

# Chapter: X

Jacob Taylor Cassady

August 26, 2023

## 1 Exercise 0

Hello world.

### 1.1 Exercise 0.0

Example Equation - Taken from “Reinforcement Learning” by Sutton and Barto:

$$Q_t(a) \doteq \frac{\text{sum of rewards when } a \text{ taken prior to } t}{\text{number of times } a \text{ taken prior to } t} = \frac{\sum_{i=1}^{t-1} R_i \cdot \mathbb{1}_{A_i=a}}{\sum_{i=1}^{t-1} \mathbb{1}_{A_i=a}} \quad (2.1)$$

where  $\mathbb{1}_{\text{predicate}}$  denotes the random variable that is 1 if *predicate* is true and 0 if it is not. If the denominator is zero, then we instead define  $Q_t(a)$  as some default value, such as 0. As the denominator goes to infinity, by the law of large numbers,  $Q_t(a)$  converges to  $q_*(a)$ . We call this the *sample-average* method for estimating action values because each estimate is an average of the sample of relevant rewards.

