18.05 Problem Set 7, Spring 2022

Problem 1. (30: 10,10,10 pts.) **(Monty Hall: Sober and drunk)**
Recall the Monty Hall problem: Monty hosts a game show. There are three doors: one hides a car and two hide goats. The contestant Shelby picks a door, which is not opened. Monty then opens another door which has a goat behind it. Finally, Shelby must decide whether to stay with her original choice or switch to the other unopened door. The problem asks which is the better strategy: staying or switching?

To be precise, let's label the door that Shelby picks by A, and the other two doors by B and C. Hypothesis H_A is that the car is behind door A, and similarly for hypotheses H_B and H_C.

(a) In the usual formulation, Monty is sober and knows the locations of the car and goats. So if the contestant picks a door with a goat, Monty always opens the other door with a goat. And if the contestant picks the door with a car, Monty opens one of the other two doors at random. Suppose that sober Monty Hall opens door B, revealing a goat. So the data is: 'Monty showed a goat behind B'. Our hypotheses are 'the car is behind door A', etc. Make a Bayes table with prior, likelihood and posterior. Use the posterior probabilities to determine the best strategy.

(b) Now suppose that Monty is drunk, i.e. he has completely forgotten where the car is and is only aware enough to randomly open one of the two doors not chosen by the contestant. It's entirely possible he might accidentally reveal the car, ruining the show.

Suppose that drunk Monty Hall opens door B, revealing a goat. Make a Bayes table with prior, likelihood and posterior. Use the posterior probabilities to determine the best strategy. (Hint: the data is the same but the likelihood function is not.)

(c) Based on Monty's pre-show behavior, Shelby thinks that Monty is sober with probability 0.7 and drunk with probability 0.3. Repeat the analysis from parts (a) and (b) in this situation.

Problem 2. (40: 10,10,10,5,5 pts.) **Prediction**
We are going to explore the dice problem from class further. I have five dice (4, 6, 8, 12, or 20 sides) and pick one at random (uniform probability). I then roll this die n times and tell you that, miraculously, every roll resulted in the value 7. As I am in a position of authority, assume that I am telling the truth!

(a) First, consider just the first roll. Find the prior predictive probability that the first roll will be a 7 and the posterior (after the first roll) predictive probability that the second roll will be a 7. Also find the posterior (after the first roll) probabilities for the chosen die.

(b) Find the posterior probability $P(H|\text{data})$ for each die given the data of all n rolls (your answers should involve n). What is the limit of each of these probabilities as n grows to infinity? Explain why this makes sense.

(c) Given that my first 10 rolls resulted in 7 (i.e., $n = 10$), rank the possible values for my next roll from most likely to least likely. Note any ties in rank and explain your reasoning carefully. You need not do any computations to solve this problem.

(d) Let x_i is the result of the ith roll.

Find the posterior predictive pmf for the $(n+1)^{\text{st}}$ roll given the data. That is, find

$P(x_{n+1}|x_1 = 7, \cdots, x_n = 7)$ for $x_{n+1} = 1, \ldots, 20$. (Hint: use part (b) and the law of total probability. Many values of the pmf coincide, so you do not need to do 20 separate computations. You should check that your answer is consistent with your ranking in part (c) for $n = 10$).

(e) What function does the pmf in part (d) converge to as n grows to infinity? Explain why this makes sense.

Problem 3. (30: 10,10,5,5 pts.) **(Odds)**
You have a drawer that contains 50 coins. 10 coins have probability $p = 0.3$ of heads, 30 coins have probability $p = 0.5$ and 10 coins have probability $p = 0.7$. You pick one coin at random from the drawer and flip it.

(a) What are the (prior) odds you chose a 0.3 coin? A 0.7 coin?

(b) What are the (prior predictive) odds of flipping a heads?

(c) Suppose the flip lands heads.

(i) What are the posterior odds the coin is a 0.3 coin?

(ii) What are the posterior odds the coin is a 0.7 coin?

(d) What are the posterior predictive odds of heads on the next (second) flip?

Problem 4. (20: 10,10 pts.) **(Courtroom fallacies)**
(a) Mrs S is found stabbed in her family garden. Mr S behaves strangely after her death and is considered as a suspect. On investigation of police and social records it is found that Mr S had beaten up his wife on at least nine previous occasions. The prosecution advances this data as evidence in favor of the hypothesis that Mr S is guilty of the murder. 'Ah no,' says Mr S's highly paid lawyer, '*statistically*, only one in a thousand wife-beaters actually goes on to murder his wife. So the wife-beating is not strong evidence at all. In fact, given the wife beating evidence alone, it's extremely *unlikely* that he would be the murderer of his wife – only a 1/1000 chance. You should therefore find him innocent.'

