# Implementation: Estimation of Action Values

In the next concept, you will write an algorithm that accepts an estimate $V$ of the state-value function $v_\pi$, along with the one-step dynamics of the MDP $p(s', r|s, a)$, and returns an estimate $Q$ the action-value function $q_\pi$.

In order to do this, you will need to use the equation discussed in the previous concept, which uses the one-step dynamics $p(s', r|s, a)$ of the Markov decision process (MDP) to obtain $q_\pi$ from $v_\pi$. Namely,

$$q_\pi(s, a) = \sum_{s' \in \mathcal{S}^+, r \in \mathcal{R}} p(s', r|s, a)(r + \gamma v_\pi(s'))$$

holds for all $s \in \mathcal{S}$ and $a \in \mathcal{A}(s)$.

You can find the associated pseudocode below.

**Estimation of Action Values**

**Input:** state-value function $V$
**Output:** action-value function $Q$
**for** $s \in \mathcal{S}$ **do**
  **for** $a \in \mathcal{A}(s)$ **do**
    $Q(s, a) \leftarrow \sum_{s' \in \mathcal{S}, r \in \mathcal{R}} p(s', r|s, a)(r + \gamma V(s'))$
  **end**
**end**
**return** $Q$

Please use the next concept to complete **Part 2: Obtain $q_\pi$ from $v_\pi$** of `Dynamic_Programming.ipynb`. Remember to save your work!

If you'd like to reference the pseudocode while working on the notebook, you are encouraged to open this sheet in a new window.

Feel free to check your solution by looking at the corresponding section in `Dynamic_Programming_Solution.ipynb`.

NEXT