Statistical Foundations of Data Science
Past Homework 3

The content of this assignment is based on the seventh lecture in the course titled "Bayes Theorem".

1. A test for a rare disease comes up positive. The test is ninety percent accurate when the person is really infected and ninety-five percent accurate when the person is not infected. Surprisingly the results of taking the same test multiple times are independent. One percent of the population has the disease.

   a Compute the probability of being infected given a positive test [3]

   b Compute the probability of being infected given a negative test [3]

   c Compute the probability of being infected given two positive tests [3]

   d Compute the probability of being infected given two negative tests [3]

   e Compute the probability of being infected given two tests one of which is positive and one of which is negative [3]

2. If X and Y are independent discrete uniform random variables on the set with parameter 10, (that is equally liekly to take on any of the values in $\{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$ and never take on any other values) Compute:

   a $\mathbb{P}(X > 4)$ [1]

   b $\mathbb{P}(XY > 75)$ [3]

   c $\mathbb{P}(XY > 65)$ [3]

   d $\mathbb{P}(X > 8|XY > 75)$ [4]

   e $\mathbb{P}(Y > 7|XY > 65)$ [4]

3. Various students take different IQ tests. These tests may be on different scales, that is may have different means and variances. However in all cases the distributions will be normal

   a Albus gets a Z-score of 3.5. The mean on his test was 110 and the variance 144. Find Albus's score [4]

   b Voldemort scores 2 standard deviations below normal with a score of 70 on a test with mean 100. Find the standard deviation of scores on the test [4]

   c Hagrid scores 0.5 standard deviations below the mean. If the tests variance is 100 and Hagrid scored 105 find the tests' mean [4]

   d Harry scores right on the mean with a of 90. Find his Z-score [4]

4. The binomial distribution is a sum of IID random variables and is therefore approximately normal large n and fixed p. In our question a fair coin ($p = 0.5$) is flipped one million times. The random variable counting the number of heads is called X.

   a Compute $\mathbb{E}[X]$ [3]

   b Compute $V(X)$ [3]

c Compute the standard deviation of X [3]

d Use the CLT to estimate $\mathbb{P}(499500 < X < 501000)$ [5]

5. We have a weighted coin which was sold as coming up heads with probability $\frac{2}{3}$.

   a We flip the coin ten times and get eight heads. Set up a one-sided hypothesis test and compute the p-value. Do not use the normal approximation. [10]

   b We continue to flip until we've flipped a total of a thousand times and we've received 690 heads. Repeat the above exercise, use the normal approximation. [10]

6. A company sells chocolate bars labelled as weighing 200 grams. We know that the weights are actually normally distributed with variance $100g^2$ We buy 20 such chocolate bars and set a sample mean of $195g$. Perform a one sided hypothesis test to determine if the chocolate bars are underweight. [10]