Homework Set 1

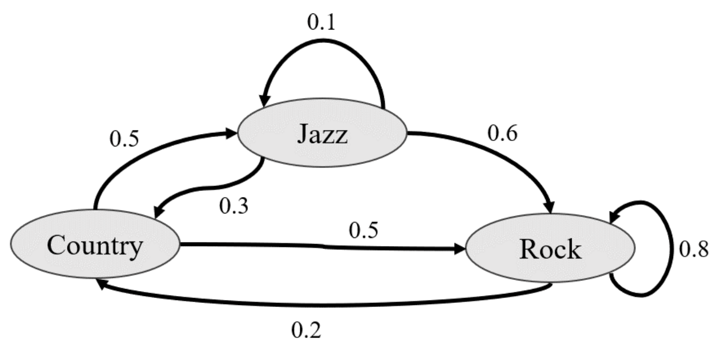
Problem 1: For a vector $\mathbf{x} \in \mathbb{R}^n$, recall the definitions of $\|\mathbf{x}\|_\infty$, $\|\mathbf{x}\|_1$, and $\|\mathbf{x}\|_2$.

1. Prove that $\|\mathbf{x}\|_\infty$ satisfies the properties of a norm over a vector space. In particular, show that:
   - $\|\mathbf{x}\|_\infty \geq 0$ for all $\mathbf{x} \in \mathbb{R}^n$, and $\|\mathbf{x}\|_\infty = 0$ if an only if $\mathbf{x} = \mathbf{0}$.
   - $\|\alpha\mathbf{x}\|_\infty = |\alpha| \|\mathbf{x}\|_\infty$ for all $\mathbf{x} \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$.
   - $\|\mathbf{x}_1 + \mathbf{x}_2\|_\infty \leq \|\mathbf{x}_1\|_\infty + \|\mathbf{x}_2\|_\infty$ for all $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^n$.
2. Prove that
$$\|\mathbf{x}\|_\infty \leq \|\mathbf{x}\|_2 \leq \|\mathbf{x}\|_1 .$$
3. Prove the following special case of the Holder's inequality for two vectors $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^n$:
$$| < \mathbf{x}_1, \mathbf{x}_2 > | \leq \|\mathbf{x}_1\|_\infty \|\mathbf{x}_2\|_1 .$$

Problem 2: Recall the law of total expectation (tower rule) for random variables X (with existing expected value) and Y, i.e.,
$$\mathbb{E}_X[X] = \mathbb{E}_Y[\mathbb{E}_X[X|Y]],$$
where the subscripts denote the randomness over which the expectations are computed.

1. Prove the tower rule for two discrete random variables X and Y with probability distributions $\mathbb{P}[X = x]$ defined for all $x \in \mathcal{X}$, $\mathbb{P}[Y = y]$ defined for all $y \in \mathcal{Y}$, and $\mathbb{P}[X = x, Y = y]$ defined for all $x \in \mathcal{X}, y \in \mathcal{Y}$.
2. Prove the tower rule for two continuous, real-valued random variables X and Y with individual and joint probability density functions $f_X(x)$, $f_Y(y)$, and $f_{X,Y}(x, y)$.
3. The tower rule also extends to conditional expectations, resulting in
$$\mathbb{E}_X[X|Z] = \mathbb{E}_Y[\mathbb{E}_X[X|Y, Z] |Z] .$$
   Apply this result to write $\mathbb{E}_R[R_t|S_t = s]$, i.e., the expected (instantaneous) reward at a given state s, based on the properties of an MDP and a policy over that. In particular, consider an MDP with state space $\mathcal{S}$, action space $\mathcal{A}$, transition function P, and stochastic reward function R with mean $\bar{R}(s, a) = \mathbb{E}_R[R_t|S_t = s, A_t = a]$, and a stochastic, stationary policy $\pi(a|s)$.

