1. This is another Matlab program.

   (a) The purpose of this problem is to write Matlab code so that given a PMF, you calculate the expectation by doing the following Matlab function:
   function E = expectedValue(pmf,indices). For a discrete random variable, you provide the support which is the set of integers k for which $P_X(k) > 0$. You also specify pmf which should be a vector of the same length as $IndicesSupport$. For example, if X takes on values 2 with value 0.6 and 3 with value 0.4 then $IndicesSupport = [2; 3]$ and $pmf = [0.6; 0.4]$. Implement this within **expectedValue.m**.

   (b) Create a **variance.m** file that calculates the variance of a random varaible, given its PMF and its support. Perform this by exploiting LOTUS:

   $$\text{var}(X) = \mathbb{E}[(X-\mu)^2] = \sum_{k \in support} (k-\mu)^2 P_X(k)$$

   In a **variance.m** file, implement **function V= variance(pmf,indices)**. Within your M-file, you should first find the expected value of X and then proceed treating with the above sum.

   First do a simple example for $[2; 3]$ and $[0.60.4]$ by hand. Verify your code is consistent with that. Hopefully this helps you understand what LOTUS means.

   (c) Revisit problem 1 from homework 2 and define X to be the random variable pertaining to the outcome of the first role and Y to be the random variable pertaining to the outcome of the second role. Build a matrix **jointPMFXY** where the (i, k) entry corresponds to $P_{X,Y}(i, k)$ *hint: recycle alot of the Matlab code in homework 2.*

   (d) Familiarize yourself with Matlab notations $A(i, :)$ as the ith row vector associated with matrix A and $A(:, k)$ as the kth column bector associated with matrix A. With that, write code to extract the marginal PMFs P_X and P_Y from the joint pmf $P_{X,Y}$. Specifically, define **function [PX,PY]=marginalizeJointPMF(jointPMFXY)** and write code to extract the two marginals. *hint, use the total probability theorem code from the Matlab problem in homework 2.*

   (e) Calculate $P_{X|Y=k}$, the PMF on X given $Y = k$. Specifically, define **function PXgivenY = conditionalPMF(PXY,k)** where the output **PXgivenY** is a vector whose ith entry is $P_{X|Y}(i|k)$. Then write an analogous function **function PYgivenX = conditionalPMF(PXY,i)** whose output **PYgivenX** is a vector whose kth entry is $P_{Y|X}(k|i)$.

   (f) use your above code to calculate $\text{var}(X|Y = 1)$ and $\mathbb{E}[X|Y = 2]$.

2. This problem involves an interpretation of the expectation as the 'smallest error guess' of a random variable.

(a) In linear algebra, for two vectors x and y, remember that the inner product and norm squared of vectors are defined as

$$\langle z, w\rangle \triangleq z^T w, \qquad \|z\|^2 \triangleq \langle z, z\rangle$$

Remember that z and w are *orthogonal* if $\langle z, w\rangle = 0$. Now let's assume that everything is in one dimension (no need to worry about transpose) and we define the inner product between **random variables** Z and W as:

$$\langle Z, W\rangle \triangleq \mathbb{E}[ZW], \qquad \|Z\|^2 \triangleq \langle Z, Z\rangle$$

where orthogonality still means zero inner product. For any constant c, define the error as $E = X - c$. Suppose we would like to identify a constant c^* that minimizes $\|E\|^2$ over all possible constants c. Use the geometry intuition for the projection theorem in linear algebra (least squares) to conclude that for the the optimal c^*, it must be that $E^* = X - c^*$ is orthogonal to all constants d:

$$\langle E^*, d\rangle = 0 \quad \text{for any d.}$$

From this, conclude that $c^* = \mathbb{E}[X]$.

(b) Now suppose we have observed that $Y = y$. Re-define the inner product as $\langle Z, W\rangle \triangleq \mathbb{E}[ZW|Y = y]$. For any function $g(y)$ define the error as $E = X - g(y)$. Now we would like to build a function $g^*(y)$ that minmizes $\|E\|^2$ over all functions $g(y)$. Use almost the exact same intuition as above to conclude that the optimal $g^*(y)$ is given by $g^*(y) = \mathbb{E}[X|Y = y]$.

