Margarita Prikhodko

1. a) $\alpha_n = \frac{1}{n}$

i) $Q_{n+1} = \frac{1}{n} \sum_{i=1}^{n} G_i$

$$Q_{n+1} = Q_n + \alpha_n (G_n - Q_n) =$$
$$= Q_n + \frac{1}{n}(G_n - Q_n) = \frac{1}{n} G_n + Q_n(1 - \frac{1}{n})$$

Now let $\tilde{n} = n+1$

$$Q_{n+2} = \frac{1}{n+1} G_{n+1} + Q_{n+1}(1 - \frac{1}{n+1}) =$$
$$= \frac{1}{n+1} G_{n+1} + \left(\frac{1}{n} G_n + Q_n (1 - \frac{1}{n})\right) \cdot (1 - \frac{1}{n+1}) =$$

$\cancel{\#}$ $Q_n = Q_{n-1} + \alpha_{n-1}(G_{n-1} - Q_{n-1})$

$$= \frac{1}{n+} G_{n+1} + \frac{1}{n} G_n + \ldots = \frac{1}{n} \sum_{i=1}^{n} G_i$$

ii) $\lim_{n \to \infty} V_n = 0$

$$V_n = E\left|(Q_n - r)^2\right| =$$
$$V_n = E\left(\left(\frac{1}{n}\right) \sum_{i=1 \text{ to } n} G_i - r)^2\right) =$$

$\lim_{n \to \infty} V_n = E(E(G) - r)^2)$, because expected value for $\frac{1}{n} \sum_{i=1}^{n} G$ appoaches to $= E(G)$
$n \to \infty$

$E(G) = r$

from here

$\lim_{n \to \infty} V_n = E(r-r)^2 = E(0) = 0$

b.) $\alpha_n = \alpha, \quad 0 < \alpha < 2$

i) $V_{n+1} = (1-\alpha)^2 V_n + \alpha^2 Var | G_n |$

$Var | G_n | = E | (G_n - r)^2 |$

$Q_{n+1} = Q_n + \alpha (G_n - Q_n)$

$Var (G_n) = E [(G_n - r)^2]$

$Q_{n+1} = Q_n + \alpha G_n - \alpha Q_n =$

$= (1-\alpha) Q_n + G_n$

$V_{n+1} = E ((Q_{n+1} - r)^2) =$

$= E (((1-\alpha) Q_n + \alpha G_n - r)^2) =$

$= E [(1-\alpha) Q_n - (1-\alpha) r + \alpha (G_n - r))^2]$

$V_{n+1} = E [((1-\alpha) Q_n - (1-\alpha) r)^2 + 2\alpha (1-\alpha) \cdot (Q_n - r)(G_n - 1) +$
$+ \alpha^2 (G_n - r)^2]$

$V_n = E ((Q_n - r)^2)$

$V_{n+1} = (1-\alpha)^2 V_n + 2\alpha (1-\alpha) E [(Q_n - r) \cdot (G_n - r)] +$
$+ \alpha^2 Var [G_n]$

$$E[(Q_n - r)\cdot(G_n - r)] = E\Big[\sum_{i=1}[\alpha_i G_i - r]\cdot[G_n - r]\Big] =$$

$$E[G_n] = r$$

$$E[r(G_n - r)] = 0 = rE[G_n - r] = 0$$

$$\Rightarrow V_{n+1} = (1-\alpha)^2 V_n + \cancel{\ldots}$$
$$+ \alpha^2 Var[G_n]$$

ii.) $$\lim_{n\to\infty} \left| V_{n+1} - \frac{\alpha}{2-\alpha} Var[G_n] \right| =$$

$$= \lim_{n\to\infty} \left| (1-\alpha)^2 V_n + \alpha^2 Var[G_n] - \frac{\alpha}{2-\alpha} Var[G_n] \right|$$

$$\lim_{n\to\infty} |(1-\alpha)^2 V_n| = 0 \quad (\text{from } a))$$

$$\lim_{n\to\infty} \left| \alpha^2 Var[G_n] - \frac{\alpha}{2-\alpha} Var[G_n] \right| =$$

$$= \lim_{n\to\infty} \left| \frac{2\alpha^2 + \alpha^3 - \alpha}{2-\alpha} Var[G_n] \right|$$

$$\lim_{n\to\infty} |Var[G_n]| = \lim_{n\to\infty} |E(G_n - r)^2| = E[\lim_{n\to\infty}(G_n - r)^2]$$
$$= 0$$

$$= \lim_{n\to\infty} \Big| \underbrace{Var[G_n]}_{0} \cdot \underbrace{\alpha^2 \cdot (1 - 1/(2-\alpha))}_{\text{const}} \Big| = 0$$

2. Exercise 2.7.

Yes, it is possible to avoid the bias of constant. Let's use a step size:

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

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

with $\bar{o}_0 = 0$

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

$$Q_{n+1}(a) = (1-\alpha) \cdot Q_n(a) + \alpha R_n$$

where

$Q_{n+1}(a)$ - updated estimate of a

$Q_n(a)$ - previous estimate of a

α is constant step size

R_n is observed reward

Let's introduce W_n as a weight at time step $n+1$:

$W_n = (1-\alpha)^n$

the

$$Q_{n+1} = W_n \cdot Q_1 + \sum_{i=1}^{n} \alpha \cdot W_{n-i} R_i$$

Let's take logarithm of W_n

$$\ln(W_n) = \ln((1-\alpha)^n) = n \ln(1-\alpha)$$

$W_n = e^{n \ln(1-\alpha)}$ (after exponetial)

$$Q_{n+1} = \underbrace{e^{n \ln(1-\alpha)}}_{\text{exponential with time}} Q_1 + \sum_{i=1}^{n} \alpha \cdot e^{(n-i)\ln(1-\alpha)} R_i$$