Homework 1

1. A group of researchers has designed a new inexpensive and painless test for detecting lung cancer. The test is intended to be an initial screening test for the population in general. A positive result (presence of lung cancer) from the test would be followed up immediately with medication, surgery, or more extensive and expensive tests. The researchers know from their studies the following facts:

   - The test gives a positive result 98% of the time when the test subject has lung cancer.
   - The test gives a negative result 96% of the time when the test subject does not have lung cancer.
   - In the general population, approximately 1 person in 1000 has lung cancer.

   The researchers are happy with these preliminary results (about 97 success rate) and wish to get the test to market as soon as possible. The test would be applied to the general population.

   (a) Use Bayes' theorem to find the probability that a person has lung cancer, given a positive test result.

   (b) How would you advise those creating and administering this test?

2. The famous "Monty Hall Problem" is named after the original host of the gameshow *Let's Make a Deal*. It begins with a contestant being presented three doors. A prize is hidden behind one of the doors and a goat is behind each of the other two. A contestant initially selects one of the doors, and Monte then directs one of the other two doors to be opened, revealing a goat. He then asks the contestant whether they wish to switch their selection to the other unopened door.

   (a) Use Bayes' theorem to show that the player is more likely to win the prize if they switch doors.

   (b) Verify your answer by simulating this game many times and summarizing the outcomes under each of the two strategies (switch doors or remain with the original door).

   (c) Let θ denote the location of the prize (behind Door 1, Door 2, or Door 3) and let a denote a contestant's choice (keep original selection or switch doors). Suppose that the prize is worth \$200 to the contestant, and the goats are worth -\$10 (mostly from shame).

      i. Use this information to create a loss (or utility) function.

      ii. Calculate the expected loss (or utility) under each of the two strategies.

3. BDA3: Ch 1, Exercise 1 (point posterior with normal).

4. BDA3: Ch 1, Exercise 8 (subjective probability).

5. Airlines routinely oversell flights knowing that a few passengers will miss their flights. (`https://time.com/6197994/airlines-overbook-flights-negotiate/`.) Suppose that a particular flight has 100 seats, all equally priced at \$500 each. The airline can choose to sell N tickets (possibly more than 100) all at this price, BUT must pay \$3,000 to reimburse any passengers they are unable to accommodate. For simplicity, assume that all passengers have the same probability of showing up, and do so independently. Then, if Y is the number of passengers who show up, we have $Y \sim \text{Binomial}(N, \theta)$. For now, assume $\theta = 0.95$.

   (a) Write a utility (or loss) function representing the net revenue, defined as price minus reimbursement cost for each ticket, aggregated for this flight. It should be a function of N and Y.

   (b) Calculate the expected utility (or loss) for $N = 99, 100, 101, \ldots, 107$. (Coding this is highly recommended.)

   (c) What is the optimal choice of N to maximize expected net revenue?

   (d) Airline executives wish to account for uncertainty in θ. Historical data for this particular flight suggest the following distribution: $\theta \sim \text{Beta}(45, 2)$. Still assuming $Y \mid \theta \sim \text{Binomial}(N, \theta)$, find the marginal distribution of Y, i.e., $p(y) = \int p(y \mid \theta)\, \pi(\theta)\, \mathrm{d}\theta$.
   Hint: $\int_0^1 z^{a-1}(1-z)^{b-1}\mathrm{d}z = \Gamma(a)\Gamma(b)/\Gamma(a+b)$ for $a > 0$, $b > 0$.

   (e) Repeat parts (5b) and (5c) using the new marginal distribution for Y.

   (f) Verify your answer in part (5e) via Monte Carlo simulation.
   Hint: Simulate many (θ, Y) pairs to obtain an empirical expected net revenue.

   (g) Use your simulation to find the largest N the airline can choose so that net revenue exceeds \$40,000 with probability at least 0.9.