Reinforcement Learning

Assignment - 1

- Arnav Kumar
- 2016228

Q1

Due to the law of large numbers, our estimated Q converges to Q^* in stationery case but in non-stationery case, the value of Q^* changes at every step. For that we use constant step-size (α) instead of averaging over all the results.
Constant step-size gives more weight to recent result.

Q3 Incremental Rule is given by:

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

Let $\beta_n = \frac{\alpha_n}{\Delta_n}$ with $\Delta_n = \Delta_{n-1} + \alpha(1 - \Delta_{n-1})$, $\Delta_0 = 0$

If we use β_n instead of α as step size

we get

$$\begin{aligned}
Q_{n+1} &= Q_n + \beta_n [R_n - Q_n] \\
&= \beta_n R_n + (1-\beta_n) Q_n \\
&= \beta_n R_n + (1-\beta_n)\left(Q_{n-1} + \beta_{n-1}[R_{n-1} - Q_{n-1}]\right) \\
&= \beta_n R_n + (1-\beta_n)\left(\beta_{n-1} R_{n-1} + (1-\beta_{n-1}) Q_{n-1}\right) \\
&= \beta_n R_n + (1-\beta_n)\beta_{n-1} R_{n-1} + (1-\beta_n)(1-\beta_{n-1}) Q_{n-1} \\
&\;\;\vdots \\
&= \beta_n R_n + (1-\beta_n)\beta_{n-1} R_{n-1} + (1-\beta_n)(1-\beta_{n-1})\beta_{n-2} R_{n-2} + \dots \\
&\quad \dots \prod_{i=0}^{n-1} (1 - \beta_{n-i}) Q_1
\end{aligned}$$

As, $\Delta_0 = 0$, $\Delta_1 = 0 + \alpha(1-0) = \alpha$

$\therefore \boxed{\beta_1 = 1}$

So, $\prod_{i=0}^{n-1} (1 - \beta_{i-n}) = 0$, Thus, initial Bias is eliminated.