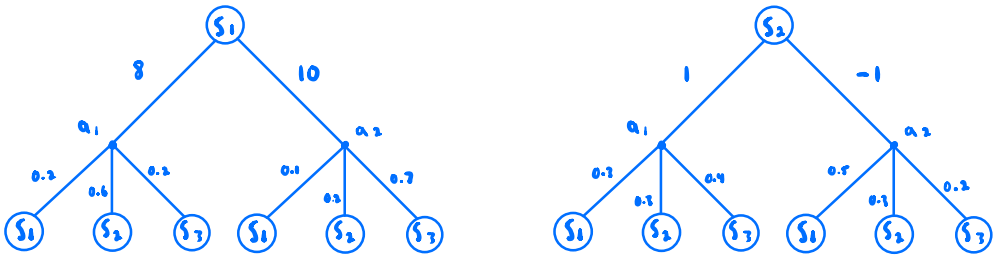
Problem 1 :

MDP with $S = \{s_1, s_2, s_3\}$, $T = \{s_3\}$, $A = \{a_1, a_2\}$

$P : N \times A \times S \longrightarrow [0, 1]$



Initialize Value functions as max over actions.

$\therefore \quad V_0(s_1) = 10, \quad V_0(s_2) = 1, \quad V_0(s_3) = 0$

Iteration 1 :

$q_1(s_1, a_1) = 8 + 0.2(10) + 0.6(1) + 0.2(0) = 10.6$

$q_1(s_1, a_2) = 10 + 0.1(10) + 0.2(1) + 0.7(0) = 11.2 \leftarrow$

$q_1(s_2, a_1) = 1 + 0.3(10) + 0.3(1) + 0.4(0) = 4.3 \leftarrow$

$q_1(s_2 \; a_2) = -1 + 0.5(10) + 0.3(1) + 0.2(0) = 4.3 \leftarrow$

$\therefore \quad V_1(s_1) = 11.2, \quad V_1(s_2) = 4.3, \quad V_1(s_3) = 0$

Iteration 2 :

$q_2(s_1, a_1) = 8 + 0.2(11.2) + 0.6(4.3) + 0.2(0) = 12.82 \leftarrow$

$q_2(s_1, a_2) = 10 + 0.1(11.2) + 0.2(4.3) + 0.7(0) = 11.98$

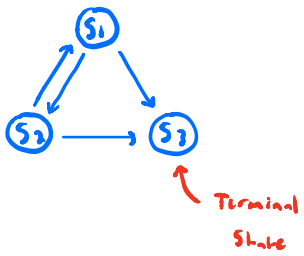
$q_2(s_2, a_1) = 1 + 0.3(11.2) + 0.3(4.3) + 0.4(0) = 5.65$

$q_2(s_2, a_2) = -1 + 0.5(11.2) + 0.3(4.3) + 0.2(0) = 5.89 \leftarrow$

$\therefore \quad v_2(s_1) = 12.82, \quad v_2(s_2) = 5.89, \quad v_2(s_3) = 0$

$\therefore \quad \pi^* = \{ s_1 : a_1, \quad s_2 : a_2 \}$

The transitions can be graphed as:



↳ The maximum distance to a terminal state is two, therefore only two iterations are required to converge on the optimal policy.

↳ Why is this the case?

Problem 3:

- Each day, a person is employed or unemployed
  - ↳ If employed, work job for the day and earn wage, but get fired with probability $\alpha \in [0,1]$, transitioning to unemployed state.
  - ↳ If unemployed, offered position at one of n jobs with salaries $w_1, w_2, \dots, w_n \in \mathbb{R}^+$ and probabilities $p_1, p_2, \dots, p_n \in [0,1]$ $(\sum_{i=1}^{n} p_i = 1)$
    - ↳ If accepting offer, proceed day as employed.
    - ↳ If rejecting offer, earn $w_0 \in \mathbb{R}^+$ unemployment for the day and enter next day as unemployed.

