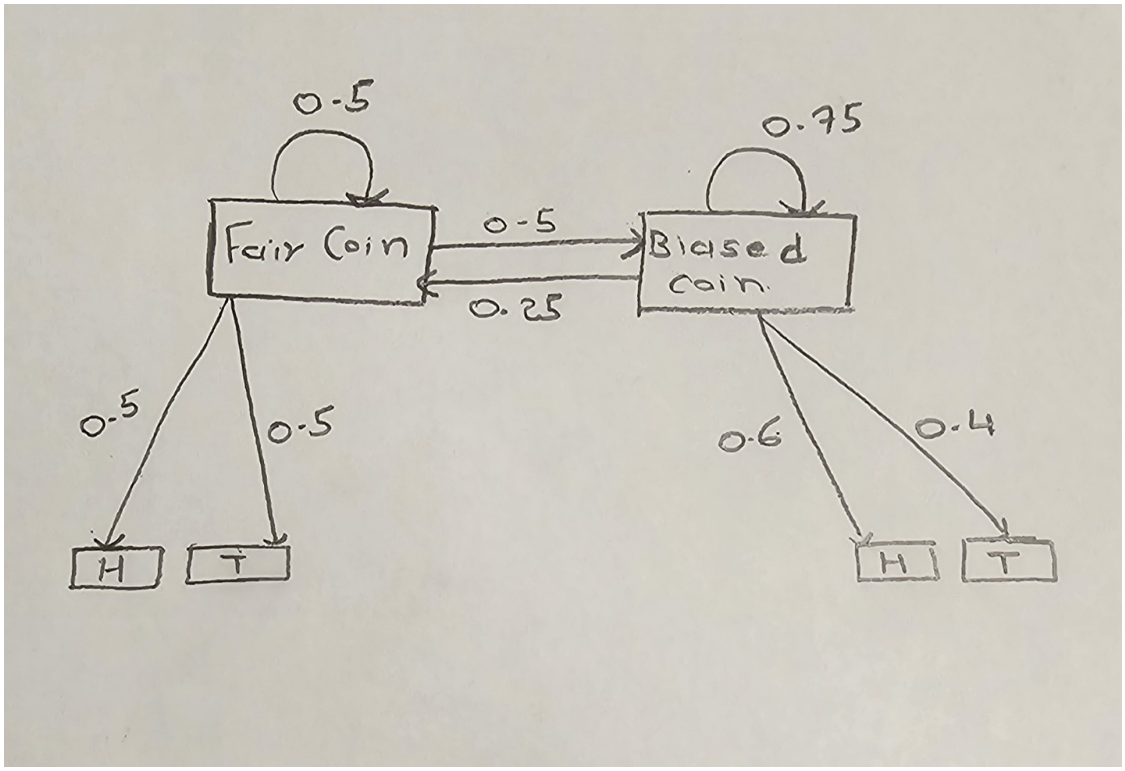

Figure 1: HMM Diagram

Hidden States

- Fair coin
- Biased coin

Observation

- H: Heads
- T: Tails

Transition Probabilities

- From Fair Coin:
  - $P(F \to F) = 0.5$
  - $P(F \to B) = 0.5$
- From Biased Coin:
  - $P(B \to B) = 0.75$
  - $P(B \to F) = 0.25$

Emission Probabilities

- From Fair Coin (F):
  - $P(H|F) = 0.5$
  - $P(T|F) = 0.5$
- From Biased Coin (B):
  - $P(H|B) = 0.6$
  - $P(T|B) = 0.4$

Q7. Predicting the Most Likely Coin Sequence

Step 1: write down given values

1.State

$$S = \{F, B\}$$

2.Observations

$$O = \{H, T\}$$

3. starting Probabilities (π)

find by stationary distribution.

