n - State MDP

(a) $V^*(n) = ?$

$$V^*(s) = \max_{a} \sum_{s'} T(s,a,s') \left[R(s,a,s') + \gamma \cdot V^*(s') \right]$$

At state n, there is only one action (GoRight).

$$V^*(n) = 1.0 \times \left[(+10) + 0.5 \times V^*(n) \right]$$

$$\Rightarrow V^*(n) = \boxed{20}$$

Another approach to arrive at this conclusion is to view $V^*(n)$ as the maximum expected discounted reward that the agent can gather from state n.

$$\Rightarrow 10 + 10 \times \left(\frac{1}{2}\right) + 10 \times \underbrace{\left(\frac{1}{2}\right)^2}_{\text{discounting}} \cdots$$

arising from $n \xrightarrow{+10} n \xrightarrow{+10} n \xrightarrow{+10} n \cdots n \xrightarrow{+10} n$

$$\Rightarrow 10 \times \left(1 + \frac{1}{2} + \left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^3 \cdots \right)$$

$$\Rightarrow 10 \times \left[\frac{1}{1 - (1/2)} \right] = \boxed{20}$$

(b) Optimal value function $V^*(k)$ for all $k \in \{1, 2, \ldots (n-1)\}$

Given $V^*(k) < V^*(k+1)$ for this MDP

Examine,

$$V^*(n-1) = \max_{\{GoRight,\ Reset\}} \begin{cases} \text{For GoRight: } 1 + \frac{1}{2} \times V^*(n) \\ \text{For Reset: } 0 + \frac{1}{2} \times V^*(1) \end{cases}$$

Since, $V^*(1) < V^*(n)$

Hence, $\max\left[\left(1 + \frac{1}{2} \times V^*(n)\right), \left(\frac{1}{2} \times V^*(1)\right)\right]$

will return $\left[1 + \frac{1}{2} \times V^*(n)\right]$

Similarly,

$$V^*(n-2) = \max_{\{GoRight,\ Reset\}} \begin{cases} \text{For GoRight: } 1 + \frac{1}{2} \times V^*(n-1) \\ \text{For Reset: } 0 + \frac{1}{2} \times V^*(1) \end{cases}$$

Since, $V^*(1) < V^*(n-1)$

Hence, $\max\left[\left(1 + \frac{1}{2} \times V^*(n-1)\right), \left(\frac{1}{2} \times V^*(1)\right)\right]$

will return $\left[1 + \frac{1}{2} \times V^*(n-1)\right]$

$$V^*(n-2) = 1 + \frac{1}{2} \times V^*(n-1) = 1 + \frac{1}{2}\left(1 + \frac{1}{2} \times V^*(n)\right)$$

$$= 1 + \frac{1}{2} + \left(\frac{1}{2}\right)^2 V^*(n)$$

$$V^*(n-k) = 1 + \left(\frac{1}{2}\right) + \left(\frac{1}{2}\right)^2 \cdots \left(\frac{1}{2}\right)^{k-1} + \left(\frac{1}{2}\right)^k V^*(n)$$

$$= \left[\frac{1 - (1/2)^k}{1 - (1/2)}\right] + \left(\frac{1}{2}\right)^k V^*(n)$$

(c) Value Iteration initialised to zero for all states.

After 1^{st} iteration:

- $V_1(n) = 10 + \frac{1}{2} \times V_0(n) = 10 + \frac{1}{2} \times 0 = \boxed{10}$

- $V_1(n-1) = \max\limits_{\{GoRight,\ Reset\}} \begin{cases} \text{For GoRight: } 1.0 + \frac{1}{2} \times V_0(n) \\ \text{For Reset: } 0.0 + \frac{1}{2} \times V_0(1) \end{cases}$

$= \max \begin{Bmatrix} 1.0 \\ 0.0 \end{Bmatrix}$

$\Rightarrow \boxed{V_1(n-1) = 1}$

Similarly, for all states $k \in \{1, \dots n-1\}$,

$\boxed{V_1(n-k) = 1}$

After 2^{nd} iteration:

- $V_2(n) = 10 + \frac{1}{2} \times 10 = \boxed{15}$

- $V_2(n-1) = \max\limits_{\{GoRight,\ Reset\}} \begin{cases} \text{For GoRight: } 1 + \frac{1}{2} \times V_1(n) \\ \text{For Reset: } 0 + \frac{1}{2} \times V_1(1) \end{cases}$

$= \max\limits_{\{GoRight,\ Reset\}} \begin{cases} \text{For GoRight: } 1 + \frac{1}{2} \times 10 \\ \text{For Reset: } 0 + \frac{1}{2} \times 1 \end{cases}$

$= \max \begin{Bmatrix} 6 \\ 1 \end{Bmatrix}$

$\Rightarrow \boxed{V_2(n-1) = 6}$

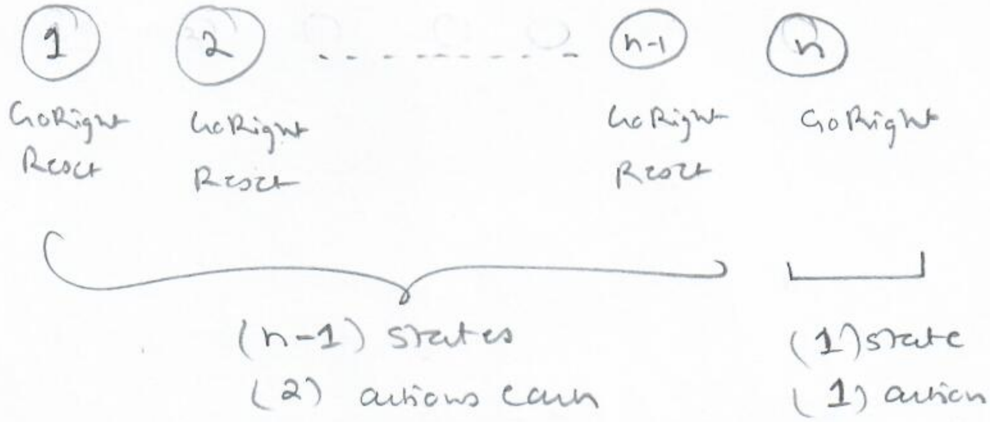
- $V_2(n-2) = \max\limits_{\{GoRight,\ Reset\}} \begin{cases} \text{For GoRight: } 1 + \frac{1}{2} \times V_1(n-1) \\ \text{For Reset: } 0 + \frac{1}{2} \times V_1(1) \end{cases}$

$= \max\limits_{\{GoRight,\ Reset\}} \begin{cases} \text{For GoRight: } 1 + \frac{1}{2} \times 1 \\ \text{For Reset: } 0 + \frac{1}{2} \times 1 \end{cases}$

$V_2(n-2) = \max\{1.5, 0.5\}$
$= \boxed{1.5}$

The same for $\boxed{V_2(n-k) = 1.5}$ $\forall k \in \{1, \dots n-1\}$

(d) How many policies are there for this MDP?



# of policies $2^{n-1} \times 1 = \boxed{2^{n-1}}$

An optimal policy:

$\pi(k) = \text{GoRight}$ for $k \in \{1, 2, \dots n\}$.