3. (a) show that if the PMF of a random variable X is symmetric around 0 (in other words, $P_X(k) = P_X(-k)$ for any k), then $\mathbb{E}[X] = 0$.

   (b) use LOTUS and the previous part of this problem to show that if the PMF of X is symmetric around some some number a, then $\mathbb{E}[X] = a$. (*hint: think about constructing a new random variable* $Y = g(X)$ *and select* g *cleverly. If you do this, you can show this easily*).

4. In Brazil, officials are trying to prepare for the World Cup and Olympics. In Rio on a given day, an almanac shows that $\frac{1}{3}$ of the time there is no smog, $\frac{1}{6}$ of the time there is mild smog, and $\frac{1}{2}$ of the time it is very smoggy. Our outside media reports can query individuals and determine if few or many people will go to the beach. When there is no smog, typically $\frac{5}{8}$ of people go to the beach. When there is mild smog, typically $\frac{1}{4}$ of people go to the beach. When it is very smoggy, typically $\frac{1}{8}$ of people go to the beach.

   (a) Draw a sequential description of the sample space

(b) Suppose that a Carbon tax is imposed on businesses, depending on the smog level of that day. The tax is 50, 25, or 10 dollars per day depending upon if there is lots of smog, a moderate level, or none. If few people come to the beach, then parking at the beach is free; otherwise parking costs 2 dollars. Construct random variables X and Y pertaining to the Carbon tax and parking costs. (*specifically, I mean draw a table where for every* $\omega \in \Omega$), you provide $X(\omega)$ and $Y(\omega)$.

(c) Calculate the PMFs $P_X(k)$, $P_Y(k)$, and the joint PMF $P_{X,Y}(k,j)$.

(d) Now suppose that a reporter in New York has observed from a video that people showed up to the beach on without their wallets. Calculate what the reporter's belief would be on the taxes imposed on that day. Represent this as a conditional PMF, do the calculation, and draw it as a PMF.

(e) Calculate the average amount of tax revenue the reporter expects the Brazilian government to incur that day.

(f) Calculate the variance in tax revenue the reporter expects the Brazilian government to incur that day.

5. Suppose that a protein either does not cross a cell membrane, it enters a cell, or it exits a cell. It does not cross a cell membrane with probability $1-p$ and it enters or exits a cell with equal likelihood. Denote X to be the change in protein count for the cell, and denote Y to be the change in protein count for the environment outside of the cell.

   (a) Provide the map in terms of a table for the random variables X and Y.

   (b) Calculate the PMF for X, $P_X(k)$ and the PMF for Y, $P_Y(k)$. Also calculate $\mathbb{E}[X]$ and $\mathbb{E}[Y]$.

   (c) Calculate the covariance between X and Y. Are they uncorrelated?

   (d) Calculate the conditional PMF for Y given X, $P_{Y|X}(k|j)$.

   (e) Calculate the joint PMF for X and Y, $P_{X,Y}(j,k)$. Are X and Y statistically independent?

6. Suppose that X_1 is a binary random variable that is equally likely to be 0 or 1. Analogously, X_2 is a random variable with the same statistical law who is statistically independent of X_1. Define the random variable $Y = g(X_1, X_2)$ as

$$Y = X_1 \oplus X_2$$

   where $\oplus$ means modulo-2 sum. Note here that clearly Y is 'physically' dependent upon X_1 and X_2 because it is a function of them.

   (a) use LOTUS to calculate $E[Y]$.

   (b) calculate $P_Y(k)$ (*hint: since* Y *is binary, you can exploit part a*).

(c) calculate the conditional PMF of Y given $X_1 = k_1$ and $X_2 = k_2$:

$$P_{Y|X_1,X_2}(j|k_1, k_2)$$

you should have four PMFs, one for each possible k_1 and k_2 values. In light of your calculations, is Y statistically independent of X_1 and X_2? Is this consistent with the notion of Y being physically dependent upon X_1 and X_2?

(d) calculate the conditional PMFs $P_{Y|X_1}(j|0)$ and $P_{Y|X_1}(j|1)$. Is Y statistically independent of X_1? Is this consistent with the notion of Y being physically dependent upon X_1?