Conditional Probability and Bayes' Rule

Lab Assignment

By: Kenneth Marenco

(10) 1. Consider the experiment of flipping three fair coins. Let $X = 1$ if the first flip is H and -1 if not. Let $Y = 1$ if the second flip is H and -1 if not. Let Z be the number of heads on the last two flips. Let W be the number of heads on the first two flips.

(a) Explain what it means for two random variables to be *independent*.

If two random variables are independent, they do not influence each other. A consequence of this that will showcase that is that P(C|D) does not equal P(C) if they are dependent.

(b) For each pairwise combination of these four random variables, explain why the pair is independent or not independent.
(For four random variables, there are 6 pair-wise combinations of random variables.)

The pair of X and Y are independent because P(X|Y) = P(X) and P(Y|X) is P(Y)

The pair X and Z are independent because Z does not take into account the outcome of the first flip (only the last two).

The pair X and W are dependent because if X equals 1 then W cannot equal 0 which means W deals with the universe of X.

The pair Y and Z are dependent because if Y equals 1 then Z cannot equal 0.

The pair Y and W are dependent because if Y equals 1 then W cannot equal 0.

The pair Z and W are dependent because if Z equals 2 then W must be at least 1.

(20) 2. Consider the experiment of rolling two six-sided die, one of them red and one of them black. Let the event A be "the red die came up 5" and let the event B be the event "the total roll came up 7".

(a) What is the sample space Ω?

(b) Compute $P(A)$.

P(A) = 1/6

(c) Compute $P(B)$.

P(B) = P(1 and 6) + P(6 and 1) + P(2 and 5) + P(5 and 2) + P(3 and 4) + P(4 and 3) = 6/36 = 1/6

(d) Compute $P(A|B)$.

[P(A) * P(B)] /P(B) = 1/6 * 1/6 / 1/6 = 1/6

(e) Compute $P(B|A)$.

[P(B) * P(A)] / P(A) = 1/6 * 1/6 / 1/6 = 1/6

(f) Are events A and B independent?

The events are dependent because the effects of A lead to limitations of the black die to satisfy event B. If the red die landed on a 1, the black die must then be a 6 in order to achieve event B.

(10) 3. Suppose a box contains three coins. Two of the coins are fair coins with H and T as sides. One of the coins has H on both sides!

(a) Pick a coin uniformly at random (don't look at it!) and toss it. What's the probability that it lands H?

The probability of heads is
P(H) = P(H|Fair) * P(Fair) + P(H|Unfair)*P(Unfair) = 1/2 * 2/3 + 1 * 1/3 = 4/6 = 2/3

(b) Suppose you picked a coin at random, tossed it, and got H. What is the probability that you chose the double-H coin?

P(Unfair) = P(Fair|H) * P(H) + P(Unfair|H) * P(H) = 1/3 * 2/3 + 1/3 * 2/3 = 4/9

(10) 4. Here we want you to actually *prove* Bayes' Rule. It's just a small amount of algebra, but as we've seen, the result is really powerful!

Suppose that we run some experiment, and that A and B are two events, each with non-zero probability.

(a) Write down the definitions of the conditional probabilities $P(A|B)$ and $P(B|A)$.

P(A|B) = [P(B|A) * P(A)]/P(B)

P(B|A) = P(A|B) * P(B)]/P(A)

(b) Using the formulae above, and some basic algebra and reasoning, prove Bayes' Rule.

Probability P(A|B) is trying to calculate the posterior belief of event A given event B occurred. This is done by multiplying the likelihood by the the prior belief P(B|A) * P(A). This product is then divided by the evidence P(B).

(20) 5. Suppose you go to the doctor for a routine checkup. At this checkup, you are randomly selected to be tested for Very Bad Syndrome (VBS). You test positive.

(a) You want to know how worried you should be. As a reasonable consumer of medical information, what questions should you now ask? (Don't turn the page yet, or it'll spoil this question!)

Questions that should be of importance is the True Positive Rate and the False Positive Rate

(b) Suppose that current studies tell us that every 1 in $10,000$ people has VBS. The false positive rate for this particular VBS test is .01, and the true positive rate is 0.999. After hearing the news about your positive test, what's your estimate of the probability that you have VBS?

