Initialize, for $a = 1$ to k:
$Q(a) \leftarrow 0$
$N(a) \leftarrow 0$

Loop forever:
$A \leftarrow \begin{cases} \text{argmax}_a\, Q(a) & \text{with probability } 1-\varepsilon \\ \text{a random action} & \text{with probability } \varepsilon \end{cases}$
$R \leftarrow bandit(A)$
$N(A) \leftarrow N(A) + 1$
$Q(A) \leftarrow Q(A) + \frac{1}{N(A)}[R - Q(A)]$

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

$$\sum_{n=1}^{\infty} \alpha_n(a) = \infty \qquad \text{and} \qquad \sum_{n=1}^{\infty} \alpha_n^2(a) < \infty$$

$$A_t \doteq \underset{a}{\text{arg max}} \left[Q_t(a) + c\sqrt{\frac{\ln t}{N_t(a)}} \right]$$

Dynamics of the MDP

$$p(s', r \,|\, s, a) \;\doteq\; \Pr\{S_t = s', R_t = r \mid S_{t-1} = s, A_{t-1} = a\}$$

State-transition probabilities

$$p(s' \,|\, s, a) \;\doteq\; \Pr\{S_t = s' \mid S_{t-1} = s, A_{t-1} = a\} \;=\; \sum_{r \in \mathcal{R}} p(s', r \,|\, s, a).$$

Expected rewards state-action

$$r(s, a) \;\doteq\; \mathbb{E}[R_t \mid S_{t-1} = s, A_{t-1} = a] \;=\; \sum_{r \in \mathcal{R}} r \sum_{s' \in \mathcal{S}} p(s', r \,|\, s, a)$$

Expected rewards state-action-state

$$r(s, a, s') \;\doteq\; \mathbb{E}[R_t \mid S_{t-1} = s, A_{t-1} = a, S_t = s'] \;=\; \sum_{r \in \mathcal{R}} r \frac{p(s', r \,|\, s, a)}{p(s' \,|\, s, a)}.$$

State-value function for policy pi

$$\begin{aligned} v_\pi(s) \;\doteq\; \mathbb{E}_\pi[G_t \mid S_t = s] \;&=\; \mathbb{E}_\pi\left[\sum_{k=0}^{\infty} \gamma^k R_{t+k+1} \,\middle|\, S_t = s \right] \\ &= \mathbb{E}_\pi[R_{t+1} + \gamma G_{t+1} \mid S_t = s] \\ &= \sum_a \pi(a|s) \sum_{s'} \sum_r p(s', r \,|\, s, a) \Big[r + \gamma \mathbb{E}_\pi[G_{t+1} | S_{t+1} = s'] \Big] \\ &= \sum_a \pi(a|s) \sum_{s', r} p(s', r \,|\, s, a) \Big[r + \gamma v_\pi(s') \Big], \quad \text{for all } s \in \mathcal{S}, \end{aligned}$$

Action-value function for policy pi

$$q_\pi(s, a) \;\doteq\; \mathbb{E}_\pi[G_t \mid S_t = s, A_t = a] \;=\; \mathbb{E}_\pi\left[\sum_{k=0}^{\infty} \gamma^k R_{t+k+1} \,\middle|\, S_t = s, A_t = a \right]$$

Bellman optimality equation

$$\begin{aligned} v_*(s) &= \max_{a \in \mathcal{A}(s)} q_{\pi_*}(s, a) \\ &= \max_a \sum_{s', r} p(s', r \,|\, s, a) \big[r + \gamma v_*(s') \big] \\ q_*(s, a) \;&=\; \mathbb{E}\Big[R_{t+1} + \gamma \max_{a'} q_*(S_{t+1}, a') \;\Big|\; S_t = s, A_t = a \Big] \\ &=\; \sum_{s', r} p(s', r \,|\, s, a) \Big[r + \gamma \max_{a'} q_*(s', a') \Big]. \end{aligned}$$

Some equations related to Bellman

$$v_\pi(s) = \sum_a \pi(a|s) \sum_{s'} p(s'|s, a) \left[r(s, a) + v_\pi(s') \right]$$

$$v_*(s) = \max_a \left\{ \sum p(s'|s, a) \left[r(s, a) + v_*(s') \right] \right\}$$

$$q_\pi(s, a) = \sum_{s'} p(s'|s, a) \left[r(s, a) + \gamma \sum_{a'} \pi(a'|s')\, q_\pi(s', a') \right]$$

$$q_*(s, a) = \sum_{s'} p(s'|s, a) \left[r(s, a) + \gamma \max_{a'} q_*(s', a') \right]$$