Reinforcement Learning
HW1
By Madhur Tandon
(2016053)

1) Comparison between Constant Step Size and Sample Averages

   Analysis:

   Constant Step Size gives more weight to recent rewards and thus easily adapts and surpasses the method using Sample Averages

2) Mysterious Spikes

   Optimistic Initial Values encourage exploration in the initial phase as every action when explored is disappointing in some sense.

   as Q(n) = Q(n-1) + alpha*(R(n) - Q(n-1))

   Since Q(n-1) is large (because optimistic initial value), Q(n) decreases.

   Thus, the agent chooses the action with max Q(n) which is essentially the action which hasn't been tried before.

   After K turns, all possible K actions would have been chosen (on an average). Thus, in the (K+1)th turn, the action which gives the maximum reward will be chosen by the agent.

   On an average, this action is the optimal action i.e. the one that gives max(q*) thereby resulting in a spike.

   <u>*Non-stationary case*</u>:

   The Mysterious Spike disappears and Optimistic Initial Value method performs equally (if not better) to Epsilon Greedy. This is because on the (K+1)th turn, the optimal action aka the one that gives max(q*) has changed since q* has changed itself.

3) Making constant step size independent of Q(0)

of $\beta_n = \frac{\alpha}{\bar{o}_n}$ and $\bar{o}_n = \bar{o}_{n-1} + \alpha(1-\bar{o}_{n-1})$ — Ⓐ

for $n \geq 0$ with $\bar{o}_0 = 0$

└ ①

Also, $\theta_n = \theta_{n-1} + \beta_n (R_n - \theta_{n-1})$ — ②

Put ① in ②

$$\theta_n = \theta_{n-1} + \frac{\alpha}{\bar{o}_n}(R_n - \theta_{n-1})$$

Substituting $\bar{o}_n$ from Ⓐ

$$\theta_n = \theta_{n-1} + \frac{\alpha(R_n - \theta_{n-1})}{\bar{o}_{n-1} + \alpha(1-\bar{o}_{n-1})}$$

Put $n = 1$

$$\therefore \theta_1 = \theta_0 + \frac{\alpha(R_1 - \theta_0)}{\bar{\theta}_0 + \alpha(1-\bar{\theta}_0)}$$

From Ⓐ, we know that $\bar{\theta}_0 = 0$

$$\therefore \theta_1 = \theta_0 + \frac{\alpha(R_1 - \theta_0)}{0 + \alpha(1-0)}$$

$$\theta_1 = \theta_0 + \frac{\cancel{\alpha}(R_1 - \theta_0)}{\cancel{\alpha}}$$

$$\theta_1 = \cancel{\theta_0} + R_1 - \cancel{\theta_0}$$

$$\therefore \theta_1 = R_1$$

Now, since $\theta_1 = R_1$

$\therefore \theta_1$ does not depend on θ_0

$\Rightarrow \theta_1$ does not have initial bias

$\therefore \theta_2$ does not have initial bias (since θ_2 depends on θ_1)

In a similar way,

θ_n does not have initial bias

$\therefore \theta_n$ is an exponential recency-weighted average without initial bias.