1. Expected Reward for all arms:

(i). Using Sample Mean:

$Q_1(1) = 5$
$Q_1(2) = 8$
$Q_1(3) = -6$
$Q_1(4) = 0$

Action: 2, 3, 4, 4, 1, 2, 3, 3, 1

Seq.: -5, 9, 5, 2, -4, 9, 10, 2, 1

* $k = 1$

(i) Action = 2, Reward = -5

$Q_2(2) = -5$

Rest same.

Formula:

$Q_{t+1} = Q_t + \frac{1}{k}(R_t - Q_t)$

k = no. of selections of the arm

(ii) Action = 3, Reward = 9

$Q_3(3) = 9$, rest same

(iii) A = 4, R = 5

$Q_4(4) = 5$, rest as prev.

(iv) A = 4, R = 2

$Q_5(4) = 5 + \frac{1}{2}(2-5) = 3.5$

(v) $A=1$, $R=-4$

$Q_6(1) = -4$, rest same

(vi) $A=2$, $R=9$

$Q_7(2) = -5 + \frac{1}{2}(9+5) = 2$

(vii) $A=3$, $R=10$

$Q_8(3) = \frac{19}{2} = 9.5$

(viii) $A=3$, $R=2$

$Q_9(3) = 9.5 + \frac{1}{3}(2-9.5)$

$= 9.5 + \frac{1}{3}(2-9.5) = 9.5 - 2.5$

$= 7$

(ix) $A=1$, $R=1$

$Q_{10}(1) = -4 + \frac{1}{2}(1+4) = -1.5$

B. For Exp. Wt. Avg. ($\alpha = 0.1$)

$$Q_{t+1} = Q_t + \alpha [R_t - Q_t]$$

We will follow same steps as above.

$Q_1(1) = 5$

$Q_1(2) = 8$

$Q_1(3) = -6$

$Q_1(4) = 0$

$A = 2, R = -5$

(i) $Q_2(2) = 8 + 0.1(-5-8) = 6.7$

(ii) $A = 3, R = 9$

$Q_3(3) = -6 + 0.1(9+6) = -4.5$

(iii) $Q_4(4) = 0 + 0.1(5) = 0.5 \rightarrow A = 4, R = 5$

(iv) $A=4,\ R=2$

$Q_5(4) = 0.5 + 0.1(2-0.5) = 0.65$

from updated value

(v) $A=1,\ R=-4$

$Q_6(1) = 5 + 0.1(-4-5) = 4.1$

(vi) $A=2,\ R=9$

$Q_7(2) = 6.7 + 0.1(9-6.7) = 6.93$

updated before

(vii) $A=3,\ R=10$

$Q_8(3) = -4.5 + 0.1(10+4.5) = -3.05$

Updated

(viii) $A=3,\ R=2$

$Q_9(3) = -3.05 + \frac{1}{10}(2+3.05) = -2.54$

(ix) $A=1,\ R=1$

$Q_{10}(1) = 4.1 + \frac{1}{10}(1+4.1) = 3.79$

updated Now

Sample mean is not affected by initial Q-values.
in the first case $k=1$.

By formula, $Q_{t+1} = Q_t + \frac{1}{k}(R_t - Q_t)$

$= Q_t + R_t - Q_t$

$= R_t$

$\therefore$ Ind. of Q_t.

In the case of the weighted avg:

$$Q_{t+1} = Q_t + \frac{1}{\alpha}(R_t - Q_t)$$

$$= Q_t\left(1 - \frac{1}{\alpha}\right) + \frac{R_t}{\alpha}$$

∴ There is clear dependence of Q_t.