Q2.7

Solution

The step-size given :-

$$\beta_n = \alpha / \bar{o}_n$$

to process the nth reward for a particular action, where $\alpha > 0$ is a conventional constant step size & $\bar{o}_n$ is a trace of 1 that starts at 0.

& $\bar{o}_n = \bar{o}_{n-1} + \alpha(1 - \bar{o}_{n-1})$ for $n \geq 0$, $\bar{o}_0 = 0$

then for values of $n = 0, 1, 2, \ldots$

$\bar{o}_0 = 0$

$\bar{o}_1 = 0 + (1-0) \times \alpha = \alpha$, $\beta_1 = \frac{\cancel{\alpha}}{\cancel{\alpha}} = 1$

$\bar{o}_2 = \alpha + \alpha(1-\alpha) = 2\alpha - \alpha^2$, $\beta_2 = \frac{\alpha}{\alpha(2-\alpha)} = \frac{1}{2-\alpha}$

$\bar{o}_3 = 2\alpha + \alpha(1-2\alpha) = 3\alpha - 2\alpha^2$, $\beta_3 = \frac{\cancel{\alpha}}{\cancel{\alpha}(3-2\alpha)} = \frac{1}{3-2\alpha}$

$\vdots$

$\bar{o}_n = n\alpha - (n-1)\alpha^{n-1}$, $\beta_n = \frac{1}{n-(n-1)\alpha}$ — (1)

This tells that step size changes with every time step.

Now, the general expression for estimating Q_{n+1} is given by :-

$Q_{n+1} = Q_n + \alpha(R_n - Q_n)$ — (2)

But here $\alpha = \beta_n$ (changing every time step).

then (2) becomes :-

$$\begin{aligned}
Q_{n+1} &= Q_n + \beta_n[R_n - Q_n] \\
&= (1-\beta_n)Q_n + \beta_n R_n \\
&= (1-\beta_n)\left[(1-\beta_{n-1})Q_{n-1} + \beta_{n-1}R_{n-1}\right] + \beta_n R_n \\
&= (1-\beta_n)(1-\beta_{n-1})Q_{n-1} + \beta_{n-1}R_{n-1}(1-\beta_n) + \beta_n R_n \\
&= (1-\beta_n)(1-\beta_{n-1})\left[Q_{n-2}(1-\beta_{n-2}) + \beta_{n-2}R_{n-2}\right] + \\
&\quad \beta_{n-1}R_{n-1}(1-\beta_n) + \beta_n R_n
\end{aligned}$$

$$= (1-\beta_n)(1-\beta_{n-1})(1-\beta_{n-2})\,Q_{n-2} + \beta_{n-2}R_{n-2}(1-\beta_{n-1})(1-\beta_n)$$
$$+ \beta_{n-1}R_{n-1}(1-\beta_n) + \beta_n R_n$$

$$\vdots$$

$$\boxed{Q_{n+1} = \underbrace{Q_1 \cdot \prod_{i=1}^{n}[1-\beta_i]}_{\text{Term 1}} + \underbrace{\sum_{j=1}^{n}\left[\beta_i R_i \prod_{j=i+1}^{n}(1-\beta_j)\right]}_{\text{Term 2}}} \quad \text{— (3)}$$

From Term 1 of (3)

$$= Q_1 \cdot \prod_{i=1}^{n}(1-\beta_i) \quad \text{— (4)}$$

Substituting values from (1) in (4) by expanding (4)

$$= Q_1 (1-\beta_1)(1-\beta_2)(1-\beta_3)\cdots(1-\beta_n)$$

$$= Q_1 \overset{0}{\cancel{(1-1)}}\left(1-\frac{1}{2-\alpha}\right)\left(1-\frac{1}{3-2\alpha}\right)\cdots\left[1-\frac{1}{n-(n-1)\alpha}\right]$$

Hence Term (1) of (3) becomes 0 which means that action-value estimation given by this step size is independent of Q_1 i.e. the initial estimates ~~for this equals~~ and hence this does not bias with initial values holding good for non-stationary case as well.

Now, equation 3 becomes:-

$$Q_{n+1} = \sum_{i=1}^{n}\left[\beta_i R_i \prod_{j=i+1}^{n}(1-\beta_j)\right] \quad \text{— (5)}$$

Further, expanding ⑤ gives :-

$$Q_{n+1} = \sum_{i=1} \beta_1 R_1 (1-\beta_2)(1-\beta_3) \cdots (1-\beta_n) + \beta_2 R_2 (1-\beta_3)(1-\beta_4) \ldots (1-\beta_n) + \ldots + \beta_{n-1} R_{n-1} (1-\beta_n) + \beta_n R_n \quad (6)$$

Substituting values of β from ① in ⑤ gives :-

$$= R_1 \times 1 \times \left(1 - \frac{1}{2-\alpha}\right)\left(1 - \frac{1}{3-2\alpha}\right) \cdots \left(1 - \frac{1}{n-(n-1)\alpha}\right) +$$

$$R_2 \times \frac{1}{(2-\alpha)} \times \left(1 - \frac{1}{3-2\alpha}\right) \cdots \left(1 - \frac{1}{n-(n-1)\alpha}\right) + \ldots +$$

$$+ \frac{1}{(n-1)-(n-2)\alpha} R_{n-1} \left(1 - \frac{1}{n-(n-1)\alpha}\right) + \frac{1}{n-(n-1)\alpha} \times R_n \quad (7)$$

This 7 can be written as :-

$$Q_{n+1} = \sum_{i=1}^{n} \left[\frac{R_i (1-\alpha)^{n-i}}{i-(i-1)\alpha} * \prod_{j=i}^{n-1} \frac{(j)}{(j+1) - j\alpha} \right]$$

Hence

$$\boxed{Q_{n+1} = \sum_{i=1}^{n} \left[\frac{R_i (1-\alpha)^{n-i}}{i-(i-1)\alpha} \prod_{j=i}^{n-1} \frac{j}{(j+1)-j\alpha} \right]} \quad (8)$$

This 8 indicates that estimating value at n+1 is independent of initial estimates Q_1 & another point is that it is an exponential recency-weighted average.