```
    end
end
$Q(s,a) \leftarrow returns\_sum(s,a)/N(s,a)$ for all $s \in \mathcal{S}$, $a \in \mathcal{A}(s)$
return $Q$
```

Both the first-visit and every-visit methods are **guaranteed to converge** to the true value function, as the number of visits to each state-action pair approaches infinity. (*So, in other words, as long as the agent gets enough experience with each state-action pair, the value function estimate will be pretty close to the true value.*)

We won't use MC prediction to estimate the action-values corresponding to a deterministic policy; this is because many state-action pairs will *never* be visited (since a deterministic policy always chooses the *same* action from each state). Instead, so that convergence is guaranteed, we will only estimate action-value functions corresponding to policies where each action has a nonzero probability of being selected from each state.

Please use the next concept to complete **Part 2: MC Prediction: Action Values** of `Monte_Carlo.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 `Monte_Carlo_Solution.ipynb`.

Search or ask questions in Knowledge.

Ask peers or mentors for help in Student Hub.

NEXT