Problem 3: Consider a jukebox that plays songs from three genres of music: "Jazz," "Rock," and "country." Once started, it plays a Jazz song and then, switches between the genres according to the time-homogeneous Markov chain $(X_t)_{t=0}^{\infty}$, depicted below.



1. Specify the state space $\mathcal{S}$, the initial distribution μ_0, and the transition function (matrix) P of this Markov chain.

2. What is the probability that the following sequence of genres is played once the jukebox starts?

   (a) (*Jazz*, *Country*, *Country*, *Rock*)

   (b) (*Jazz*, *Rock*, *Rock*, *Country*, *Jazz*)

   (c) (*Jazz*, *Rock*, *Country*, *Rock*, *Country*, $\ldots$, *Rock*, *Country*)
   where $X_0 = \textit{Jazz}$, $X_{2k-1} = \textit{Rock}$, and $X_{2k} = \textit{Country}$ for all $k \in \{1, 2, 3, \ldots m\}$.

   (d) (*Jazz*, *Rock*, *Country*, *Rock*, *Country*, $\ldots$)
   where $X_0 = \textit{Jazz}$, $X_{2k-1} = \textit{Rock}$, and $X_{2k} = \textit{Country}$ for all $k \in \mathbb{N} = \{1, 2, 3, \ldots\}$, i.e., as $m \to \infty$ in the previous part.

3. What is the probability that the third song (at $t = 2$) belongs to any of these three genres?

4. This Markov chain is ergodic and hence, has a unique stationary (steady-state) distribution $\bar{\mu}$. Compute $\bar{\mu}$ by hand, calculator, or code; show the steps you use in the computation.

5. Consider the consecutive update of the state distribution μ_t from $t = 0$ to $t = 100$. Plot the sequence of differences between the state distribution μ_t and the stationary distribution $\bar{\mu}$, measured using L1 norm, with respect to the time steps.

6. If we start the jukebox and let it play for a very long time, which genre do you expect to be played most often? Explain your reasoning.

[**Note:** For this problem, if you have any code in a programming language of your choice, please attach it to the end of your submission.]

Problem 4: Consider a discrete-time, time-homogeneous Markov chain $(X_t)_{t=0}^{\infty}$ that satisfies the Markov property, i.e.,

$$\mathbb{P}[X_{t+1}|X_0, X_1, \dots, X_t] = \mathbb{P}[X_{t+1}|X_t] , \qquad \forall t \in \mathbb{N}_0.$$

1. Prove the following holds:

$$\mathbb{P}[X_{t+1}, X_{t+2}, \dots, X_{t+m}|X_0, X_1, \dots, X_t] = \mathbb{P}[X_{t+1}, X_{t+2}, \dots, X_{t+m}|X_t] , \\ \forall t \in \mathbb{N}_0 , \forall m \in \mathbb{N}.$$

2. Prove the following holds:

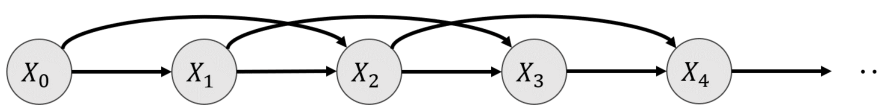
$$\mathbb{P}[X_{t+k}, X_{t+k+1}, \dots, X_{t+m}|X_0, X_1, \dots, X_t] = \mathbb{P}[X_{t+k}, X_{t+k+1}, \dots, X_{t+m}|X_t] , \\ \forall t \in \mathbb{N}_0 , \forall k, m \in \mathbb{N} , m \geq k.$$

Problem 5: A Markov chain of order k defines a stochastic process $(X_t)_{t=0}^{\infty}$ such that

$$\mathbb{P}(X_t|X_0, X_1, X_2, \dots, X_{t-1}) = \mathbb{P}(X_t|X_{t-k}, X_{t-k+1}, \dots, X_{t-1}) \quad \text{for all } t \in \mathbb{N}_0, t \geq k,$$

i.e., the probability of moving to the next state depends on the past k states. Higher-order Markov chains can be transformed into an equivalent simple Markov chain (Markov chain of order 1).