Is the lawyer right to imply that the history of wife-beating does not point to Mr S's being the murderer? Or is this a legal trick? If the latter, what is wrong with his argument?

Use the following scaffolding to reason precisely:
Hypothesis: M = 'Mr S murdered Mrs S'
Data: K = 'Mrs S was killed', B = 'Mr S had a history of beating Mrs S'
How is the above probability 1/1000 expressed in these terms? How is the (posterior) probability of guilt expressed in these terms? How are these two probabilities related? Hint: Bayes' theorem, conditioning on B throughout.

(b) [True story] In 1999 in Great Britain, Sally Clark was convicted of murdering her two sons after each child died weeks after birth (the first in 1996, the second in 1998). Her conviction was largely based on the testimony of the pediatrician Professor Sir Roy Meadow. He claimed that, for an affluent non-smoking family like the Clarks, the probability of a single cot death (SIDS) was 1 in 8543, so the probability of two cot deaths in the same family was around "1 in 73 million." Given that there are around 700,000 live births in Britain each year, Meadow argued that a double cot death would be expected to occur once every hundred years. Finally, he reasoned that given this vanishingly small rate, the far more likely scenario is that Sally Clark murdered her children.

Carefully explain at least two errors in Meadow's argument.

Problem 5. (15 pts.) **(Bayes at the movies)**
A local theater employs two ticket collectors, Oscar and Emmy, although only one of them works on any given day. The number of tickets X that a ticket collector can collect in an hour is modeled by a distribution which has mean λ, and probability mass function

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

for $k = 0, 1, 2, \ldots$. (This distribution is called a *Poisson distribution*. It is an important *discrete* distribution in biology and physics.)

Suppose that Oscar collects, on average, 10 tickets an hour and Emmy collects, on average, 15 tickets an hour. One day the manager stays home sick. They know Emmy is supposed to work that day, but thinks there are 1 to 10 odds that Oscar is filling in for Emmy (based on Emmy's prior history of taking advantage of Oscar's generous nature when the manager is away). The suspicious manager monitors ticket sales online and observes that over the span of 5 hours there are 12, 10, 11, 4, and 11 tickets collected. What are the manager's posterior odds that Oscar is filling in for Emmy?

Problem 6. (30: 10,10,10 pts.) **(Normal is the new normal)**
Your friend transmits an unknown value θ to you over a noisy channel. The noise is normally distributed with mean 0 and a known variance 4, so the value x that you receive is modeled by $x \sim N(\theta, 2^2)$. Based on previous communications, your prior on θ is $N(6, 3^2)$.

(a) Suppose your friend transmits a value to you that you receive as $x = 5$. Use the formulas for normal-normal updating (given in the reading), to find the posterior pdf for θ.

(b) Suppose your friend transmits the same value θ to you 8 times. You receive these signals plus noise as $x_1, \ldots, x_8$ with sample mean $\bar{x} = 5$. Using the same prior and know variance σ^2 as in part (a), show that the posterior on θ is $N(5.05, 0.47)$. Plot the prior and posterior on the same graph. Describe how the data changes your belief about the true value of θ.

(c) IQ in the general population follows a $N(100, 15^2)$ distribution. An IQ test is unbiased with a known normal variance of 10^2; that is, if the same person is tested multiple times, their measured IQ will differ from their true IQ according to a normal distribution with mean 0 and variance 100.

(i) Randall Vard scored an 80 on the test. What is the posterior expected value of their true IQ?

(ii) Mary I. Taft scored a 150 on the test. What is the posterior expected value of their true IQ?

Problem 7. (15: 10,5 pts.) **(Censored data)**
Sometimes data is not reported in full. This can mean only reporting values in a certain range or not reporting exact values. We call such data censored.

We have a 4-sided die and a 6-sided die. One of them is chosen at random and rolled 5 times. Instead of reporting the number of spots on a roll we censor the data and just report

1 if the roll is a 1;   0 if the roll is not a 1.

(a) Suppose the data for the five rolls is 1, 0, 1, 1, 1. Starting from a flat prior on the choice of die, update in sequence and report, after each roll, the posterior odds that the chosen die is the 4-sided die.

(b) A censored value of 1 is evidence in favor of which die? What about 0? How is this reflected in the posterior odds in part (a)?

MIT OpenCourseWare
https://ocw.mit.edu

18.05 Introduction to Probability and Statistics
Spring 2022

For information about citing these materials or our Terms of Use, visit: https://ocw.mit.edu/terms.