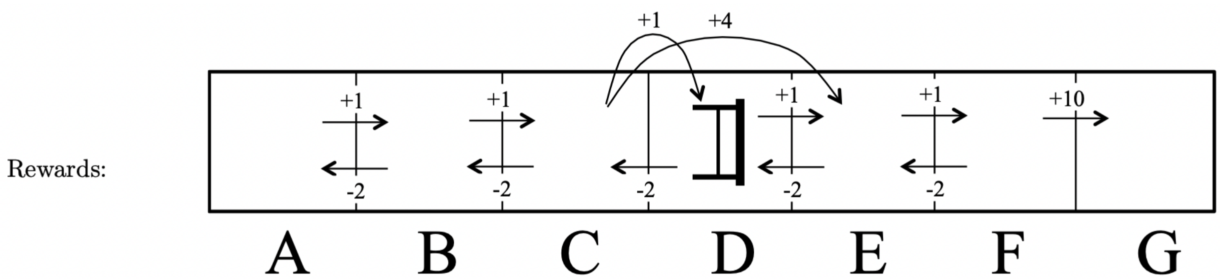
DS 102 Discussion 9
Wednesday, April 6, 2022

1. **A Track Meet for Robots**

Consider an MDP modeling a hurdle race track, shown below.[1] There is a single hurdle in square D. The terminal state is G. The agent can run `left` or `right`. If the agent is in square C, it cannot run `right`. Instead, it can jump, which either results in a fall to the hurdle square D, or a successful hurdle jump to square E. Rewards are shown below. Assume that the discount factor γ is 1. Note that the agent cannot move `right` at state C.



The possible actions are:

- `right`: Deterministically move to the right
- `left`: Deterministically move to the left
- `jump`: Stochastically jump to the right. This action is only available in state C. The transition probabilities are given by:

$$T(C, \texttt{jump}, E) = 0.5$$
$$T(C, \texttt{jump}, D) = 0.5$$

(a) *Computing $V^\pi(s)$*

For the policy π of always moving forward (i.e., using actions `right` or `jump`), compute $V^\pi(C)$.

[1]Source: UC Berkeley CS 188, Fall 2011

(b) *Value Iteration by Hand*

Perform two rounds of value iteration on the MDP and compute $V_2(B)$, $Q_2(B, \texttt{right})$, and $Q_2(B, \texttt{left})$. At iteration 0, assume all values have been initialized to 0.

(c) *Q-Learning by Hand*

Suppose you are given the trajectories in the table below. Derive the relevant Q-value updates using Q-learning with a fixed learning rate α of 0.5. Assume all Q-values are initialized to 0.

s	a	r	s	a	r	s	a	r	s	a	r	s
C	*jump*	+4	E	*right*	+1	F	*left*	-2	E	*right*	+1	F

2. **Bounding the Chance of Extreme Events**

   Suppose you are a data scientist working with a city planning organization to assess the risk of major earthquakes on infrastructure. The city planning team informs you that many of the city's buildings will be severely damaged if struck by 10 or more high-magnitude earthquakes over a 50-year period. You are given the following task: make a recommendation for significant investment into earthquake protection if the chance of severe damage can exceed 2%.

   In this problem, let X be a random variable representing the number of high-magnitude earthquakes over a 50-year period.

   (a) *An Initial Bound*

   Suppose you collect historical data on the occurrence of earthquakes in the area and estimate that the mean is 2 high-magnitude earthquakes over a 50-year period. Assuming this is the true mean, find a bound on the chance of severe damage to buildings via earthquakes. If this is the only information you had available, what recommendation would you make to the city planning organization?

   Hint: The distribution of X is unknown.

(b) *A Tighter Bound*

Not satisfied with your initial research, you additionally estimate that the variance of the number of high-magnitude earthquakes over a 50-year period is 2. Assuming this is the true variance, find a tighter upper bound on the chance of severe damage to buildings via earthquakes. Does this change your recommendation to the city planning organization?

(c) *A Final Bound*

To make your estimation even better, you need to make some assumptions. Specifically, you notice that the mean and the variance of X match, which is characteristic of a Poisson random variable. As a result, you believe that the moments of the distribution of X match that of a Poisson random variable. Given that this assumption holds, find an even tighter upper-bound on the chance of severe damage to buildings via earthquakes. What is your final recommendation to the city planning organization?

Hint: Recall that if $X \sim \text{Poisson}(\lambda)$, then $\mathbb{P}[X = k] = \frac{e^{-\lambda}\lambda^k}{k!}$.