Implementation: MC Control: Constant-alpha

The pseudocode for (first-visit) constant-α MC control can be found below. (*Feel free to implement either the first-visit or every-visit MC method. In the game of Blackjack, both the first-visit and every-visit methods return identical results.*)

Constant-α GLIE MC Control

Input: positive integer $num\_episodes$, small positive fraction α
Output: policy π ($\approx \pi_*$ if $num\_episodes$ is large enough)
Initialize Q arbitrarily (e.g., $Q(s,a) = 0$ for all $s \in \mathcal{S}$ and $a \in \mathcal{A}(s)$)
for $i \leftarrow 1$ ***to*** $num\_episodes$ **do**
  $\epsilon \leftarrow \frac{1}{i}$
  $\pi \leftarrow \epsilon$-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**
      $Q(S_t, A_t) \leftarrow Q(S_t, A_t) + \alpha(G_t - Q(S_t, A_t))$
  end
end
return π

Please use the next concept to complete **Part 4: MC Control: Constant-alpha** 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`.