Announcements

- Mid-semester feedback is open! Due Feb 21^{st}.
  - 5 points **extra credit** on the midterm if you fill it out.
  - See Canvas announcement and assignment for details.

CS 4300/6300: Artificial Intelligence

Monte Carlo Methods

Instructor: Daniel Brown --- University of Utah

- Monte Carlo
  - District in Monaco famous for gambling

- Monte Carlo Methods
  - Simulation method
  - Used when there is uncertainty
  - Involves (usually many) randomized simulations

Example 1

- What is the probability of rolling a 7 with two dice?

Example 2

- Estimating the value of π (ratio of circle circumference to diameter)

Example 3

- Estimate the value of a state V(s) given a policy π without complete knowledge of the transition function T

Monte Carlo Value Estimation

- Use actual *experience* of interactions with the environment.
  - Environment could be the real world or a simulation.
  - Building a simulator is often easier than fully specifying T(s,a,s')
  - Works with continuous states and actions

Initialize:

$\pi \leftarrow$ policy to be evaluated

$V \leftarrow$ an arbitrary state-value function

$Returns(s) \leftarrow$ an empty list, for all $s \in \mathcal{S}$

Repeat forever:

(a) Generate an episode using π

(b) For each state s appearing in the episode:

$R \leftarrow$ return following the first occurrence of s

Append R to $Returns(s)$

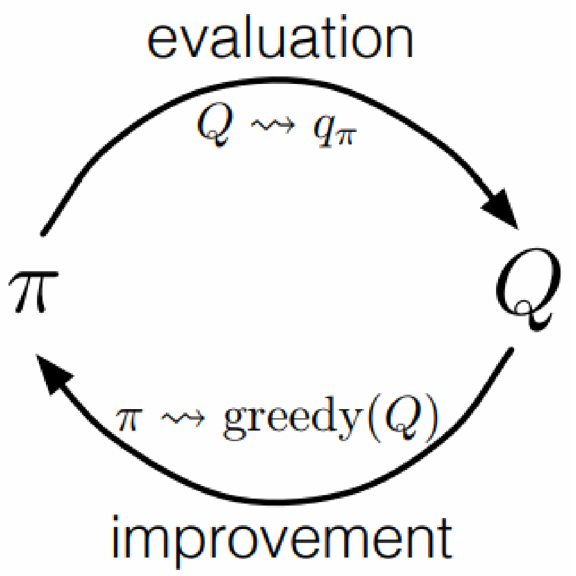
$V(s) \leftarrow$ average$(Returns(s))$

Monte Carlo Value Estimation

- How would we extend this to estimate Q(s,a)?
- What problems might we encounter with either V(s) or Q(s,a) estimation using MC methods?
  - What happens if we never visit (or rarely) visit a state or state-action pair? How could we try and fix this?
  - How do we do control?

Monte Carlo Control

- We want to do something like policy iteration



$$\pi(s) \doteq \arg\max_a q(s, a)$$

$$\pi_0 \xrightarrow{E} q_{\pi_0} \xrightarrow{I} \pi_1 \xrightarrow{E} q_{\pi_1} \xrightarrow{I} \pi_2 \xrightarrow{E} \cdots \xrightarrow{I} \pi_* \xrightarrow{E} q_*$$

Initialize, for all $s \in \mathcal{S}$, $a \in \mathcal{A}(s)$:
  $Q(s, a) \leftarrow$ arbitrary
  $\pi(s) \leftarrow$ arbitrary
  $Returns(s, a) \leftarrow$ empty list

Repeat forever:
  (a) Generate an episode using exploring starts and π
  (b) For each pair s, a appearing in the episode:
    $R \leftarrow$ return following the first occurrence of s, a
    Append R to $Returns(s, a)$
    $Q(s, a) \leftarrow \text{average}(Returns(s, a))$
  (c) For each s in the episode:
    $\pi(s) \leftarrow \arg\max_a Q(s, a)$