Date created: 2025-07-16-Wed

2.5 Tracking a Nonstationary Problem

Exponential recency-weighted average

In many RL problems, the action values are nonstationary. In such cases, it makes sense to give more weight to recent rewards than to long-past rewards. We can achieve this by setting the step size parameter to a constant $\alpha \in (0, 1]$:

$$Q_{n+1} := Q_n + \alpha[R_n - Q_n]$$

This results in Q_{n+1} to be a weighted average of past rewards and the initial estimate Q_1:

$$\begin{aligned} Q_{n+1} &= Q_n + \alpha[R_n - Q_n] \\ &= \alpha R_n - (1-\alpha)Q_n \\ &= \alpha R_n - (1-\alpha)(\alpha R_{n-1} - (1-\alpha)Q_{n-1}) \\ &= \alpha R_n - (1-\alpha)\alpha R_{n-1} - (1-\alpha)^2 Q_{n-1}) \\ &\vdots \\ &= (1-\alpha)^n Q_1 + \sum_{i=1}^{n} \alpha(1-\alpha)^{n-i} R_i \end{aligned}$$

This is a weighted average because $(1-\alpha)^n + \sum_{i=1}^{n} \alpha(1-\alpha)^{n-i} = 1$, which can be proved by induction. Note that the weight $\alpha(1-\alpha)^{n-i}$ on R_i depends on how many rewards ago, $n-i$, it was observed. This weight decays exponentially, thus this is also known as *exponential recency-weighted average*.

Variable step size

Sometimes it is convenient to vary the step size on each step. Let $\alpha_n(a)$ denote the step size after the nth selection of action a. Note that $\alpha_n(a) = \frac{1}{n}$ is the sample-average method, which guarantees convergence to the true action values by the Law of Large Numbers. In general, a well-known result from stochastic approximation theory gives us the following convergence conditions:

$$\sum_{n=1}^{\infty} \alpha_n(a) = \infty, \text{ and } \sum_{n=1}^{\infty} \alpha_n^2(a) < \infty$$

The first condition ensures that steps are large enough to overcome initial conditions or random fluctuations, while the second condition ensures that eventually the steps become small enough to converge.

Convergence is not always a desired property, as in nonstationary problems we want continuous updates to our estimates. These convergence conditions provide theoretical guarantees, but in practice are seldom checked since it often requires tedious tuning to find a satisfactory convergence rate.

Exercises

2.4:

If the step-size parameters, α_n, are not constant, then the estimate Q_n is a weighted average of previously received rewards with a weighting different from that given by (2.6). What is the weighting on each prior reward for the general case, analogous to (2.6), in terms of the sequence of step-size parameters?

Solution

$$\begin{aligned} Q_{n+1} &= Q_n + \alpha_n[R_n - Q_n] \\ &= \alpha_n R_n - (1-\alpha_n)Q_n \\ &= Q_1 \prod_{i=1}^{n}(1-\alpha_i) + \sum_{i=1}^{n} \alpha_i \prod_{j=i+1}^{n} (1-\alpha_j) R_i \end{aligned}$$

2.5: Programming

Design and conduct an experiment to demonstrate the difficulties that sample-average methods have for nonstationary problems. Use a modified version of the 10-armed testbed in which all the $q_*(a)$ start out equal and then take independent random walks (say by adding a normally distributed increment with mean zero and standard deviation 0.01 to all the $q_*(a)$ on each step). Prepare plots like Figure 2.2 for an action-value method using sample averages, incrementally computed, and another action-value method using a constant step-size parameter, $\alpha = 0.1$. Use $\epsilon = 0.1$ and longer runs, say of 10,000 steps.

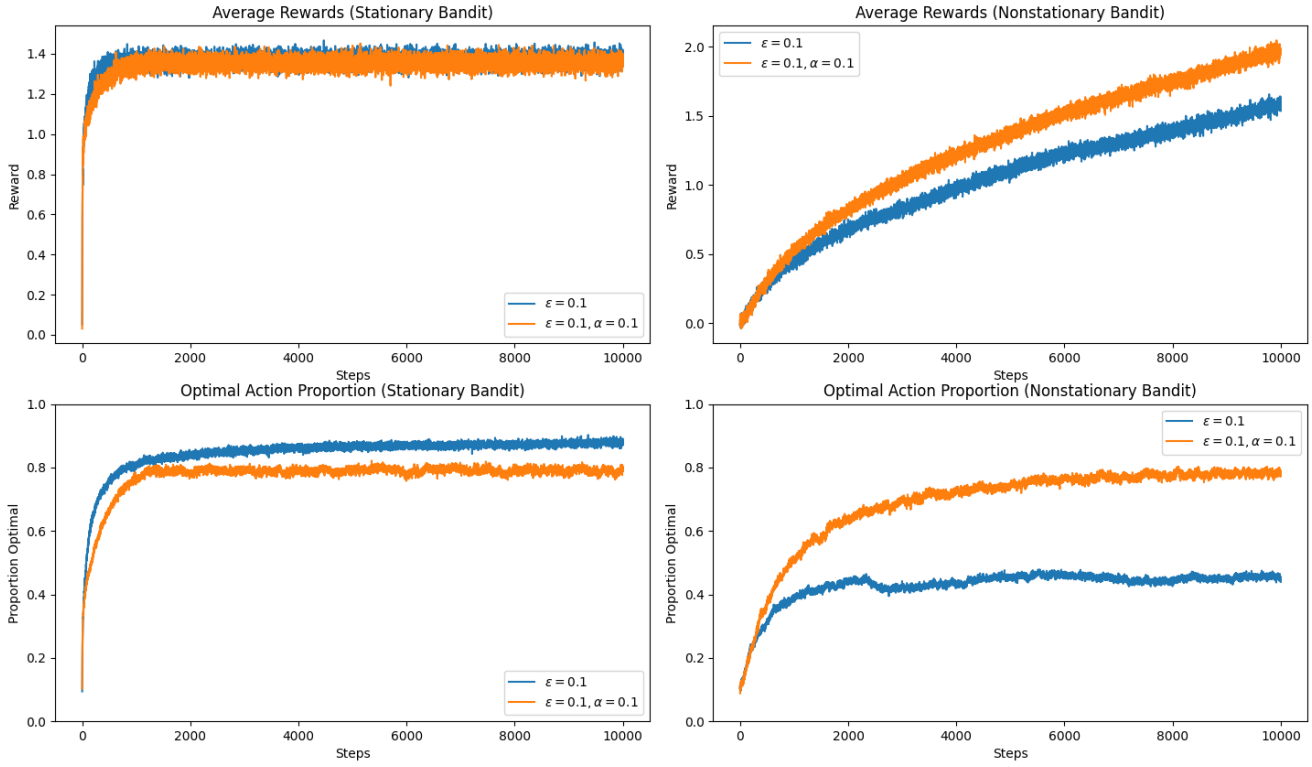

Figure 1: Experiment Results

Solution As shown in the figure, the non-constant step size method becomes less and less able to adapt to the nonstationary environment as time goes on.