18.05 Problem Set 7, Spring 2022 Solutions

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.*

Solution: In all three parts to this problem we have 3 hypotheses:

H_A = 'the car is behind door A'
H_B = 'the car is behind door B'
H_C = 'the car is behind door C'.

In all three parts the data is D = 'Monty opens door B and reveals a goat'.

The key to our Bayesian update table is the likelihoods: For part (a), since Monty is sober he always reveals a goat.

$P(D|H_A)$: H_A says the car is behind A. So, assuming H_A is true, Monty is equally likely to pick B or C and reveal a goat. Thus $P(D|H_A) = 1/2$.

$P(D|H_B)$: If H_B is true, the car is behind B and sober Monty will never choose B (and if he did it would not reveal a goat). So $P(D|H_B) = 0$.

$P(D|H_C)$: If H_c is true, the car is behind C. Since sober Monty doesn't make mistakes he will open door B and reveal a goat. So $P(D|H_C) = 1$.

Here is the table for this situation.

Hyp.	Prior	Likelihood	Bayes numerator	Posterior
H	$P(H)$	$P(\text{data}|H)$	$P(H)P(\text{data}|H)$	$P(H|\text{data})$
H_A	1/3	1/2	1/6	1/3
H_B	1/3	0	0	0
H_C	1/3	1	1/3	2/3
Total:	1	–	1/2	1

Therefore, Shelby should switch to door C, since the posterior probability H_C is true is twice that of H_A.

(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.)

Solution: Some of the likelihoods change in this setting.

$P(D|H_A)$: If H_A is true, then the car is behind A. So Monty is equally likely to show B or C and reveal a goat. Thus $P(D|H_A) = 1/2$. (Remember, D is the event 'Monty opens door B and reveals a goat'.)

$P(D|H_B)$: If H_B is true, then the car is behind B, drunk Monty might show B, but if he does we won't reveal a goat. (He will ruin the game.) So $P(D|H_B) = 0$.

$P(D|H_C)$: H_c says the car is behind C. Drunk Monty is equally likely to show B or C. If he chooses B he'll reveal a goat. So $P(D|H_C) = 1/2$.

Our table is now:

Hyp. H	Prior $P(H)$	Likelihood $P(\text{data}|H)$	Bayes numerator $P(H)P(\text{data}|H)$	Posterior $P(H|\text{data})$
H_A	1/3	1/2	1/6	1/2
H_B	1/3	0	0	0
H_C	1/3	1/2	1/6	1/2
Total:	1	–	1/3	1

So, in this case, switching is just as good (or as bad) as staying with the original choice.

The main lesson here is that, in the Bayesian framework, probability represents the uncertainty of our knowledge. Sober Monty gave us information about the door he opened (it hid a goat). But, since he would avoid the other door if the car was behind it, not picking the other door gives us additional information. Since drunk Monty was choosing a door at random, he gave us information about the door he opened, but, since he wasn't avoiding the other door, there was no information in the fact that he didn't choose it. (Imagine there are 100 doors. If you choose 1 and sober Monty opens 98 with goats, then the door he avoided almost certainly hides a car. On the other hand, drunk Monty is virtually guaranteed to spoil the show if he randomly opens 98 doors.)

(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.*

Solution: We have to recompute the likelihoods. Remember the data is that Monty chooses door B <u>and</u> reveals a goat.

$P(D|H_A)$: If the car is behind A then sober or drunk Monty is equally likely to choose door B and reveal a goat. Thus $P(D|H_A) = 1/2$.

$P(D|H_B)$: If the car is behind door B, then if he chooses it he will reveal a car, not a goat. So the probability of the data given H_B is 0, i.e., $P(D|H_B) = 0$.

