Northeastern University
College of Engineering

EECE 5698 - ST: Reinforcement Learning Spring 2023

HW1

Problem 1.

Consider a random variable X whose pdf is:

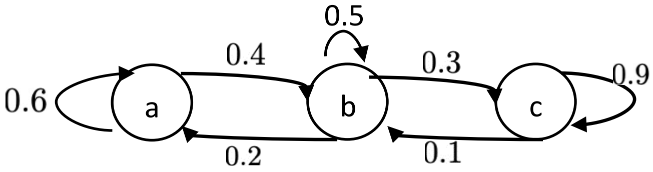
$$P_X(x) = \begin{cases} 1/2 & x = -1 \\ 1/4 & x = 0 \\ 1/4 & x = 1 \end{cases}$$

a) Find E[X] and E[X^2]

b) Find Var[X] and σ

Problem 2.

Consider a Markov chain $\{x_n, n = 0, 1, ...\}$ with a transition diagram:



a) Compute the transition matrix, given x={a,b,c}

b) Compute $p(x_k = b|x_{k-1} = a)$ and $p(x_k = b|x_{k-2} = a)$

Problem 3.

Consider two-bandit problem with the following reward distributions:

$$R(a^1) \sim Uniform[0 \;\; 1.4]$$

$$R(a^2) \sim \mathcal{N}(\mu = 0.5, \sigma = 1)$$

a) Compute the optimal $Q^*(a^1)$, $Q^*(a^2)$ and π^*.

b) Consider the reward distributions are unknown. Use the learning rate $\alpha = 0.5$ to estimate $Q(a^1)$, $Q(a^2)$ and π given the following:

	k=1	k=2	k=3	k=4	k=5
Action	a^1	a^2	a^1	a^2	a^1
Reward	1	0.5	0	1.25	1.35

c) Repeat part b for optimistic initial value Given $Q(a^1) = Q(a^2) = 5$.

Problem 4.

Given the following interaction and reward sequence, set $\alpha = 0.5$, $H_1(a^1) = H_1(a^2) = 0$ and use the gradient-bandit policy to compute $H_4(a^1)$, $H_4(a^2)$, $\pi_4(a^1)$ and $\pi_4(a^2)$.

	k=1	k=2	k=3
Action	a^1	a^2	a^1
Reward	1	0.5	0

Questions about the HW should be directed to TA, Begum Taskazan, at taskazan.b@northeastern.edu.