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 $\|\mathbf{x}\|_\infty = 0$ <u>if and only if</u> $\mathbf{x} = \mathbf{0}$.
   - $\|\alpha\mathbf{x}\|_\infty = |\alpha| \, \|\mathbf{x}\|_\infty$ 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 that
$$\|\mathbf{x}\|_1 \leq \sqrt{n} \, \|\mathbf{x}\|_2 .$$
   This inequality, combined with Part 2 yields an equivalency between L1 norm and L2 norm, in particular, $\|\mathbf{x}\|_2 \leq \|\mathbf{x}\|_1 \leq \sqrt{n} \, \|\mathbf{x}\|_2$ and $\frac{1}{\sqrt{n}} \|\mathbf{x}\|_1 \leq \|\mathbf{x}\|_2 \leq \|\mathbf{x}\|_1$.
4. Prove the following special case of the Hölder'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: The law of total expectation (tower rule) for random variables X (with existing expected value) and Y states that

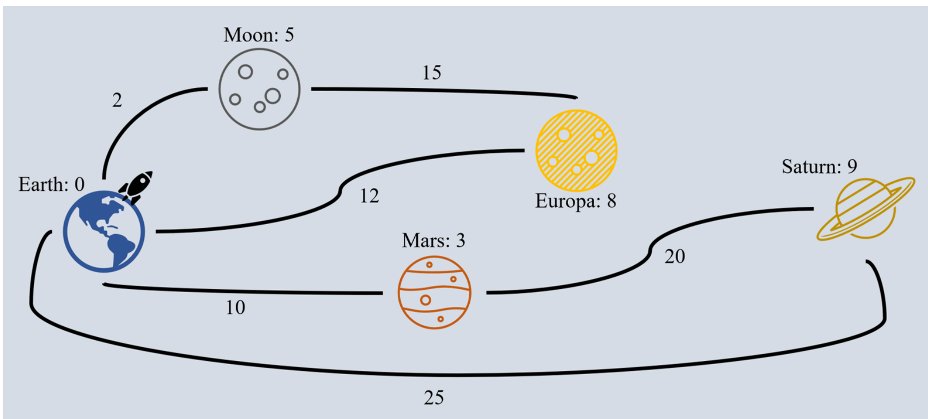
$$\mathbb{E}_X[X] = \mathbb{E}_Y[\mathbb{E}_X[X|Y]],$$

where the subscripts denote the randomness over which the expectations are computed, i.e., $\mathbb{E}_X$ denotes expectation (average) with respect to randomness of X and $\mathbb{E}_Y$ denotes expectation (average) with respect to randomness of Y. Notice that $\mathbb{E}_X[X|Y]$ is a function of Y that maps any realization (outcome) y of Y to $\mathbb{E}_X[X|Y=y]$.

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
$$\mathbb{E}_X[X|Z] = \mathbb{E}_Y[\mathbb{E}_X[X|Y,Z]\,|Z] .$$
   Prove the above tower rule for three discrete random variables X, Y, and Z with probability distributions $\mathbb{P}[X=x]$ defined for all $x \in \mathcal{X}$, $\mathbb{P}[Y=y]$ defined for all $y \in \mathcal{Y}$, $\mathbb{P}[Z=z]$ defined for all $z \in \mathcal{Z}$, and all joint probabilities and conditional probabilities defined for all realizations of X, Y, and Z.

4. The tower rule can facilitate computing expectations. To see this, consider a robot that is doing a one-dimensional random walk, starting from 0 and taking steps X_t at times $t \in \{1, 2, 3, \ldots\}$. Each step $X_t = +1$ with probability p and $X_t = -1$ with probability $1 - p$, where $p \in (0, 1)$. The robot's energy will be depleted after H steps, where $H \in \{1, 2, 3, \ldots\}$ is a geometrically distributed random variable with parameter $q \in (0, 1)$ and is independent of X_t variables. When the robot runs out of energy, its location will be $L = \sum_{t=1}^{H} X_t$. Compute the expected (average) robot's location at the end, that is $\mathbb{E}_L[L]$, using the tower rule.

Problem 3: An earthling is on a space exploration mission consisting of random voyages in the solar system. The map below shows the possible bidirectional travel routes.



1. Each voyage consists of a sequence of stops that follow these guidelines:
   - It starts on Earth and ends as soon as Earth is reached again.
   - It randomly goes from one stop to the next, in particular:
     - If currently on Earth, its next stop will be Moon, Mars, Europa, or Saturn with probabilities 0.1, 0.4, 0.25, and 0.25, respectively.
     - If currently on Moon, its next stop will be Earth or Europa with probabilities 0.6 and 0.4, respectively.
     - If currently on Mars, its next stop will be Earth or Saturn with probabilities 0.3 and 0.7, respectively.
     - If currently on Europa, its next stop will be Earth.
     - If currently on Saturn, its next stop will be Earth.

