Date created: 2025-07-14-Mon

2.2 Action-value Methods

Action-value methods are those for estimating the values of actions and for using the estimates to make action selection decisions.

Sample-average Estimate

Recall that the true value of an action is defined as the expected reward when it is selected. One natural way to estimate it is to take the sample mean of rewards received:

$$Q_t(a) := \frac{\sum_{i=1}^{t-1} R_i \mathbb{1}_{A_i=a}}{\sum_{i=1}^{t-1} \mathbb{1}_{A_i=a}}$$

If $\sum_{i=1}^{t-1} \mathbb{1}_{A_i=a} = 0$, we define some default value for the initial estimate. It is easy to see that as t approaches 0, the estimate converges to the true expectation by the Law of Large Numbers.

Greedy Action Selection

Our action selection is *greedy* if we select the action with the highest estimated value:

$$A_t := \underset{a}{\text{argmax}} Q_t(a)$$