GLIE MC Control

Input: positive integer $num\_episodes$
Output: policy π ($\approx \pi_*$ if $num\_episodes$ is large enough)
Initialize $Q(s,a) = 0$ for all $s \in \mathcal{S}$ and $a \in \mathcal{A}(s)$
Initialize $N(s,a) = 0$ for all $s \in \mathcal{S}, a \in \mathcal{A}(s)$
for $i \leftarrow 1$ ***to*** $num\_episodes$ **do**
    $\epsilon \leftarrow \frac{1}{i}$
    $\pi \leftarrow \epsilon\text{-greedy}(Q)$
    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, A_t) *is a first visit (with return* G_t*)* **then**
            $N(S_t, A_t) \leftarrow N(S_t, A_t) + 1$
            $Q(S_t, A_t) \leftarrow Q(S_t, A_t) + \frac{1}{N(S_t, A_t)}(G_t - Q(S_t, A_t))$
    end
end
return π

MC Control: Constant-alpha

- (In this concept, we derived the algorithm for **constant-α MC control**, which uses a constant step-size parameter α.)
- The step-size parameter α must satisfy $0 < \alpha \leq 1$. Higher values of α will result in faster learning, but values of α that are too high can prevent MC control from converging to π_*.