

Lecture 8: Planning versus Learning, and Certainty Equivalence

CS 639: Introduction to Reinforcement Learning

Tengyang Xie

Tuesday, September 29, 2026

Quiz 2

- 15 minutes, 11:00–11:15. Closed book; the formulas you need are printed on the sheet.
- Name, UW NetID, and Campus ID at the top.
- Final answers go in the table at the top: numbers for 1(a)–(c), one letter for 2(a)–(b).
 - Exact fractions or decimals are fine. No partial credit.
- Work alone. No notes, phones, or laptops.
- Finished early? Turn the sheet face down and wait. Sheets are collected at 11:15.

Housekeeping

- Quiz 2 is in. Solutions and scores follow on Canvas.
 - Best 5 of 6 quizzes count; no make-ups.
- HW1 is due Fri 10/02 on Gradescope; hard close 48 hours later.
 - Four slip days shared by HW1–HW4, at most two on one assignment; applied automatically.
- HW2 comes out Tue 10/06.
- Office hours
 - Tengyang Xie: Wed 1–2pm, MH 5530
 - Sarina Xi: Mon 9–11am, MH 5683
 - Yurun Yuan: Fri 9–11am, MH 5683
- Questions: Piazza first; private posts are fine.

Loss against sample size, by computer

n per pair	P(learned policy $\neq \pi^*$)	expected loss	mean value error	Hoeffding radius, $\delta = 0.1$
1	0.667	5.19	28.97	1.224
3	0.296	2.31	15.29	0.707
10	0.104	0.81	7.32	0.387
30	0.036	0.28	3.93	0.223
60	0.003	0.025	2.73	0.158
100	0.0002	0.002	2.10	0.122
300	0.0000	0.000	1.21	0.071

The machine, $\gamma = 0.9$, n samples at each of the six pairs; exact expectations over the two binomial counts. The only wrong policy is repair when worn (loss 7.79), returned when the estimated wear probability at G falls below $11/63 = 0.175$. The value error falls like $21/\sqrt{n}$ from $n = 20$ on; Hoeffding's radius certifies the margin 0.159 from $n = 60$.

Before you leave, and next time

- Reading: Sutton and Barto, Section 8.1, and the certainty-equivalence paragraph of Section 6.3.
- Lecture 9 (Thu 10/01): no model at all. Monte Carlo: average the returns.
- HW1 is due Fri 10/02 on Gradescope; hard close 48 hours later.
- HW2 comes out Tue 10/06 (planned).