$P(D|H_C)$: Let S be the event that Monty is sober and S^c the event he is drunk. From the table in (a), we see that $P(D|H_C, S) = 1$ and from the table in (b), we see that $P(D|H_C, S^c) = 1/2$. Thus, by the law of total probability

$$P(D|H_C) = P(D|H_C, S)P(S) + P(D|H_C, S^c)P(S^c) = 0.7 + \frac{1}{2}(0.3) = 0.85 = \frac{17}{20}.$$

Hyp. H	Prior $P(H)$	Likelihood $P(\text{data}|H)$	Bayes numerator $P(H)P(\text{data}|H)$	Posterior $P(H|\text{data})$
H_A	1/3	1/2	1/6	10/27
H_B	1/3	0	0	0
H_C	1/3	17/20	17/60	17/27
Total:	1	–	9/20	1

Thus, switching gives a probability of 17/27 of winning. So switching is still the best strategy. The intuitive feel for this is that even a little bit sober, Monty is giving some information by picking B, or, more precisely, avoiding C.

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.*

Solution: We start by making a Bayesian update table. Let D_4, D_6, D_8, D_{12}, and D_{20} are the hypotheses that we have selected the 4, 6, 8, 12, or 20 sided die respectively.

Hyp. H	Prior $P(H)$	Likelihood $P(\text{data}|H)$	Bayes numerator $P(H)P(\text{data}|H)$	Posterior $P(H|\text{data})$
D_4	1/5	0	0	0
D_6	1/5	0	0	0
D_8	1/5	1/8	$\frac{1}{40}$	$\frac{1}{40T} \approx 0.4839$
D_{12}	1/5	1/12	$\frac{1}{60}$	$\frac{1}{60T} \approx 0.3226$
D_{20}	1/5	1/20	$\frac{1}{100}$	$\frac{1}{100T} \approx 0.1935$
Total:	1	–	$T = \frac{31}{600} \approx 0.0517$	1

Let R_1 be the event 'the first roll is a 7'. The prior predictive probability $P(R_1) = T \approx 0.0157$, i.e. the total probability in the Bayes numerator column of the table.

The posterior probabilities $P(D_4|R_1)$, $P(D_6|R_1)$, $P(D_8|R_1)$, $P(D_{12}|R_1)$, $P(D_{20}|R_1)$ are given in the last column of the above table.

Let R_2 be the event 'the second roll is a 7'. The posterior predictive probability $P(R_2|R_1)$ is also computed using the law of total probability, where we must use the posterier probabilities for each of the dice.

$$P(R_2|R_1) = P(R_2|D_4)P(D_4|R_1) + P(R_2|D_6)P(D_6|R_1) + P(R_2|D_8)P(D_8|R_1)$$
$$+ P(R_2|D_{12})P(D_{12}|R_1) + P(R_2|D_{20})P(D_{20}|R_1)$$
$$\approx 0 \cdot 0 + 0 \cdot 0 + \frac{1}{8} \cdot 0.4839 + \frac{1}{12} \cdot 0.3226 + \frac{1}{20} \cdot 0.1935$$
$$\approx 0.0970$$

(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.

Solution: We make the Bayesian update table. This is similar to the table in part (a). The main difference is that the likelihood column needs to account for all n rolls.

Hyp. H	Prior $P(H)$	Likelihood $P(\text{data}\|H)$	Bayes numer. $P(H)P(\text{data}\|H)$	Posterior $P(H\|\text{data})$
D_4	1/5	0	0	0
D_6	1/5	0	0	0
D_8	1/5	$(1/8)^n$	$\frac{1}{5} \cdot (1/8)^n$	$\frac{1}{5T}(1/8)^n$
D_{12}	1/5	$(1/12)^n$	$\frac{1}{5} \cdot (1/12)^n$	$\frac{1}{5T}(1/12)^n$
D_{20}	1/5	$(1/20)^n$	$\frac{1}{5} \cdot (1/20)^n$	$\frac{1}{5T}(1/20)^n$
Total:	1	–	$T = \frac{1}{5} \cdot ((1/8)^n + (1/12)^n + (1/20)^n)$	1

The posterior probabilities are given in the table. To find what happens as n grows large, we rewrite the posterior probabilities by multiplying numerator and denominator by 8^n:

$$P(D_8|\text{data}) = \frac{1}{1 + \left(\frac{2}{3}\right)^n + \left(\frac{2}{5}\right)^n}$$

$$P(D_{12}|\text{data}) = \frac{\left(\frac{2}{3}\right)^n}{1 + \left(\frac{2}{3}\right)^n + \left(\frac{2}{5}\right)^n}$$

$$P(D_{20}|\text{data}) = \frac{\left(\frac{2}{5}\right)^n}{1 + \left(\frac{2}{3}\right)^n + \left(\frac{2}{5}\right)^n}$$

