Greedy in the Limit with Infinite Exploration (GLIE)

In order to guarantee that MC control converges to the optimal policy π_*, we need to ensure that two conditions are met. We refer to these conditions as **Greedy in the Limit with Infinite Exploration (GLIE)**. In particular, if:

- every state-action pair s, a (for all $s \in \mathcal{S}$ and $a \in \mathcal{A}(s)$) is visited infinitely many times, and
- the policy converges to a policy that is greedy with respect to the action-value function estimate Q,

then MC control is guaranteed to converge to the optimal policy (in the limit as the algorithm is run for infinitely many episodes). These conditions ensure that:

- the agent continues to explore for all time steps, and
- the agent gradually exploits more (and explores less).

One way to satisfy these conditions is to modify the value of ϵ when specifying an ϵ-greedy policy. In particular, let ϵ_i correspond to the i-th time step. Then, both of these conditions are met if:

- $\epsilon_i > 0$ for all time steps i, and
- ϵ_i decays to zero in the limit as the time step i approaches infinity (that is, $\lim_{i\to\infty} \epsilon_i = 0$).

For example, to ensure convergence to the optimal policy, we could set $\epsilon_i = \frac{1}{i}$. (You are encouraged to verify that $\epsilon_i > 0$ for all i, and $\lim_{i\to\infty} \epsilon_i = 0$.)

Setting the Value of ϵ, in Practice

As you read in the above section, in order to guarantee convergence, we must let ϵ_i decay in accordance with the GLIE conditions. But sometimes "guaranteed convergence" *isn't good enough* in practice, since this really doesn't tell you how long you have to wait! It is possible that you could need trillions of episodes to recover the optimal policy, for instance, and the "guaranteed convergence" would still be accurate!