Question 3

When we use sample average as our action value method, Q_n is independent of Q_1.

However, if we use a constant step size parameter, we have

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

$$Q_{n+1} = \alpha R_n + (1-\alpha) Q_n$$
$$= \alpha R_n + (1-\alpha)(\alpha R_{n-1} + (1-\alpha) Q_{n-1})$$
$$= (1-\alpha)^n Q_1 + \sum_{i=1}^{n} \alpha (1-\alpha)^{n-i} R_i$$

Therefore we can see that Q_{n+1} still depends on Q_1 irrespective of the actions picked by the agent.

To get an independent estimate using a constant step size, we can use

$$\beta_n = \frac{\alpha}{\bar{o}_n} \quad (\alpha \Rightarrow \text{constant step size})$$

$$\bar{o}_n = \bar{o}_{n-1} + \alpha (1 - \bar{o}_{n-1}) \text{ for } n > 0$$

$$\bar{o}_0 = 0$$

Scanned with CamScanner

$$Q_2 = Q_1 + \beta_1 [R_1 - Q_1]$$

$$\beta_1 = \frac{\alpha}{Q_1} \quad \frac{\alpha}{(1-\alpha) \times Q_0 + \alpha} \quad \frac{\alpha}{\alpha} = 1$$

$$\therefore Q_2 = Q_1 + R_1 - Q_1$$

$$= R_1$$

∴ We can see that Q_2 is independent of our initial estimate Q_1 and only depends on R_1 [reward]

We can extend it further :

$$Q_3 = Q_2 + \beta_2 [R_2 - Q_2]$$

$$\beta_2 = \frac{\alpha}{Q_2} = \frac{\alpha}{(1-\alpha) \times Q_1 + \alpha} = \frac{\alpha}{2\alpha - \alpha^2} = \frac{1}{2-\alpha}$$

$$\therefore Q_3 = Q_2 + \frac{1}{2-\alpha} [R_2 - Q_2]$$

$$= \frac{R_2}{2-\alpha} + Q_2 \left[1 - \frac{1}{2-\alpha}\right]$$

$$= \frac{R_2}{2-\alpha} + Q_2 \left[\frac{1-\alpha}{2-\alpha}\right]$$

Scanned with CamScanner

We know that $Q_2 = R_1$,

$$\therefore Q_3 = \frac{R_3}{2-\alpha} + R_1 \left[\frac{1-\alpha}{2-\alpha}\right]$$

which is independent of Q_1.

Similarly we can know for $Q_4, Q_5 \ldots\ldots Q_n$.

Hence $\beta = \frac{\alpha}{Q_n}$, $Q_n = Q_{n-1}(1-\alpha) + \alpha$

is a parameter that removes bias and provides constant step size.

Scanned with CamScanner