As $n \to \infty$, we know that $\left(\frac{2}{3}\right)^n \to 0$ and $\left(\frac{2}{5}\right)^n \to 0$. Thus, as n grows to infinity, $P(D_8|\text{data})$ approaches 1 and the posterior probability of all the other hypotheses goes to 0.

This makes sense because, as unlikely as it is, rolling n sevens is vastly more probable with the 8-sided die is vastly than with the bigger ones.

(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.*

Solution: Having observed n 7's already, we know that we could not have selected the 4-sided or the 6-sided die. We have three different groups of numbers: we can roll 1 to 8 with all three remaining dice; 9 to 12 with the 12 and 20-sided dice; and 13 to 20 with only the 20-sided die. Thus, rolling 1 to 8 are all equally likely, likewise 9 to 12 and 13 to 20. Since we can get 1 to 8 from all three dice each of these values is in the most likely group. The next most likely values are 9 to 12 which can happen on two dice. Least likely values are 13 to 20.

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

Find the posterior predictive pmf for the $(n + 1)^{st}$ 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 $n = 10$).

Solution: Let T be the total probability in the table from part (b). By the law of total probability, for $x_{n+1} = 1, 2, \ldots, 8$, we have

$$P(x_{n+1}|\text{data}) = \frac{1}{5T} \cdot \left(\frac{1}{8}\right)^n \cdot \frac{1}{8} + \frac{1}{5T} \cdot \left(\frac{1}{12}\right)^n \cdot \frac{1}{12} + \frac{1}{5T} \cdot \left(\frac{1}{20}\right)^n \cdot \frac{1}{20}.$$

For $x_{n+1} = 9, 10, 11, 12$, we have

$$P(x_{n+1}|\text{data}) = \frac{1}{5T} \cdot \left(\frac{1}{12}\right)^n \cdot \frac{1}{12} + \frac{1}{5T} \cdot \left(\frac{1}{20}\right)^n \cdot \frac{1}{20}.$$

Finally, for $x_{n+1} = 13, 14, \ldots, 20$, we have

$$P(x_{n+1}|\text{data}) = \frac{1}{5T} \cdot \left(\frac{1}{20}\right)^n \cdot \frac{1}{20}.$$

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

Solution: As $n \to \infty$, we see that $P(D_{n+1} = x|\text{data} = \text{all sevens}) \to 1/8$ for $x = 1, 2, \ldots, 8$, and 0 for $9 \le x \le 20$. This makes sense because, in the limit, all sevens makes the probability the die is 8-sided equal to 1.

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?*

Solution: Odds of A are $P(A)/P(A^c)$. So both types of coin have odds $0.2/0.8 = 1/4$.

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

Solution: To answer parts b-d we make a likelihood table and a Bayesian update table. We label our hypothesis $C_{0.3}$, $C_{0.5}$ and $C_{0.7}$ meaning that the chosen coin has that probability of heads. Our data from the first flip, D_1, is the event 'heads on the first flip'.

Likelihood table		
Hypothesis	outcomes	
Type of coin	Heads	Tails
$C_{0.3}$	0.3	0.7
$C_{0.5}$	0.5	0.5
$C_{0.7}$	0.7	0.3

Bayesian update table				
Hypoth.	Prior	likelihood	Bayes numer.	posterior
H	$P(H)$	$P(D_1|H)$	$P(H)P(D_1|H)$	$P(H|D_1)$
$C_{0.3}$	0.2	0.3	0.06	0.12
$C_{0.5}$	0.6	0.5	0.30	0.60
$C_{0.7}$	0.2	0.7	0.14	0.28
Total:	1	—	0.50	1

The prior probability of heads is the total in the Bayes numerator column: $P(\text{heads}) = 0.50$

So the prior probability of tails is $1 - P(\text{heads}) = 0.50$

So the prior odds of heads are $O(\text{heads}) = 1$, i.e. 50-50 odds.