4. transition Probabilities (A)

- If F flipped:

$$P(F \to F) = 0.5$$
$$P(F \to B) = 0.5$$

- If B flipped:

$$P(B \to B) = \frac{3}{4}$$
$$P(B \to F) = \frac{1}{4}$$

so transition matrix A is:

$$A = \begin{bmatrix} 0.5 & 0.5 \\ 0.25 & 0.75 \end{bmatrix}$$

5. emission Probabilities (B)

- For F:

$$P(H|F) = 0.5$$
$$P(T|F) = 0.5$$

- For B:

$$P(H|B) = 0.6$$
$$P(T|B) = 0.4$$

Step 2: calculayte stationary distribution

calculate stationary distribution (π_F, π_B) satisfying:

$$\begin{cases} \pi_F = \pi_F \cdot P(F \to F) + \pi_B \cdot P(B \to F) \\ \pi_B = \pi_F \cdot P(F \to B) + \pi_B \cdot P(B \to B) \\ \pi_F + \pi_B = 1 \end{cases}$$

1. Substitute $\pi_B = 1 - \pi_F$ into the first equation:

$$\pi_F = \pi_F \cdot 0.5 + (1 - \pi_F) \cdot 0.25$$

$$\pi_F = 0.5\pi_F + 0.25 - 0.25\pi_F 0.75\pi_F = 0.25\pi_F = \frac{1}{3}$$

$$\pi_B = 1 - \pi_F = \frac{2}{3}$$

Conclusion: The stationary distribution is $\pi = \left(\frac{1}{3}, \frac{2}{3}\right)$.

Step 3: Viterbi Algorithm

it use to find the most probable sequence of hidden state that results in the observed sequence.

Initialization $(t = 1)$

$$\delta_1(i) = \pi_i \cdot b_i(o_1)$$

Where $o_1 = H$.

Compute $\delta_1(F)$ and $\delta_1(B)$:

$$\delta_1(F) = \frac{1}{3} \cdot 0.5 = \frac{1}{6} \approx 0.1667$$
$$\delta_1(B) = \frac{2}{3} \cdot 0.6 = 0.4$$

Recursion $(t = 2, 3)$

For each time step t, compute:

$$\delta_t(j) = \max_i [\delta_{t-1}(i) \cdot a_{ij}] \cdot b_j(o_t)$$
$$\psi_t(j) = \arg\max_i [\delta_{t-1}(i) \cdot a_{ij}]$$

At $t = 2$ **(Observation** $o_2 = T$**)**

Compute $\delta_2(F)$ and $\delta_2(B)$:

- For F:

$$\delta_2(F) = \max(0.0833, 0.1) \cdot 0.5 = 0.05 \quad \psi_2(F) = B$$

- For B:

$$\delta_2(B) = \max(0.0833, 0.3) \cdot 0.4 = 0.12 \quad \psi_2(B) = B$$

At $t = 3$ (Observation $o_3 = T$)

Compute $\delta_3(F)$ and $\delta_3(B)$:

- For F:
$$\delta_3(F) = \max(0.025, 0.03) \cdot 0.5 = 0.015 \quad \psi_3(F) = B$$
- For B:
$$\delta_3(B) = \max(0.025, 0.09) \cdot 0.4 = 0.036 \quad \psi_3(B) = B$$

termination chose the final state with the highest probability:

$$\delta_3(B) = 0.036 > \delta_3(F) = 0.015 \Rightarrow s_3 = B$$

Backtracking to find the most likely state sequence

$$\begin{aligned} s_3 &= B \\ s_2 &= \psi_3(s_3) = \psi_3(B) = B \\ s_1 &= \psi_2(s_2) = \psi_2(B) = B \end{aligned}$$

Final ans:

- **First Flip:** Biased Coin (B)
- **Second Flip:** Biased Coin (B)
- **Third Flip:** Biased Coin (B)

Stationary distribution shows that at any given time there is a $\frac{2}{3}$ probability of using the biased coin and a $\frac{1}{3}$ probability of using the fair coin.

Initial probability match with our calculations in the Viterbi algorithm shows the biased coin was most likely used in the first round.

