Exercises
Reinforcement Learning: Chapter 5

Exercise 5.1. Consider the diagrams on the right in Figure 5.2. Why does the value function jump up for the last two rows in the rear? Why does it drop off for the whole last row on the left? Why are the frontmost values higher in the upper diagrams than in the lower?

Exercise 5.2. What is the backup diagram for Monte Carlo estimation of $Q^\pi$?

Exercise 5.4. Modify the algorithm for first-visit MC policy evaluation (Figure 5.1) to use the incremental implementation for stationary averages described in Section 2.5.

Exercise 5.5. Derive the weighted-average update rule (5.5) from (5.4). Follow the pattern of the derivation of the unweighted rule (2.4) from (2.1).

Exercise 5.6. Modify the algorithm for the off-policy Monte Carlo control algorithm (Figure 5.7) to use the method described above for incrementally computing weighted averages.

Exercise 5.7. Let $p$ be a density over an interval $[a, b]$ of the real line, $f : [a, b] \to \mathbb{R}$ be integrable and assume that we want to compute the integral $I(p) = \int f(x)p(x)dx$. If $X_1, \ldots, X_n$ is an i.i.d. sample from $p$, we know that $I_n = 1/n \sum_{i=1}^{n} f(X_i)$ approximates $\mathbb{E}[f(X)] = \int f(x)p(x)dx$ well when $n$ is large enough. In particular, we know that by the law of large numbers, $I_n \to I(p)$ w.p.1. since the random variable $Z = f(X)$ has a finite expectation.

Often sampling from $p$ is not feasible, but we can sample from a “similar” density $q$. One method (called importance sampling) then computes

$$J_n = \frac{1}{n} \sum_{i=1}^{n} f(X_i) \frac{p(X_i)}{q(X_i)},$$

where now $\{X_i\}_i$ is an i.i.d. sample from $q$. Assume that if $p(x) > 0$ then also $q(x) > 0$ holds. Show that in this case (i) $\mathbb{E}[J_n] = I(p)$ and (ii) $J_n \to I(p)$ w.p.1 as $n \to \infty$.

Exercise 5.8. The previous exercise was considered with importance sampling. This time let us consider the so-called weighted importance sampling method. For this let $p, q, f, X_i$ be as in the the previous exercise and consider

$$K_n = \frac{\sum_{i=1}^{n} f(X_i)w(X_i)}{\sum_{i=1}^{n} w(X_i)},$$

where $w(X_i) = p(X_i)/q(X_i)$. Show that $K_n$ converges to $I(p)$ w.p.1 as $n \to \infty$. 

1
**Hint:** Remember that if $a_n \to a$ and $b_n \to b$, $b \neq 0$ then $a_n/b_n \to a/b$. Combine this with the result of the previous exercise.

**Exercise 5.9.** (Programming) Compare importance sampling and weighted importance sampling in the following simple setting: Let $p$ be the density of the standard normal distribution but truncated to $[-1, 1]$ (to generate a sample from $p$ you would draw a sample from the standard normal distribution and accept it only if it is inside the interval $[-1, 1]$, otherwise repeat the procedure). Let $q$ be the density of the uniform distribution over $[-1, 1]$. Let $f(x) = x + 1$. Write a program that estimates these procedures bias and variance, as a function of the number of samples. Plot the results and draw conclusions.