1. Consider a Markov model of order 2 as shown in the graphical model below. Let the state



space be $\{1, 2\}$ and the transition probabilities be

$$\mathbb{P}(X_t = 1|X_{t-2} = 1, X_{t-1} = 1) = 0.8 , \quad \mathbb{P}(X_t = 1|X_{t-2} = 1, X_{t-1} = 2) = 0.1,$$
$$\mathbb{P}(X_t = 1|X_{t-2} = 2, X_{t-1} = 1) = 0.3 , \quad \mathbb{P}(X_t = 1|X_{t-2} = 2, X_{t-1} = 2) = 0.7,$$

if $t \geq 2$, and

$$\mathbb{P}(X_1 = 1|X_0 = 1) = 0.2 , \quad \mathbb{P}(X_1 = 1|X_0 = 2) = 0.4,$$

if $t = 1$. Let the initial state distribution be $\mathbb{P}(X_0 = 1) = 0.5$. Create an equivalent simple Markov chain over the state space $\{1, 2\}^2 = \{(1, 1), (1, 2), (2, 1), (2, 2)\}$. Write all transition probabilities between the states and the initial state distribution, or alternatively, show them on a graphical representation.

2. Based on the example in the previous part, describe the process for creating an equivalent simple Markov chain for a Markov chain of order k.

Problem 6: Consider a movie recommendation system, interacting with a human user, that aims to suggest movies that the user would like to watch. In each step of the interaction, the human communicates its current desired movie genre to the system, which may be *Action*, *Horror*, or *Comedy*. Upon receiving this communication, the system selects a movie among *Movie A*, *Movie B*, *Movie C*, and *Movie D*. The genre of these movies is listed below, where 1 shows that the movie is categorized under that genre while 0 shows that it is not categorized under that genre.

	Action	Horror	Comedy
Movie A	1	0	1
Movie B	1	1	0
Movie C	0	1	1
Movie D	0	1	0

Upon receiving the recommendation, the user watches the movie and provides a *like* if the movie belongs to the desired genre and provides a *dislike* otherwise. If the movie belongs to the desired genre, in the next interaction, the user's desired movie genre will be selected uniformly at random from the other two genres. For instance, if the user wanted Horror in this interaction and Movie B, C, or D was suggested, the user will want Action or Comedy in the next interaction with equal probability. However, if the movie does not belong to the desired genre, in the next interaction, the user's desired movie genre will remain the same. At the beginning of the interaction, the user may select any of the genres uniformly at random.

1. Define an MDP that models this sequential decision-making scenario; in particular, specify all elements of the MDP. Consider an infinite-horizon setting with a discount factor of $\gamma = 0.95$. Let the reward value be bounded between $[-1, +1]$.

2. Consider the following recommendation strategy by the system:

   - If the user asks for Action movies $\longrightarrow$ The system will recommend Movies A or B, each with probability of 0.4, or Movie D with probability of 0.2.
   - If the user asks for Horror movies $\longrightarrow$ The system will recommend any of the four movies with probability 0.25.
   - If the user asks for Comedy movies $\longrightarrow$ The system will recommend Movie B with probability of 0.1, or Movie C with probability of 0.9.

   Determine whether this policy is stationary or not. Also, determine whether this policy is deterministic or randomized (stochastic).

3. Define the Markov chain that is induced by the policy given in Part 2 over the MDP you introduced in Part 1; in particular, specify all elements of the Markov chain.

4. Now, suppose the user's next desired movie genre was dependent not only on its last desired genre (and the recommended movie) but also the one before that. Show a graphical representation of the temporal evolution of the model in this case.

[**Note:** For this problem, if you have any code in a programming language of your choice, please attach it to the end of your submission.]