Q8. Tota actions

Step 1: available actions from each state

- S_1:
  - Available actions: A_1, A_2
  - 2 possible actions
- S_2:
  - Available actions: A_1, A_3
  - 2 possible actions
- S_3:
  - Available action: A_4
  - 1 possible action
- S_4:
  - No available actions (indicated by X)
  - 1 possible action (indicating no action can be taken)

Step 2: total number of unique policies

Total number of unique policies = Actions from S_1×Actions from S_2×Actions from S_3×Actions from S_4

$$\text{Total number of unique policies} = 2 \times 2 \times 1 \times 1 = 4$$

Final ans:
there are **4 unique policies**.

Step 3: List all policies

4 policies are:

- **Policy 1**: $S_1 \to A_1$, $S_2 \to A_1$, $S_3 \to A_4$, $S_4 \to X$
- **Policy 2**: $S_1 \to A_1$, $S_2 \to A_3$, $S_3 \to A_4$, $S_4 \to X$
- **Policy 3**: $S_1 \to A_2$, $S_2 \to A_1$, $S_3 \to A_4$, $S_4 \to X$
- **Policy 4**: $S_1 \to A_2$, $S_2 \to A_3$, $S_3 \to A_4$, $S_4 \to X$

these are the four unique policies.

Q9

Valid Paths and probabilities

Paths to S_4 from S_1:

1. **Length 2 Path:**

- $S_1 \to A_2 \to S_3 \to A_4 \to S_4$
- **Total:** $0.5 \times 0.7 \times 1 \times 0.7 = 0.245$

2. **Length 3 Paths:**

- $S_1 \to A_1 \to S_2 \to A_3 \to S_3 \to A_4 \to S_4$
- **Total:** $0.5 \times 0.5 \times 0.5 \times 0.6 \times 1 \times 0.7 = 0.0525$
- $S_1 \to A_2 \to S_2 \to A_3 \to S_3 \to A_4 \to S_4$
- **Total:** $0.5 \times 0.3 \times 0.5 \times 0.6 \times 1 \times 0.7 = 0.0315$
- $S_1 \to A_1 \to S_1 \to A_2 \to S_3 \to A_4 \to S_4$
- **Total:** $0.5 \times 0.5 \times 0.5 \times 0.7 \times 1 \times 0.7 = 0.06125$

Total Probability Calculation:
$P_{\text{total}} = 0.245 + 0.0525 + 0.0315 + 0.06125 = 0.39.$

Que10

probabilities:

- From S_1:

$$P(S_2|S_1, A_1) = 0.5, \quad P(S_1|S_1, A_1) = 0.5$$
$$P(S_2|S_1, A_2) = 0.3, \quad P(S_3|S_1, A_2) = 0.7$$

- From S_2:

$$P(S_3|S_2, A_3) = 0.6, \quad P(S_2|S_2, A_3) = 0.4$$

- From S_3:

$$P(S_4|S_3, A_4) = 0.7, \quad P(S_1|S_3, A_4) = 0.3$$

Step 1: Bellman eq

eq for the value function $V_{\text{opt}}(S)$ in a given state S is:

$$V_{\text{opt}}(S) = \max_a \left(\sum_{s'} P(s'|S, a) \left[R(S, a, s') + \gamma V_{\text{opt}}(s') \right] \right)$$

Step 2: Calculating $V_{\text{opt}}(S_1)$

action A_1 in S_1:

From S_1, taking action A_1 leads to:

- S_1 with probability 0.5 and reward 0,
- S_2 with probability 0.5 and reward 0.

$$V_{\text{opt}}(S_1 \mid A_1) = 0.5 \cdot (0 + \gamma V_{\text{opt}}(S_1)) + 0.5 \cdot (0 + \gamma V_{\text{opt}}(S_2))$$
$$V_{\text{opt}}(S_1 \mid A_1) = 0.5\gamma V_{\text{opt}}(S_1) + 0.5\gamma V_{\text{opt}}(S_2)$$

action A_2 in S_1:

From S_1, taking action A_2 leads to:

- S_2 with probability 0.3 and reward 0,
- S_3 with probability 0.7 and reward 5.