(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?

Solution: (i) From the table we see the posterior probability the coin is the 0.3 coin is 0.12 so the posterior odds are $\boxed{\dfrac{0.12}{0.88} = \dfrac{12}{88} = 0.136}$.

(ii) Likewise the posterior odds it's the 0.7 coin are $\boxed{\dfrac{0.28}{0.72} = \dfrac{28}{72} = 0.389}$.

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

Solution: The posterior predictive probability of heads is found by summing the product of the posterior column in the Bayesian update table and the heads column in the likelihood table. We get $P(\text{heads}|D_1) = 0.12 \cdot 0.3 + 0.60 \cdot 0.5 + 0.28 \cdot 0.7 = 0.532$.

The posterior predictive probability of tails $P(\text{tails}|D_1) = 1 - 0.532 = 0.468$. So the posterior predictive odds of heads are $\boxed{O(\text{heads}|D_1) = \dfrac{0.532}{0.468} = 1.1368}$.

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.

Solution: This problem is taken from [Mackay, *Information Theory, Inference, and Learning Algorithms*].

The lawyer may correctly state that $P(M|B) = 1/1000$, but the lawyer then conflates this with the probability of guilt given all the relevant data, which is really $P(M|B, K)$. The short counterargument is that while only one in a thousand abused wives are murdered, the vast majority of those that are murdered, are killed by their abusers.

One way to format the solution is with a Bayes table. Alternatively we could use Bayes'

theorem written out long hand. Of course, they are equivalent.

Let's condition throughout on the a priori known fact B that Mr S beat Mrs S. A priori, that is before Mrs S was murdered, there are three hypotheses. They are
M: Mr S will murder Mrs S
E: Someone else will murder Mrs S
A: Mrs S will not be murdered.
(It's important to specify all three so the prior probabilities sum to 1.)

We'll take the lawyer's value of $P(M|B) = 1/1000$. Our Bayes update table is then

Hyp. H	prior $P(H\|B)$	likelihood $P(K\|H, B)$	Bayes numer.	posterior $P(H\|K, B)$
M	$1/1000$	1	0.001	$\dfrac{0.001}{p + 0.001}$
E	p	1	p	$\dfrac{p}{p + 0.001}$
A	q	0	0	0
Total:	1	$-$	$p + 0.001$	1

Here we have used $P(K|M, B) = 1$, i.e. given she was murdered by her husband the probability she was killed is 1. Likewise $P(K|E, B) = 1$ and $P(K|A, B) = 0$.

We now see that the posterior odds that Mr S is the murderer are

$$O(M|K, B) = \frac{0.001}{p}.$$

Thus, the odds depend on the value of p. Now, in most countries, $P(M|B)$ is much greater than $p = P(E|B)$, which tells us the odds overwhelmingly favor the hypothesis that Mr S is the murderer.

In fact, let's make one more assumption: $p = P(E|B) = P(E)$, i.e. that Mrs S being murdered by someone else is independent of the fact that her husband beat her. (We should acknowledge that this assumption might not be warranted without further study.) Now, our formula for the odds that Mr S is the murderer is

$$O(M|K, B) = \frac{P(M|B)}{P(E)}$$

Let's accept the lawyer's statistic that $P(M|B) = 1/1000$. A quick Wikipedia search gives that the murder rate in the US in 2020 was about 6.5/100000.* A further google search shows that the murder rate of women in the US in 2010 was about 1/100000** If Mr and Mrs S lived in the US that would put the odds at greater than 100 to 1 that he is the murderer. I would say the lawyer's argument is not credible.

The murder rate seems to be a hard statistic to pin down. Different sources give slightly different rates. Other countries have much lower murder rates. In 2017, the worldwide average was about 6.1/100000. The highest murder rates by country in 2021* was about 1/2000. This data was taken from the following Wikipedia article

https://en.wikipedia.org/w/index.php?title=List_of_countries_by_intentional_homicide_rate&oldid=1143421050

**https://en.wikipedia.org/w/index.php?title=Homicide_statistics_by_gender&oldid=1139890572

(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.

