- Each occurrence of state $s \in \mathcal{S}$ in an episode is called a **visit to** s.
- There are two types of Monte Carlo (MC) prediction methods (for estimating v_π):
  - **First-visit MC** estimates $v_\pi(s)$ as the average of the returns following *only first* visits to s (that is, it ignores returns that are associated to later visits).
  - **Every-visit MC** estimates $v_\pi(s)$ as the average of the returns following *all* visits to s.

First-Visit MC Prediction (for State Values)

Input: policy π, positive integer *num_episodes*
Output: value function V ($\approx v_\pi$ if *num_episodes* is large enough)
Initialize $N(s) = 0$ for all $s \in \mathcal{S}$
Initialize $returns\_sum(s) = 0$ for all $s \in \mathcal{S}$
for $i \leftarrow 1$ ***to*** *num_episodes* **do**
  Generate an episode $S_0, A_0, R_1, \ldots, S_T$ using π
  for $t \leftarrow 0$ ***to*** $T-1$ **do**
    if S_t *is a first visit (with return* G_t*)* **then**
      $N(S_t) \leftarrow N(S_t) + 1$
      $returns\_sum(S_t) \leftarrow returns\_sum(S_t) + G_t$
  end
end
$V(s) \leftarrow returns\_sum(s)/N(s)$ for all $s \in \mathcal{S}$
return V

MC Prediction: Action Values

- Each occurrence of the state-action pair s, a ($s \in \mathcal{S}, a \in \mathcal{A}$) in an episode is called a **visit to** s, a.
- There are two types of MC prediction methods (for estimating q_π):
  - **First-visit MC** estimates $q_\pi(s, a)$ as the average of the returns following *only first* visits to s, a (that is, it ignores returns that are associated to later visits).
  - **Every-visit MC** estimates $q_\pi(s, a)$ as the average of the returns following *all* visits to s, a.

First-Visit MC Prediction (for Action Values)