$$V_{\text{opt}}(S_1 \mid A_2) = 0.3 \cdot (0 + \gamma V_{\text{opt}}(S_2)) + 0.7 \cdot (5 + \gamma V_{\text{opt}}(S_3))$$

$$V_{\text{opt}}(S_1 \mid A_2) = 0.3\gamma V_{\text{opt}}(S_2) + 3.5 + 0.7\gamma V_{\text{opt}}(S_3)$$

Step 3: value unction for $V_{\mathbf{opt}}(S_1)$

eq for $V_{\text{opt}}(S_1)$ as the maximum of these two values:

$$V_{\text{opt}}(S_1) = \max\left(0.5\gamma V_{\text{opt}}(S_1) + 0.5\gamma V_{\text{opt}}(S_2), 0.3\gamma V_{\text{opt}}(S_2) + 3.5 + 0.7\gamma V_{\text{opt}}(S_3)\right)$$

Final eq:

$$V_{\text{opt}}(S_1) = \max\left(0.5\gamma V_{\text{opt}}(S_1) + 0.5\gamma V_{\text{opt}}(S_2), 0.3\gamma V_{\text{opt}}(S_2) + 3.5 + 0.7\gamma V_{\text{opt}}(S_3)\right)$$

Que. 11

1. Initialization

- **Q-values**: Initialize $\hat{Q}_{\text{opt}}(s, a) = 0$
- **States**: S_1, S_2, S_3, S_4
- **Actions**: A_1, A_2, A_3, A_4.

starting Q-values and Update Counts

State s	Action a	$\hat{Q}_{\text{opt}}(s, a)$	Updates $n(s, a)$
S_1	A_1	0	0
S_1	A_2	0	0
S_2	A_3	0	0
S_3	A_4	0	0

2. Episode 1

ep 1 transitions:

1. $(S_1, A_1, 0, S_2)$
2. $(S_2, A_3, 5, S_3)$
3. $(S_3, A_4, 100, S_4)$

Transition 1: $(S_1, A_1, 0, S_2)$

- **Current Q-value**: $\hat{Q}^{\text{old}}_{\text{opt}}(S_1, A_1) = 0$
- **Number of Updates**: $n(S_1, A_1) = 0$
- **Learning Rate**:
$$\eta_{S_1, A_1} = \frac{1}{1 + n(S_1, A_1)} = 1$$
- **Next State Value**:
$$\hat{V}^{\text{old}}_{\text{opt}}(S_2) = 0$$
- **Q-value Update**:
$$\hat{Q}^{\text{new}}_{\text{opt}}(S_1, A_1) = 0$$
- **Update Q-value and Counts**:
  - $\hat{Q}_{\text{opt}}(S_1, A_1) = 0$
  - $n(S_1, A_1) = 1$

transition 2: $(S_2, A_3, 5, S_3)$

- **Current Q-value**: $\hat{Q}^{\text{old}}_{\text{opt}}(S_2, A_3) = 0$
- **Number of Updates**: $n(S_2, A_3) = 0$
- **Learning Rate**:
$$\eta_{S_2, A_3} = 1$$
- **Next State Value**:
$$\hat{V}^{\text{old}}_{\text{opt}}(S_3) = 0$$
- **Q-value Update**:
$$\hat{Q}^{\text{new}}_{\text{opt}}(S_2, A_3) = 5$$
- **Update Q-value and Counts**:
  - $\hat{Q}_{\text{opt}}(S_2, A_3) = 5$
  - $n(S_2, A_3) = 1$

transition 3: $(S_3, A_4, 100, S_4)$

- **Current Q-value**: $\hat{Q}^{\text{old}}_{\text{opt}}(S_3, A_4) = 0$
- **Number of Updates**: $n(S_3, A_4) = 0$
- **Learning Rate**:
$$\eta_{S_3, A_4} = 1$$
- **Next State Value**:
$$\hat{V}^{\text{old}}_{\text{opt}}(S_4) = 0$$
- **Q-value Update**:
$$\hat{Q}^{\text{new}}_{\text{opt}}(S_3, A_4) = 100$$
- **Update Q-value and Counts**:
  - $\hat{Q}_{\text{opt}}(S_3, A_4) = 100$
  - $n(S_3, A_4) = 1$