Solution: Here are four errors in the argument

1. The prosecutor arrived at "1 in 73 million" as follows: The probability that 1 child from an affluent non-smoking family dies of SIDS is $1/8543$, so the probability that 2 children die is $(1/8543)^2$. However, this assumes that the SIDS death among siblings are independent. Due to genetic or environmental factors, we suspect that this assumption is invalid.

2. The use of the figure "700,000 live births in Britain each year." The prosecutor had restricted attention only to affluent non-smoking families when (erroneously) computing the probability of two SIDS deaths. However, he does not similarly restrict his attention when considering the number of births.

3. The rate "once every hundred years" is not valid: The prosecutor arrived at this by multiplying the number of live births by the probability that two children die from SIDS. The result is a non-sensical rate.

4. While double SIDS is very unlikely, double infanticide may be even more unlikely. It is the odds of one explanation relative to the other given the deaths that matters, and not just how unlikely one possibility is.

The Sally Clark case is an example of the "Prosecutor's Fallacy." You can read about it at
https://en.wikipedia.org/w/index.php?title=Sally_Clark&oldid=629645024

There is also a video that discusses legal fallacies at
https://www.ted.com/talks/peter_donnelly_shows_how_stats_fool_juries

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 $k = 0, 1, 2, \dots$. (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?

Solution: Since we are computing odds, we will only give the hypothesis and likelihood columns of the update table. Our hypotheses are Oscar, i.e. Oscar is working and Emmy, i.e. Emmy is working. Denoting our data as D, we have the table

hypothesis	likelihood $P(D|\text{hypothesis})$
Oscar	$P(D|\text{Oscar}) = \dfrac{10^{12+10+11+4+11}e^{-50}}{12!10!11!4!11!}$
Emmy	$P(D|\text{Emmy}) = \dfrac{15^{12+10+11+4+11}e^{-75}}{12!10!11!4!11!}$

Thus the Bayes factor (likelihood ratio) is $\dfrac{P(D|\text{Oscar})}{P(D|\text{Emmy})} \approx 254.09$.

We are given the prior odds that Oscar is working, $O(\text{Oscar}) = \frac{1}{10}$. Thus, our posterior odds are

$$O(\text{Oscar}|D) = \text{prior odds} \times \text{Bayes factor} \approx \boxed{25.409}$$

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.*

Solution: For the normal-normal updating formulas given in the reading we have

$$n = 1 \text{ (number of data values)}, \quad \overline{x} = x = 5, \quad \mu_{\text{prior}} = 6, \quad \sigma^2_{\text{prior}} = 9, \quad \sigma^2 = 4$$

So, using the normal-normal update formulas:

$$a = \frac{1}{\sigma^2_{\text{prior}}} = \frac{1}{9}, \quad b = \frac{n}{\sigma^2} = \frac{1}{4}, \quad \Rightarrow \mu_{\text{post}} = \frac{a\mu_{\text{prior}} + b\overline{x}}{a+b} = \frac{69}{13}, \quad \sigma^2_{\text{post}} = \frac{1}{a+b} = \frac{36}{13}.$$

Thus, the posterior $f(\theta\,|\,x) \sim N\left(\frac{69}{13}, \frac{36}{13}\right) \approx N(5.31, 1.66^2)$.

(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.*

Solution: We have

$$n = 8, \quad \overline{x} = x = 5, \quad \mu_{\text{prior}} = 6, \quad \sigma^2_{\text{prior}} = 9, \quad \sigma^2 = 4$$

