1. (30, 30%) Consider an undiscounted Markov Reward Process with two states $A$ and $B$. The transition matrix and reward function are unknown, but you have observed two sample episodes:

$A + 3 \to A + 2 \to B - 3 \to A + 4 \to B - 4 \to$ terminate

$B - 2 \to A + 3 \to B - 4 \to$ terminate

In the above episodes, sample state transitions and sample rewards are shown at each step, e.g., $A + 3 \to A$ indicates a transition from state $A$ to state $A$, with a reward of $+3$.

(a) (5, 10%) Using first-visit Monte-Carlo evaluation, estimate the state-value function $V(A), V(B)$.

$$V(A) = \frac{(3 + 2 - 3 + 4 - 4) + (3 - 4)}{2}$$

$$= \frac{2 - 1}{2} = 0.5$$

$$V(B) = \frac{(-3 + 4 - 4) + (-2 + 3 - 4)}{2}$$

$$= \frac{-3 - 3}{2} = -3$$

(c) (5, 10%) Draw a diagram of the Markov Reward Process that best explains these two episodes (i.e. the model that maximises the likelihood of the data - although it is not necessary to prove this fact). Show rewards and transition probabilities on your diagram.

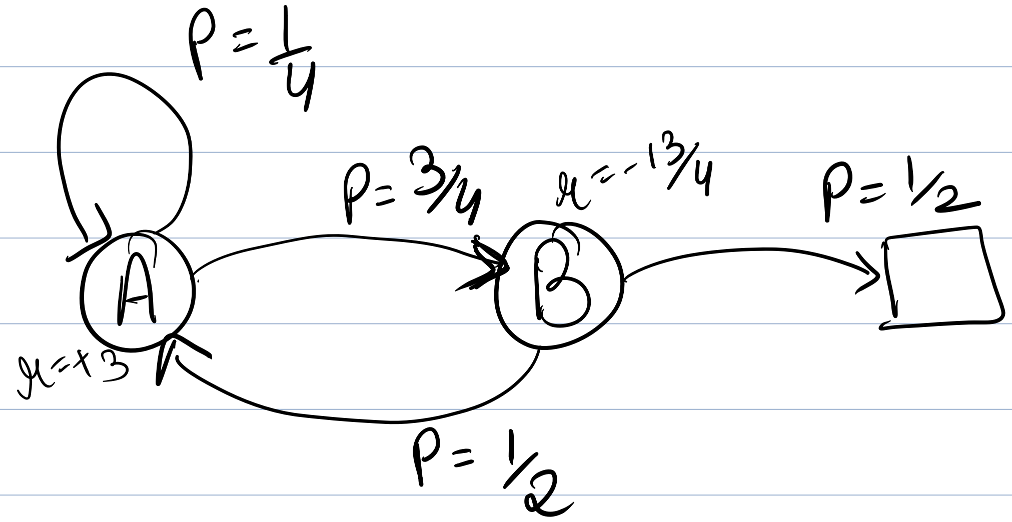
$A \xrightarrow{+3} A,\ A \xrightarrow{+2} B,\ B \xrightarrow{-3} A,\ A \xrightarrow{+4} B,\ B \xrightarrow{-4} \square$ term

$B \xrightarrow{-2} A,\ A \xrightarrow{+3} B,\ B \xrightarrow{-4} \square$ term

reward = avg??

In state A = 4 times $\frac{1}{4}$ back to A, $\frac{3}{4}$ to B

In state B = 4 times $\frac{2}{4}$ terminates, $\frac{2}{4}$ to A




(d) (5, 10%) Define the Bellman equation for your above Markov reward process. Solve the Bellman equation **directly**, rather than iteratively, to find the true state-value function $V(A), V(B)$

$$V(s) = E\{G_t \mid S_t = s\}$$

undiscounted

$$= E\{r_{t+1} + V(s_{t+1}) \mid S_t = s\} \quad (\gamma = 1)$$

$$= E\{r_{t+1} \mid S_t = s\} + E\{V(S_{t+1}) \mid S_t = s\}$$

$$= r_s + \sum_{s'} p(s' \mid s)\, V(s')$$

$$\therefore V(A) = 3 + \frac{1}{4} V(A) + \frac{3}{4} V(B) \quad — \ (1)$$

$$\therefore V(B) = -\frac{13}{4} + \frac{1}{2} V(A) \quad — \ (2)$$

(1)

$$V(A) - \frac{1}{4}V(A) = 3 + \frac{3}{4}V(B)$$

$$3V(A) = 12 + 3V(B)$$

$$V(A) = 4$$

(2)

$$V(B) = -\frac{13}{4} + \frac{1}{2}V(A)$$

$$\Rightarrow V(B) = -\frac{13}{4} + 2 + \frac{1}{2}V(B)$$

$$\Rightarrow \frac{1}{2}V(B) = -\frac{5}{4}$$

$$\Rightarrow V(B) = -\frac{5}{2} = \underline{-2.5}$$

$$\therefore V(A) = 4 - 2.5 = \underline{1.5}$$