Assignment 3: Policy evaluation

Published | Edit

Instructions

All the problems in this assignment should be solved and handed in **individually**. You should be prepared to answer questions about your solutions yourself. The full set of solutions should be submitted as a single PDF document in Canvas. Feel free to use any software of your choosing (or pen and paper) for preparing illustrations and drawings.

Problem 1

Consider the SCM below on the variables

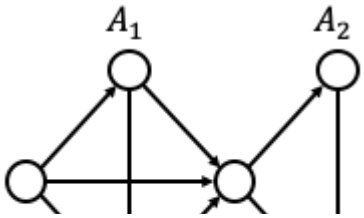
- $Age = U_A$, where $U_A \sim Uniform(\{18, \ldots, 66\})$ (integer valued)
- $Employed = U_E$, where $U_E \sim Bernoulli(0.8)$
- $Salary = Employed * [(Age - 18) * 1000 + 15000 + U_S]$, where $U_S \sim U(\{-5000, \ldots, 10000\})$ (integer valued)
- $Support = 0$
- $Income = Salary + Support$

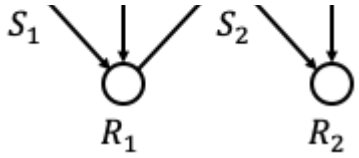
What is the *value,* in terms of $Income$, of the policy π for distributing financial $Support$, defined below?

$$\pi = \left\{ \begin{array}{ll} Support = 5000, & \text{if } Age < 25 \\ Support = 10000, & \text{if } 25 \leq Age < 35 \text{ and } unemployed \\ Support = 2000, & \text{if } 25 \leq Age < 35 \text{ and } employed \\ Support = 0, & \text{if } Age \geq 35 \end{array} \right\}$$

Problem 2

Consider the causal graph representing a Markov decision process (MDP) below.





Now, assume that you could access samples from the distribution $p_\mu$defined by the policy μ with

$$p_\mu\,(S_1,\ A_1,\ R_1,\ S_2,\ A_2,\ R_2) = p\,(S_1)\,p_\mu\,(A_1 \mid S_1)\,p\,(R_1 \mid S_1,\ A_1)\,p\,(S_2 \mid S_1,\ A_1,\ R_1)\,p_\mu$$

Consider evaluating a new policy π with action probabilities $p_\pi\,(A_t \mid S_t)$ under *the same transition and reward probabilities* as above (i.e., same conditional distributions for states and rewards).

Recall that $V\,(\pi)$ is defined as the expected sum of rewards under $p_\pi\,(S_1,\ A_1,\ R_1,\ S_2,\ A_2,\ R_2)$, that is $\mathbb{E}_\pi[R_1 + R_2]$.

A) Identify (derive) a statistical estimand of the value $V\,(\pi)$ that uses importance weighting (or inverse-propensity weighting), derived as expectation over the distribution p_μ

B) Propose a finite-sample estimator of your estimand which makes use of samples from p_μ.

Problem 3

In the sessions on off-policy evaluation, we argued that a difficulty with off-policy evaluation of sequential decision-making policies was to find enough samples that follow the proposed policy in data. We expand on this argument in the technical report **Evaluating Reinforcement Learning Algorithms in Observational Health Settings** . Read chapters 1–5 (at least) of this paper and briefly summarize the main findings of chapter 5 (~1/2 page)

Points	20
Submitting	a file upload
File types	pdf

Due	For	Available from	Until
15 Oct	Everyone	-	-

+ Rubric