Q3. Sample averages assign equal weights to all the rewards which does not work well in non stationary problems as the environment is constantly changing.

Step size $= \beta_n = \frac{\alpha}{\bar{O}_n}$

$$Q_{n+1} = Q_n + \beta_n (R_n - Q_n)$$

$$= Q_n + \frac{\alpha}{\bar{O}_n} (R_n - Q_n)$$

$$= Q_n \left(1 - \frac{\alpha}{\bar{O}_n}\right) + R_n \frac{\alpha}{\bar{O}_n}$$

$$\Rightarrow \bar{O}_n Q_{n+1} = Q_n (\bar{O}_n - \alpha) + R_n \alpha.$$

$$\bar{O}_n = \bar{O}_{n-1} + \alpha(1 - \bar{O}_{n-1})$$

$$\therefore \bar{O}_n Q_{n+1} = \bar{O}_{n-1}(1-\alpha)Q_n + \alpha R_n$$

Further:

$$Q_{n+1} = (1-\alpha)^n Q_1 + \sum \alpha (1-\alpha)^{n-1} R_i$$

$$\therefore \bar{O}_n Q_{n+1} = (1-\alpha)^n \bar{O}_0 Q_1 + \sum \alpha (1-\alpha)^{n-1} R_i$$

($\bar{O}_0 \to 0$)

$$= \sum \alpha (1-\alpha)^{n-1} R_i$$

which is free of initial bias