   For example, *(Earth, Moon, Europa, Earth)* is a possible voyage while *(Earth, Moon, Europa)*, *(Earth, Mars, Europa, Earth)*, and *(Earth, Moon, Earth, Moon, Europa, Earth)* are not.

   Let V represent the voyage, i.e., sequence of stops, as a random variable. We would like to measure the probability of each voyage. Given that, specify a probability space $(\Omega_V, \mathcal{F}_V, \mathbb{P}_V)$ for V. For the probability function $\mathbb{P}_V$, only list a subset of its values that are sufficient to infer the rest of its values.

2. Each trip in a voyage incurs an energy cost, indicated by the values over the routes on the map. At each stop, the earthling collects samples with values indicated in front of the names on the map. Each voyage is associated with a return determined by

$$\textit{return} = \textit{sum of sample values} - \textit{sum of energy costs}.$$

   For example, the return of voyage *(Earth, Moon, Europa, Earth)* is $(0+5+8+0)-(2+15+12)$.

   Let $R = R(V)$ represent the return of voyage V (as described in Part 1) as a random variable. Specify a probability space $(\Omega_R, \mathcal{F}_R, \mathbb{P}_R)$ for R.

3. Compute the true (theoretical) expectation (average) of the return of this earthling's voyage, i.e., evaluate $\mathbb{E}[R]$.

4. Instead of computing the theoretical expectation (average) of return R, our voyager wants to estimate the expectation empirically from n voyages. To find out how many voyages are needed to make a confident estimate, follow the steps below.

   (a) Let R_1, R_2, $\ldots$, R_n denote the returns of the n voyages. Find an upper-bound and lower-bound on R_i, i.e., find α and β such that

$$\alpha \leq R_i \leq \beta \;, \quad \forall i \in \{1, 2, \ldots, n\}.$$

   (b) Let $S_n = \sum_{i=1}^{n} R_i$ denote the sum of n returns. Derive and express $\mathbb{E}[S_n]$ in terms of $\mathbb{E}[R]$, where $\mathbb{E}[R]$ is the theoretical expectation (average) of return R. (There is no need to replace $\mathbb{E}[R]$ with its numerical value.)

   (c) Apply the Hoeffding's inequality to upper-bound the probability that S_n deviates from its expected value obtained in Part b, i.e., bound $\mathbb{P}\left(|S_n - \mathbb{E}[S_n]| \geq \epsilon\right)$ for any $\epsilon > 0$. Notice that the samples R_i are independent random variables.

   (d) Considering $\bar{S}_n = \frac{1}{n} S_n$ to denote the average of n returns, derive and express $\mathbb{E}[\bar{S}_n]$ in terms of $\mathbb{E}[R]$, where $\mathbb{E}[R]$ is the theoretical expectation (average) of return R. (There is no need to replace $\mathbb{E}[R]$ with its numerical value.)

   (e) <u>Use the result in Part c to derive</u> an upper-bound on the probability that $\bar{S}_n$ deviates from $\mathbb{E}[\bar{S}_n]$, i.e., bound $\mathbb{P}\left(|\bar{S}_n - \mathbb{E}[\bar{S}_n]| \geq \epsilon'\right)$ for any $\epsilon' > 0$. (Notice that you should show how the Hoeffding's inequality for the summation of random variables leads to its version for the average of random variables.)

   (f) The voyager wants to use $\bar{S}_n$ as an estimate of $\mathbb{E}[R]$. Based on the result in Part 5, determine the number n of voyages required to guarantee with probability 0.9 that $\bar{S}_n$ will be within the range $[\mathbb{E}[R] - 2, \mathbb{E}[R] + 2]$.

Problem 4: Consider a discrete-time, time-homogeneous Markov process $(X_t)_{t=0}^{T}$. Recall that the Markov property states that, given the present, (1 step in) the future is independent of the past; formally,

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

1. Show that, given the present, the m steps in the future are independent of the past. Formally, prove that

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

   [Hint: You may find it helpful to start by decomposing the joint probability distribution (chain rule of probability).]

2. Show that, given the present, the d-th step in the future is independent of the past. Formally, prove that

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

   [Hint: You may find it helpful to introduce a missing random variable and marginalize over it; for instance, by introducing W and marginalizing over it, one obtains $\mathbb{P}[Y|Z] = \sum_{w \in \mathcal{W}} \mathbb{P}[Y, W = w|Z]$.]

3. Show that, given the present, the k-th step to the m-th step in the future are independent of the past. Formally, prove that

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

   [Hint: The techniques used in Part 2 could be helpful here as well.]