Problem 1

Consider a random variable X whose probability mass function (pmf) is given by:

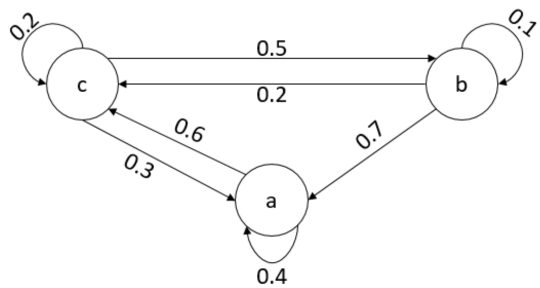
$$P_X(x) = \begin{cases} \frac{3}{8}, & x = -2, \\ \frac{1}{8}, & x = -1, \\ \frac{1}{4}, & x = 0, \\ \frac{1}{4}, & x = 1, \end{cases}$$

(a) Find $\mathbb{E}[X]$ and $\mathbb{E}[X^2]$.

(b) Find $\mathrm{Var}[Y]$ where $Y = 3X + 1$.

Problem 2

Consider a Markov chain $\{x_n, n = 0, 1, \ldots\}$ with state space $\{a, b, c\}$ and the following transition diagram:



a)Compute the transition matrix for this Markov chain, using the state order (a, b, c).

b)Compute $p(x_k = c \mid x_{k-1} = a)$ and $p(x_k = c \mid x_{k-2} = a)$. Briefly explain why these two probabilities can differ.

Problem 3

Consider a two-bandit problem with the following reward distributions:

$$R(a^1) \sim \mathcal{N}\big(\mu = 0.6,\ \sigma = 1.2\big) \qquad R(a^2) \sim \text{Uniform}[-0.2,\ 1.2]$$

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

b)Suppose the reward distributions are unknown. Use the learning rate $\alpha = 0.6$ to estimate $Q(a^1)$, $Q(a^2)$, and the policy π, given the following data:

k	1	2	3	4	5
Action	a^1	a^2	a^1	a^1	a^2
Reward	0.2	0.9	0	1.3	0.1

c)Repeat part (b) for **optimistic initial values** given $Q(a^1) = Q(a^2) = 5$.

d)After part (c), you now have estimates for $Q(a^1)$ and $Q(a^2)$ at $k = 5$. Suppose at $k = 6$, you want to pick your *next action* according to ϵ-greedy policy with $\epsilon = 0.3$. What would be the probability of tacking each action?

Problem 4

Given the following data, set $\alpha = 0.6$, and initialize $H_1(a^1) = H_1(a^2) = 0$.

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

Use the **gradient-bandit** policy to compute $H_4(a^1)$, $H_4(a^2)$, $\pi_4(a^1)$, and $\pi_4(a^2)$.