4 Analytical (15 Points)

1 (3 points) A basket contains black and green grapes, 30% of the grapes are sweet, the rest are sour. 40% of the grapes that are sweet are black, 20% of the sour grapes are black. What is the probability that a black grape is sweet?

- $P(Sweet) = 30\%$
- $P(Black|Sweet) = 40\%$
- $P(Black|Sour) = 20\%$
- $P(Black) = P(BlackandSweet) + P(BlackandSour) =$
- $= P(Black|Sweet) * P(Sweet) + P(Black|Sour) * P(Sour) = (0.4 * 0.3) + (0.2 * 0.7) = 0.26$
- $P(Sweet|Black) = \frac{P(Black|Sweet)*P(Sweet)}{P(Black)} = \frac{0.4*0.3}{0.26} = \frac{0.12}{0.26} = \mathbf{0.461538}$

2 (3 points) You are provided a computer program that produces random integers between 1 and 6, i.e. a die. The programmer advises you that the die results are not chosen IID. You are told that the die is biased. In order to determine its bias, you run the program for many trials, recording the number of times each number is returned. Suppose that out of n trials, there were m 1s. You are then asked to compute the probability that the next roll of the die will produce a 1. In terms of m and n, can you estimate the probability of a 1? If yes, what is it? If not, why not?

Yes you can estimate the probability of getting a 1 as $\frac{m}{n}$ which is the total number of $1's$ divided by the total number of rolls. This is because we can estimate the probability of the die, even if it is loaded, based on a large amount of previous trials.

3 (4 points) For each of the following, state if the function is a valid loss function. If it is not, why not? Note that $\hat{y}$ is the predicted label and y is the correct label.

1. $\ell(y, \hat{y}) = y - \hat{y}$
   This is a valid loss function because it can measure how far off the predicted label is from the correct label. This loss function can be minimized in order to increase accuracy since this loss function gives a reasonable measure of the cost of predicting incorrectly. However, using absolute value would make it a better function.
2. $\ell(y, \hat{y}) = (y - \hat{y})^2$
   This is a valid loss function because it can measure how incorrect the predicted label is relative to the correct label. This is a minimizable function which would reasonably evaluate the accuracy of an algorithm.
3. $\ell(y, \hat{y}) = |(y - \hat{y})|/\hat{y}$
   This is not a valid loss function. If we were to minimize this loss function, it could be minimized, for example, by always predicting a large number close to infinity. Then the value of the loss function would go to 0 and would be minimized regardless of the actual correct label values. So, this loss function is not a valid loss function.

4 (5 points) Give an example of an optimal hypothesis, a finite hypothesis class that contains the optimal hypothesis, and an infinite class that does not contain the optimal hypothesis.

Suppose there a jar with 42 marbles and we have picture data that sometimes shows all of the marbles and a few pictures where some marbles cannot be seen and counted. We want to predict the number of marbles in the jar given the picture data. For example, the data could be $x_1 = 42, x_2 = 42, x_3 = 41$ and so the optimal hypothesis would be to predict that all $x_i == 42$.

A finite hypothesis class that contains the optimal hypothesis would be to guess values in the range [40,42]
An infinite class that does not contain the optimal hypothesis would be to predict a random number N, no matter what the input is, on the range [100, ∞). (This is for someone who possibly does not know how to count and tends to be dramatic).