The probability of having VBS is going to be the true positive rate since this is in the world of receiving a positive test result or P(having VBS| positive test).

P(positive) = .999 * 1/10,000 + .01 * 9,999/10,000 = .0100989

P(positive | have disease) = .999

P(positive | no disease) = .01

$$P(\text{have disease} \mid \text{positive}) = \frac{P(\text{positive} \mid \text{have disease}) * P(\text{have disease})}{P(\text{positive})} = \frac{.999 * 1/10{,}000}{.0100989} = .0098922$$

(c) A different person comes into the office and also tests positive. It turns out that this person was recently on a cruise ship, where there was a VBS outbreak. Estimates from the ship's doctors tell us that 1 in 20 of the passengers on that ship disembarked with VBS. How worried should that person be before the positive test? After the positive test?

The person should be more worried than I was since the likelihood of him actually contracting it is much higher. His probability before the test would be P(cruise rate) = 1/20. After receiving the positive test, his new probability is the world he receives a positive test.

$$P(\text{have disease} \mid \text{positive}) = \frac{.999 * 1/20}{.05945} = .840202$$

P(positive) = .999 * 1/20 + .01 * 19/20 = .05945

(d) Why are the probabilities of actually having VBS in these two scenarios so different?

The biggest reason is that the probability of them testing positive in any case drastically changes once the exposure rate changes from 1/10,000 to 1/20

P(having VBS | cruise) = .999 * 1/20 = .04995

P(having VBS | average life) = .999 * 1/10,000 = .0000999

(20) 6. (a) Write pseudo-code for Monte Carlo simulations to experimentally verify your answers to both previous parts of this question.

using numpy
create a sample size large enough to lower the standard deviation of each run
list probabilities known such as VBS rate for the cruise and regular exposure
create for loops that will check the random values generated from np.random.random_sample between the range (0,1) are lower then their given percentage. If they do, they are considered to have received a positive VBS test.
These individuals are then compared against the false positives and the true positives
The probabilities of each run are then averaged

(b) Code and run your Monte Carlo simulations. How do the probabilities estimated via Monte Carlo simulation compare to your hand-calculated probabilities? (Be sure to upload your code as part of submitting this lab.)

The experimental values were very similar to those calculated. However, there was a quite a bit of variance. The cruise's probability of having the disease had values ranging from .81-.864. On a similar note, the values seen for normal exposure had values from .0078 - .0158. These probability runs were averaged using a sample size of 50,000 individuals.

(20) 7. The color distribution in a bag of M&M's has changed over the years, no doubt in response to intense market research! Your friend comes to you with a 1994 bag and a 1996 bag. She gives you one yellow M&M and one green M&M, and tells you that exactly one came from each bag.

In 1994, the bags contained the following colors with percentages in parentheses: brown (30), yellow (20), red (20), green (10), orange (10), and tan[1] (10).

By 1996, tan was gone and had been replaced by blue. The new color distribution was brown (13), yellow (14), red (13), green (20), orange (16), and blue (24).

What is the probability that your friend picked the yellow candy from the 1994 bag? (Hint: this problem is not as easy as it might look! Make sure that you are in fact using all of the evidence available to you!)

P(yellow | 1994) = 20/100 = 1/5

P(green | 1994) = 10/100 = 1/10

P(yellow | 1996) = 14/100 = 7/50

P(green | 1996) = 20/100 = 1/5

P(green) = P(green | 1994) * P(1994) + P(green | 1996) * P(1996) = 1/10 * 1/2 + 1/5 * 1/2 = 3/20

$$P(1996 \mid green) = \frac{P(green \mid 1996) * P(1996)}{P(green)} = \frac{1/5 * 1/2}{3/20} = \frac{1/10}{3/20} = 2/3$$

$$P(green \mid 1996) = \frac{P(1996 \mid green) * P(green)}{P(1996)} = \frac{2/3 * 3/20}{1/2} = 1/5$$

P(yellow 1994 | green 1996) = 1/5 * 1/5 = 1/25

[1] Yes, Tan!