Q4 Show that the sample mean is not influenced by the initial choice of $Q_1(a)$, $\forall a$, where as when using a constant step-size α the estimate $Q_t(a)$ is function of $Q_1(a)$. Also, show that the dependence is larger for a smaller α. Propose a method such that we can have a constant step-size but no dependence of $Q_t(a)$ on $Q_1(a)$.

from equation (2.5),

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

where $\alpha = 1/n$ for sample average method

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

take $Q_1 = x$; $x \in \mathbb{R}$

Putting $n = 1$, we get:

$$Q_2 = Q_1 + \frac{1}{1} [R_1 - Q_1]$$

$$Q_2 = \cancel{x} + R_1 - \cancel{x} = R_1$$

$$Q_3 = Q_2 + \frac{1}{2}(R_2 - Q_2)$$

$$Q_3 = R_1 + \frac{1}{2}(R_2 - R_1)$$

In case of sample averages

from here, it is clear that Q_n is independent of Q_1 once an action is chosen, no matter whatever the value of Q_1 is.

Case II: α is constant s.t $\alpha \in (0,1]$

take $Q_1 = x$; $x \in \mathbb{R}$

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

$Q_2 = Q_1(1-\alpha) + \alpha R_1$

$\Rightarrow$ Q_2 is dependent on Q_1

Slly $Q_3 \ldots Q_n$ will be dependent on Q_1 with the weight $(1-\alpha)^n$ from (2.6)

Since $\alpha \in (0,1]$, smaller α implies larger dependence b/c from (2.6)

$$Q_{n+1} = (1-\alpha)^n Q_1 + \sum_{i=1}^{n} \alpha(1-\alpha)^{n-i} R_i \qquad (2.6)$$

If we take $\beta_n = \alpha / O_n$ as our time step where α is a constant time and

$$O_n = O_{n-1}(1-\alpha) + \alpha \quad ; \quad O_0 = 1$$

<u>Claim</u>: Using β_n as a time-step we can have a constant time-step such that Q_n is independent of Q_1.

$$\beta_1 = \alpha / O_1$$

$$O_1 = O_0(1-\alpha) + \alpha = \alpha$$

$$\beta_1 = \alpha/\alpha = 1$$

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

$$Q_2 = Q_1 + R_1 - Q_1 = R_1$$

$Q_3 = Q_2 + \beta_2 [R_2 - Q_2]$ which is independent of Q_1

So on $Q_4, Q_5 \ldots, Q_n$ will be independent of Q_1.