Monte Carlo vs Temporal Difference

(an episode can be represented as a character dying and the steps he took)

- With Monte Carlo, we update the value function from a complete episode, and so we use the actual accurate discounted return of this episode
- with TD, we update the value function from a step, so we replace G_t with an estimated return

$$MC: V(S_t) \leftarrow V(S_t) + \alpha[G_t - V(S_t)]$$

α is learning rate

$$TD: V(S_t) \leftarrow V(S_t) + \alpha[R_{(t+1)} + \gamma V(S_{(t+1)}) - V(S_t)]$$

Q-Learning

Q-Learning is the algorithm used to train a Q-Function which is an action-value function.

Internally, our Q-Function has a Q-table, a table where each cell corresponds to a state-action pair value.

example:

S \ A	←	→	↑	↓
S_0	0	0.1	10	-4
S_1	0	0	10	0
S_2	5	-3	0	0.5
S_3	3	-1	1.5	0
S_4	1	2	-3	1
S_5	0	-2	3	0