CSE190 Final Report

Zeyu Chen
Jiapeng Li

Problem Formulation:

- Implement reinforcement learning without the knowledge of the robot's motion CPT
  - Simulate the CPT by repeating one action
  - The ***Law of Large Number*** indicates that if we repeat the experiment many times we will be able to approximate the true distribution of the experiment

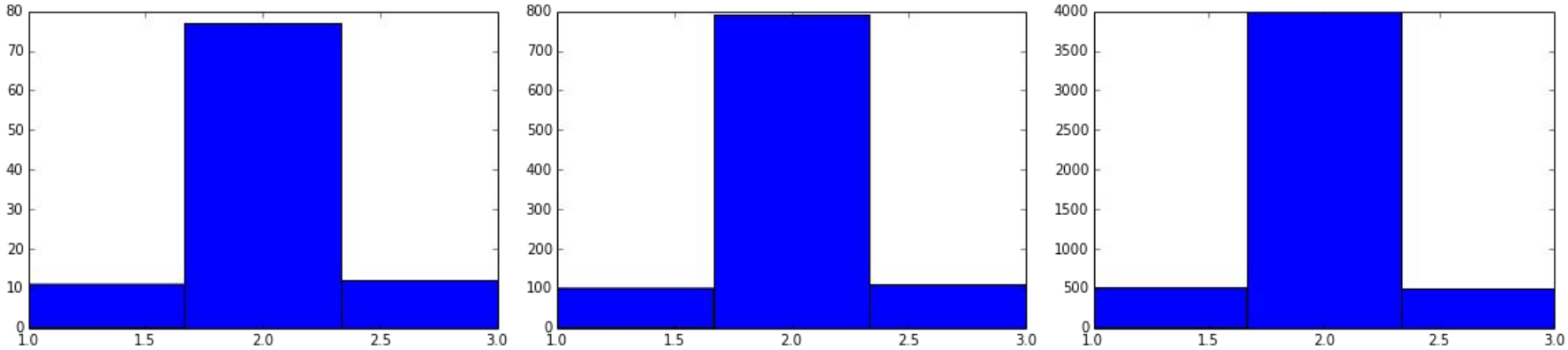


Figure: Simulation of the robot motion by sampling 500/1000/4000 times, the expected outcome has the ratio of: 0.1,0.8,0.1

Problem Formulation (Cont.)

$$
\begin{aligned}
U_1 &= R(S, Up, A) + \gamma V^{\pi}(A) \\
U_2 &= R(S, Up, S) + \gamma V^{\pi}(S) \\
U_3 &= R(S, Up, A) + \gamma V^{\pi}(A) \\
&\ldots \\
U_N &= R(S, Up, A) + \gamma V^{\pi}(A)
\end{aligned}
$$

$$
V^{\pi}(S) = \frac{1}{N} \sum_{i=1}^{N} U_i
$$

Problem Formulation (Cont.)

$$V^{\pi}(S) = \sum_{S'} P(S'|S, \pi(S))\{R(S, \pi(S), S') + \gamma V^{\pi}(S')\}$$

$$V^{\pi}_{new}(S) \leftarrow (1 - \alpha)V^{\pi}_{old}(S) + \alpha U^{\pi}_{i}$$

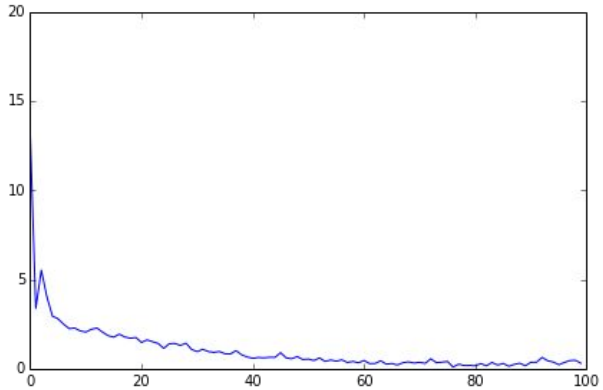
Evaluation of Algorithm Convergence

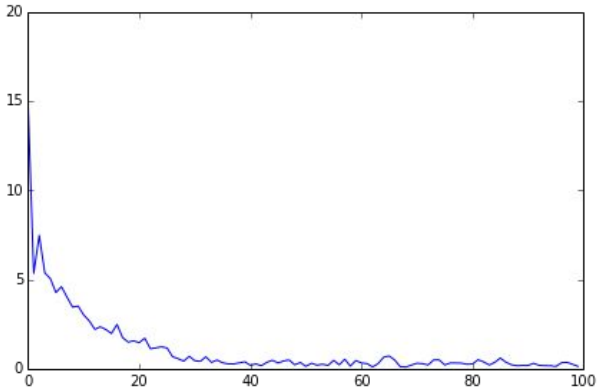
- We evaluated the convergence of our algorithm based on various situations:
  - Different learning rate
  - Different sampling numbers
  - Different Maps

