Chapter 5: Monte Carlo Methods

5.1 Monte Carlo Prediction

- **First-visit MC method**: the return is taken by the first occurrence of s in the episode.
- **Every-visit MC method**: the return is taken by every occurrences of s in the episode.

In first-visit MC method, every sample of $V(s)$ are independent with each other.

1. Estimates for states are independent. (do not *bootstrap*)
2. Ability to learn from actual experience and from simulated experience.
3. Its computational expense of estimating the value of
   a single state is independent of the number of states.

5.2 Monte Carlo Estimation of Action Values

Estimation of action values is useful when the model is not completely known.

Complication (vs estimation for value function): many state–action pairs may never be visited. $\Rightarrow$ *maintain exploration*

- **Exploring starts**: specifying that the episodes start in a state–action pair, and that every pair has a nonzero probability of being selected as the start.
- Consider only policies that are stochastic with a nonzero probability of selecting all actions in each state.

5.3 Monte Carlo Control

Monte Carlo ES (with exploring starts): alternate between evaluation and improvement on an episode-by-episode basis

5.4 Monte Carlo Control without Exploring Starts

- **On-policy methods**: to evaluate or improve the policy that is used to make decisions.
- **Off-policy methods**: to evaluate or improve the policy different from the policy that is used to make decisions.