Using the update formulas we have

$$a = \frac{1}{\sigma_{\text{prior}}^2} = \frac{1}{9}, \quad b = \frac{n}{\sigma^2} = 2, \qquad \Rightarrow \quad \mu_{\text{post}} = \frac{a\mu_{\text{prior}} + b\overline{x}}{a + b} = \frac{96}{19} \approx 5.05, \quad \sigma_{\text{post}}^2 = \frac{1}{a + b} = \frac{9}{19} \approx 0.47.$$
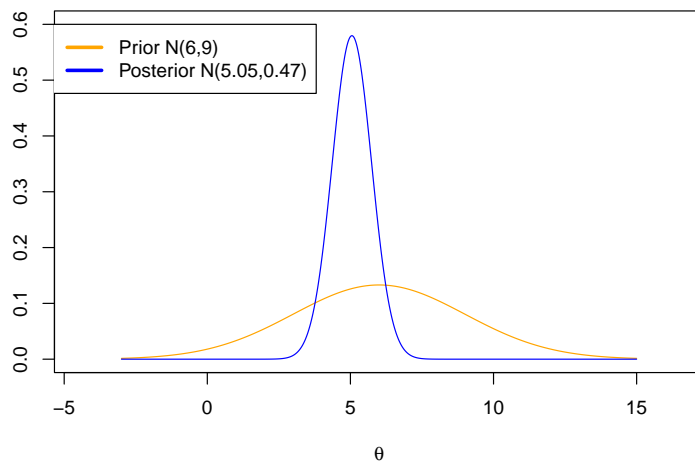
After observing $x_1, \ldots, x_4$, we see that the posterior mean is close to $\overline{x}$ and the posterior variance is much smaller than the prior variance. The data has made us more certain about the location of θ.

(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?

Solution: With no data we assume the prior mean for θ is the population average of 100, i.e. our prior is $f(\theta) \sim N(100, 15^2)$. For data $x = $ score on the IQ test we have the likelihood $f(x|\theta) \sim N(\theta, 10^2)$. Using the update formulas we have

$$n = 1, \quad \mu_{\text{prior}} = 100, \quad \sigma_{\text{prior}}^2 = 15^2, \quad \sigma^2 = 10^2.$$

So $a = 1/225$, $b = 1/100$.

(i) Randall, $x = 80$: $\mu_{\text{post}} = \dfrac{a \cdot 100 + b \cdot 80}{a + b} = 86.15.$ (This is the posterior expected value.)

(ii) Mary, $x = 150$: $\mu_{\text{post}} = \dfrac{a \cdot 100 + b \cdot 150}{a + b} = 134.62$ (This is the posterior expected value.)

Both their posterior expected values are closer to the population mean than their scores. This is the Bayesian version of regression towards the mean!

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.*

Solution: First note that we assume that, given a particular die, the rolls are independent. Let x be the censored value on one roll. The Bayes factor for x is

$$\text{Bayes factor} = \frac{p(x|\text{4-sided})}{p(x|\text{6-sided})} = \begin{cases} \frac{3/4}{5/6} = 9/10 & \text{if } x = 0 \\ \frac{1/4}{1/6} = 3/2 & \text{if } x = 1 \end{cases}$$

Starting from the prior odds of 1, we multiply by the appropriate Bayes factor and get the posterior odds after rolls 1–5 are

$$\text{Prior odds: } 1$$
$$\text{Posterior odds after roll } 1 = 1 \cdot \frac{3}{2} = 1.5$$
$$\text{Posterior odds after roll } 2 = \frac{3}{2} \cdot \frac{9}{10} = \frac{27}{20} = 1.35$$
$$\text{Posterior odds after roll } 3 = \frac{27}{20} \cdot \frac{3}{2} = \frac{81}{40} = 2.025$$
$$\text{Posterior odds after roll } 4 = \frac{81}{40} \cdot \frac{3}{2} = \frac{243}{80} = 3.0375$$
$$\text{Posterior odds after roll } 5 = \frac{243}{80} \cdot \frac{3}{2} = \frac{729}{160} = 4.55625$$

(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)?*

Solution: In part (a) we saw the Bayes factor when $x = 1$ is $3/2$. Since this is more than 1 it is evidence in favor of the 4-sided die. When $x = 0$ the Bayes factor is $9/10$, which is evidence in favor of the 6-sided die.

We saw this in part (a) because after every value of 1 the odds for the 4-sided die went up and after the value of 0 the odds went down.

MIT OpenCourseWare

https://ocw.mit.edu

18.05 Introduction to Probability and Statistics

Spring 2022