3. episode 2

ep 2 transitions:

1. $(S_1, A_2, 5, S_3)$
2. $(S_3, A_4, 100, S_4)$

transition 1: $(S_1, A_2, 5, S_3)$

- **Current Q-value**: $\hat{Q}^{\text{old}}_{\text{opt}}(S_1, A_2) = 0$
- **Number of Updates**: $n(S_1, A_2) = 0$
- **Learning Rate**:
$$\eta_{S_1, A_2} = 1$$
- **Next State Value**:
$$\hat{V}^{\text{old}}_{\text{opt}}(S_3) = 100$$
- **Q-value Update**:
$$\hat{Q}^{\text{new}}_{\text{opt}}(S_1, A_2) = 105$$
- **Update Q-value and Counts**:
  - $\hat{Q}_{\text{opt}}(S_1, A_2) = 105$
  - $n(S_1, A_2) = 1$

transition 2: $(S_3, A_4, 100, S_4)$

- **Current Q-value**: $\hat{Q}^{\text{old}}_{\text{opt}}(S_3, A_4) = 100$
- **Number of Updates**: $n(S_3, A_4) = 1$
- **Learning Rate**:
$$\eta_{S_3, A_4} = 0.5$$
- **Next State Value**:
$$\hat{V}^{\text{old}}_{\text{opt}}(S_4) = 0$$
- **Q-value Update**:
$$\hat{Q}^{\text{new}}_{\text{opt}}(S_3, A_4) = 100$$
- **Update Q-value and Counts**:
  - $\hat{Q}_{\text{opt}}(S_3, A_4) = 100$ (remains the same)
  - $n(S_3, A_4) = 2$

4. final Q-values

State s	Action a	$\hat{Q}_{\text{opt}}(s, a)$	Updates $n(s, a)$
S_1	A_1	0	1
S_1	A_2	105	1
S_2	A_3	5	1
S_3	A_4	100	2

5. calculate state value estimate

State-Value Estimates:

- S_1**:**
$$\hat{V}_{\text{opt}}(S_1) = \max\{\hat{Q}_{\text{opt}}(S_1, A_1), \hat{Q}_{\text{opt}}(S_1, A_2)\} = \max\{0, 105\} = 105$$
- S_2**:**
$$\hat{V}_{\text{opt}}(S_2) = \hat{Q}_{\text{opt}}(S_2, A_3) = 5$$
- S_3**:**
$$\hat{V}_{\text{opt}}(S_3) = \hat{Q}_{\text{opt}}(S_3, A_4) = 100$$
- S_4**:**
$$\hat{V}_{\text{opt}}(S_4) = 0 \quad \text{(terminal state)}$$

6. final optimal policy

selects the action with the highest Q-value at each state.

Optimal Actions at Each State:

- S_1**:**
  - Compare $\hat{Q}_{\text{opt}}(S_1, A_1) = 0$ and $\hat{Q}_{\text{opt}}(S_1, A_2) = 105$.
  - **Optimal Action:** A_2
- S_2**:**
  - **Optimal Action:** A_3.
- S_3**:**
  - **Optimal Action:** A_4.
- S_4**:**
  - Terminal state, no action needed.

final Optimal Policy:

- $\pi^*(S_1) = A_2$
- $\pi^*(S_2) = A_3$
- $\pi^*(S_3) = A_4$
- $\pi^*(S_4) =$ No action (terminal state)

Q12

I have read and understood the academic integrity policy as outlined in the course syllabus for CS4100/CS5100. By pasting this acknowledgement in my submission, I declare that all work presented here is my own, and any conceptual discussions I may have had with classmates have been fully disclosed. I declare that generative AI was not used to answer any questions in this assignment. Any use of generative AI to improve writing clarity alone is accompanied by an appendix with my original, unedited answers..