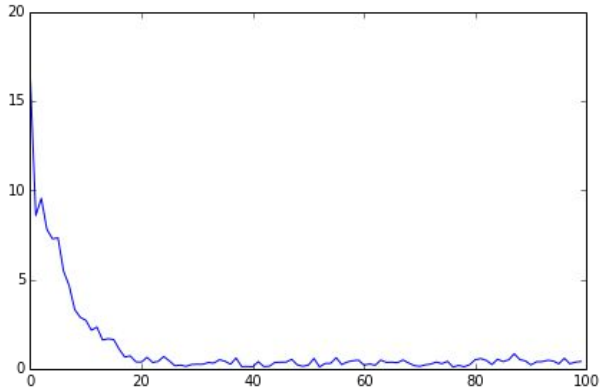
Evaluation Configuration:

$$Convergence = \sum_{i=1}^{n} |(V_{k+1}^{(\pi)}(S_i) - V_k^{(\pi)}(S_i))|$$

Different Learning Rate:

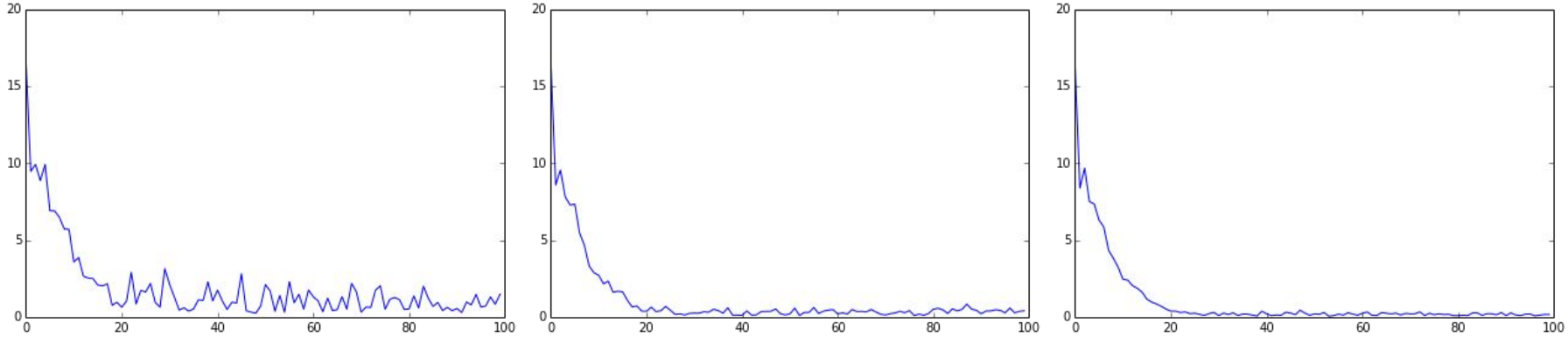






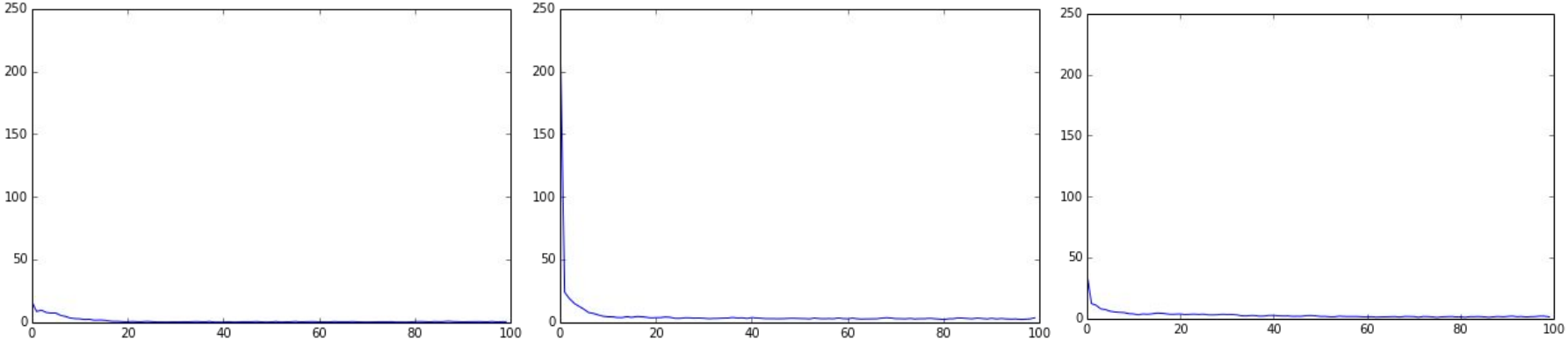
The convergence of the algorithm with a learning rate of 0.4, 0.6, 0.8 with a CPT constructed on 1000 samplings on a 3x4 map

Different Number of Samplings:



The convergence of the algorithm with a sampling number of 100, 1000, 5000 with a fixed learning rate of 0.8 on a 3x4 map

Different Map Configurations



The Convergence of the algorithm on 3x4, 10x10, 10x10 map configurations. The configurations for the 10x10 maps are different. The learning rate is fixed at 0.8 with 1000 samples calculated.