- **Objective:** Determine optimal choice of accepting or rejecting offers in a manner which maximizes infinite horizon expected, discounted sum of wage utility.
  - ↳ Discount Factor: $\gamma \in [0,1)$
  - ↳ Wage Utility: $U(w) = \log(w)$ for wage $w \in \mathbb{R}^+$

$$\Rightarrow \text{Maximize} \quad \mathbb{E}\left[\sum_{u=t}^{\infty} \gamma^{u-t} \cdot \log(w_{iu}) \right]$$

$t \equiv$ Start day

$w_{iu} \equiv$ Wage earned on day u in job i

- **State Space:** $S = \{(s_1, s_2)\}$ where $s_1 \in \{e_0, e_1, e_2, \dots, e_n\}$ and $s_2 \in \{e_1, e_2, \dots, e_n\}$ with e_0 representing unemployment and e_i representing employment at position i for $i \in \{1,2,\dots,n\}$. By this setup, if employed, $s_1 = s_2$

  $T = \{\}$ (there are no terminal states)

  $N = S$ (all states are non-terminal states)

- **Action Space:** $A = \{\text{accept offer}, \text{reject offer}\}$

- **Reward Function:** $R_T((s_i, s_j), a) : N \times A \longrightarrow \mathbb{R}$
  - ↳ **Curried form:** $R_T : N \to (A \to \mathbb{R})$
    - ↳ For each state (s_i, s_j), if you accept the offer of employment, you get the salary of position s_j
    - ↳ If you reject the offer at hand, you get the salary of position s_i
    - ↳ Recall, if employed, $s_i = s_j$ and action does not affect outcome

$$\therefore R((s_i, s_j), a) = \begin{cases} w_i & \text{if } a = \text{reject} \\ w_j & \text{if } a = \text{accept} \end{cases}$$

where $i \in \{0, 1, 2, \dots, n\}$, $j \in \{1, 2, \dots, n\}$

- **State Transition Function:** $P(s, a, s') : N \times A \times S \longrightarrow [0, 1]$
  - ↳ **Curried form:** $P : N \to (A \to (S \to [0, 1]))$
    - ↳ $P((s_i, s_j), a, (s_k, s_\ell)) = 0 \quad \forall\ s_i, s_j, s_k, s_\ell \in \{e_1, e_2, \dots, e_n\}$, $a \in A$ where not $s_i = s_j = s_k = s_\ell$.
    - ↳ $P((s_i, s_j), a, (s_k, s_\ell)) = 1 - \alpha \quad \forall\ s_i, s_j, s_k, s_\ell \in \{e_1, e_2, \dots, e_n\}$, $a \in A$ where $s_i = s_j = s_k = s_\ell$
    - ↳ $P((s_i, s_j), a, (s_k, s_\ell)) = 0 \quad \forall\ s_i, s_j, s_k, s_\ell \in \{e_0, e_1, \dots, e_n\}$, $a \in A$ if $s_j = e_0$ or $s_\ell = e_0$
    - ↳ $P((s_i, s_j), a, (s_k, s_\ell)) = \alpha p_\ell \quad \forall\ s_i, s_j, s_\ell \in \{e_1, e_2, \dots, e_n\}$, $s_k = e_0$, $a \in A$
    - ↳ $P((s_i, s_j), a, (s_k, s_\ell)) = \begin{cases} 1 & \text{if } a = \text{accept} \\ 0 & \text{if } a = \text{reject} \end{cases}$ $\forall\ s_i = e_0$, $s_j = s_k = s_\ell \in \{e_1, \dots, e_n\}$
    - ↳ $P((s_i, s_j), a, (s_k, s_\ell)) = 0 \quad \forall\ a \in A$, $s_i = e_0$, $s_j, s_k, s_\ell \in \{e_1, \dots, e_n\}$ where not $s_j = s_k = s_\ell$.

- **Bellman Optimality Equation:**

$$V^*((s_i, s_j)) = \max_{a \in A} \left\{ R((s_i, s_j), a) + \gamma \sum_{(s_k, s_\ell) \in N} P((s_i, s_j), a, (s_k, s_\ell)) \cdot V^*((s_k, s_\ell)) \right\}$$

$\forall\ s_i, s_j \in N$