Q3. Sample averages don't perform well in nonstationary tracking due to the fact that they assign equal priority to all rewards received previously, no matter how long ago the reward was received. The changing nature of the nonstationary problem makes this a foolish thing to do.

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

is the update formula

we can take $X_n = \alpha / \alpha_n'$ where $\alpha_n' = \alpha_{n-1}' + \alpha(1 - \alpha_{n-1}')$

to prove this is free of initation bias,

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

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

$$\Rightarrow Q_{n+1} = Q_n\left(1 - \frac{\alpha}{\alpha_n'}\right) + R_n\left(\frac{\alpha}{\alpha_n'}\right)$$

$$\Rightarrow \alpha_n' Q_{n+1} = Q_n(\alpha_n' - \alpha) + R_n(\alpha)$$

$$\Rightarrow \alpha_n' Q_{n+1} = (\alpha)R_n + (\alpha_{n-1}' + \alpha - \alpha \cdot \alpha_{n-1}' - \alpha) Q_n$$

$$\Rightarrow \alpha_n' Q_{n+1} = (\alpha)R_n + \alpha_{n-1}'(1 - \alpha) Q_n$$

$$\Rightarrow \alpha_n' Q_{n+1} = (1-\alpha)^n \alpha_0' Q_1 + \sum_{i=1}^{n} \alpha(1-\alpha)^{n-i} R_i$$

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

$\Rightarrow$ free of initial bias