Math 4600/6600

1. More problems involving Bayes' Theorem

Here are some additional problems involving Bayes' theorem. Some are computational, while others are more theoretical. Note: Please be cautioned that some of these are based on medicine, including problems about cancer rates and stillbirths.

(1) Elvis Presley had a twin brother who died at birth. At the time, 15 births of every 1000 were births of twins, while 3 of every 1000 births were births of identical twins. What is the probability that Elvis was an identical twin? Please write several sentences to explain your reasoning.

(2) According to the CDC, "Compared to nonsmokers, men who smoke are about 23 times more likely to develop lung cancer and women who smoke are about 13 times more likely." If you learn that a woman has been diagnosed with lung cancer, and you know nothing else about her, what is the probability that she is a smoker? Note that you will have to do some research to solve this problem. Please determine what additional information you need, find it, and cite a source. Again, write several sentences to explain your reasoning.

(3) The Bayesian update over time. You've already analyzed the data in "CoinFlipData.csv" on the course webpage to arrive at a Bayesian estimate of the probability that the coin is biased. Now plot your confidence level $P(\text{biased})$ as a function of the number of flips that you have seen. Turn in the graph.

(4) The Bayesian Filter. Suppose that we are given a sequence of coin flips from a coin which has probability p of returning "heads" and probability $1 - p$ of returning tails. We will consider 11 different options for modeling the coin: $p = 0, \frac{1}{10}, \frac{2}{10}, \ldots, 1$. We will call these events $A_0, \ldots, A_{10}$ and assume that they are the only possibilities for the coin (that is, that they are a partition of the probability space).

At the start of the trial, we assign initial probabilities $P(A_i) = \frac{1}{11}$. After each flip B, we use Bayes' theorem to update all 11 of these probabilities using the formula

$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{\sum_{j=0}^{1} 0P(A_j)P(B|A_j)}$$

The file "biasedcoindata.csv" on the course webpage contains a large collection of data from such a coin.

Use a computer to analyze the data and plot the probabilities $P(A_i)$ (as a function) of i after 0 flips, 10 flips, 100 flips, 1,000 flips, 10,000 flips, and all 100,000 flips. Be sure to include a printout and explanation of your code or your spreadsheet.

If you're using a spreadsheet, it might help to set things up like this:

	A	B	C	D	E	F	G	H	I	J	K	L
1	Flip	$P(A_0)$	$P(A_1)$	$P(A_2)$	$P(A_3)$	$P(A_4)$	$P(A_5)$	$P(A_6)$	$P(A_7)$	$P(A_8)$	$P(A_9)$	$P(A_{10})$
2	T	$1/11$	$1/11$	$1/11$	$1/11$	$1/11$	$1/11$	$1/11$	$1/11$	$1/11$	$1/11$	$1/11$
3	H	(f0)	(f1)	(f2)	(f3)	(f4)	(f5)	(f6)	(f7)	(f8)	(f9)	(f10)

Here, the $1/11$ are the initial probabilities you've assigned to each of the options before seeing any data, and the different update formulae f0 through f10 for the probabilities $P(A_0)$ through $P(A_{10})$ each depend on *all* the prior probabilities in the row above (that is, on cells B2 through L2) *and* on the value of the current flip (cell A2)[1].

(5) The (over?)confident Bayesian. Suppose that you are completely certain that a belief you hold is true: your prior belief $P(A) = 1$. You observe an event B which is extremely unlikely, given your theory of the world: $P(B|A) = 0.0001$. Compute your updated level of belief in A. Does $P(B|A^c)$ matter to you? Why or why not? Can any piece of evidence change your mind?

[1]It might help to look up the IF statement if you're wondering how to make the outcome depend on whether A2 contains an H or a T.

(6) An unlikely story. You are confident that there is a $1/1000$ chance that Nessie can be seen swimming in Loch Ness on any given day. Around the Loch, there live two clans: the MacWaffles, who always tell the truth, and the MacGriddles, who always claim to see the monster (whether they did so or not). You know that 99% of the residents are MacWaffles, while 1% of the residents are MacGriddles.

At the Loch, you meet a local who states that he saw Nessie out for a swim last Tuesday. The next day, he claims to have seen Nessie swimming on Wednesday. We now have four events:

- A: Your new friend is a MacWaffle (and definitely is telling the truth).
- A^c: Your new friend is a MacGriddle (and would claim to see Nessie either way).
- B: Your new friend **claims** to have seen Nessie swimming on Tuesday.
- C: Your new friend **claims** to have seen Nessie swimming on Wednesday.

Your prior estimates of $P(A)$ and $P(A^c)$ are 0.99 and 0.01.

(a) Compute P("your new friend **actually did** see Nessie swimming on Tuesday") using your prior estimates of $P(A)$ and $P(A^c)$. (Note: This is not a Bayes theorem problem.)

(b) Compute $P(A|B)$, the updated probability that your friend is a MacWaffle given his claim about the monster sighting. (Note: This is a Bayes theorem problem.)

(c) Compute P("your new friend **actually did** see Nessie swimming on Wednesday") using your updated estimate $P(A) \to P(A|B)$ that your friend was a MacWaffle after his first monster report.

(d) Compute a new update $P(A|C)$ of the probability that your friend is a MacWaffle after the second report.

(e) Now go back and recompute P("your new friend **actually did** see Nessie swimming on Tuesday") using your current (twice-updated) value for $P(A)$. Did the second claimed sighting increase or decrease your belief in the original sighting? Why?

Describe your conclusions in a brief paragraph. What should we make of (apparently) reliable witnesses who describe (apparantly) very unlikely events? Draw conclusions for your day-to-day life.