STAT 5444G
Homework Assignment 1

Colburn Hassman

February 7, 2021

[3] 1. What are the two ways of quantifying probabilities and give an example for each.

Probabilities can be defined as either *discrete* or *continuous*, with discrete probabilities taking 1 of a set of n distinct values from the sample space. This can be, for example, the number of letters I receive in the mail everyday. The sample space would be all positive real numbers, but it is still discrete, there is no possibility for 0.2 letters. Continuous variables exist on a continuum. An example of this would be the percent daily change in the price of a bushel of soybeans (assuming infinite precision).

---

[3] 2. For the example in Unit 1, what if the result of the mammogram instead had been negative? Write out a table like in slide 23 of Unit 1 for a negative mammogram. What is the posterior probability that the friend will be diagnosed with breast cancer in this case?

Model	Prior	Likelihood for M-	Prior X Likelihood	Post
Cancer	0.0045	0.276	0.001242	0.00013
No Cancer	0.9955	0.973	0.968622	0.99987

As displayed in the table, if the result of the mammogram is negative, it is very unlikely (0.013%) that the friend has breast cancer.

---

[10] 3. This problem is based on an example on pages 9 –11 of Gelman et al. (2004). Blue Hemophilia is a rare hereditary bleeding disorder caused by a defect in genes that control the body's production of blood-clotting factors. It occurs almost exclusively in males. However, women may be carriers of the hemophilia gene. Females carriers of the hemophilia gene usually show no physical symptoms of hemophilia. A son born of a woman who is a hemophilia carrier and a man who does not have hemophilia has a 0.5 probability of inheriting hemophilia from his mother. A son born of a woman who is not a carrier and a man who does not have hemophilia has zero probability of inheriting hemophilia.

Danielle is a young married woman. Her husband does not have hemophilia. Because Danielle's mother is known to be a carrier of hemophilia, there is a 0.5 probability that Danielle inherited a hemophilia gene from her mother and is also a carrier. We may consider two possible "models": Danielle is a carrier, and Danielle is not a carrier. Danielle gives birth to three sons. None of them are identical twins, and we will consider their hemophilia outcomes to be independent conditional on her carrier status. For each of the sons, we will define a random variable Yi that takes on the value 1 if the son has hemophilia and 0 if he does not.

A. Use the information given above to determine

(**i**) Prior probabilities for the two possible "models" (carrier and not carrier) evaluated before Danielle gives birth to her first son.

Model	Prior
Carrier	0.5
Not Carrier	0.5

The prior probabilities are based off the fact that Danielle's mother is a known carrier, and thus Danielle has a 50% likelihood of being a carrier herself.

(**ii**) Two different sets of likelihood probabilities, one for each model Using the notation above, for the ith son, $Y_i = 0$ indicates the son is not affected by hemophilia; Y_i =1 indicates that he is affected. For each model, you will need P($Y_i = 0$|model) and P($Y_i = 1$|model).

Model	Prior	P($Y_i = 1$ |Model)	P($Y_i = 0$ |Model)
Carrier	0.5	0.5	0.5
Not Carrier	0.5	0.0	1.0

B. Now you learn that the three outcomes: $Y_1 = 0$, $Y_2 = 1$, $Y_3 = 0$. Do a sequential Bayesian analysis in which you compute the posterior probability that the woman is a carrier using the data from each son one at a time. For each step, use Bayes' rule and make a table with columns for model, prior probabilities, likelihood given observed data, product, and posterior probabilities.

Iteration 1

Model	Prior	P(Y_1 |Model)	Prior x Likelihood	Posterior
Carrier	0.5	0.5	0.25	1/3
Not Carrier	0.5	1	0.5	2/3

Iteration 2

Model	Prior	P(Y_2 |Model)	Prior x Likelihood	Posterior
Carrier	1/3	0.5	1/6	1
Not Carrier	2/3	0	0	0

Iteration 3

Model	Prior	P(Y_3 |Model)	Prior x Likelihood	Posterior
Carrier	1	0.5	0.5	1
Not Carrier	0	1	0	0

C. Also answer the following questions:

(**i**) What was the posterior probability that the woman was a carrier after the first son's status became known?

(**ii**) Did the posterior probability change based on the data from the second son? Why or why not?

(**iii**) What is the posterior probability that the woman was a carrier based on the data after the third son?

The posterior probability that the women carried Hemophilia was $\frac{1}{3}$ after her first son's status was know. Her posterior probability was smaller than her prior probability because having a son without Hemophilia provided evidence that she may not be a carrier. The data from her second son changed her posterior probability from $\frac{1}{3}$ to 1. Given P(Y_i=1|Not Carrier) = 0, P(Carrier|Y_i=1) = 1. Because her son has Hemophilia, she *must* be a carrier. The posterior probability after Iteration 3 did not change from Iteration 2, because having one son with Hemophilia confirmed without a doubt that she was a carrier: her prior P(Carrier) = 1 and P(Not Carrier) = 0. Even if she had multiple more sons without Hemophilia, one son with is proof that she is a carrier, despite lack of further evidence.

---

[4] 4. In the early 1980s, HIV had just been discovered and was rapidly expanding. There was major concern with the safety of the blood supply. Also, virtually no cure existed making an HIV diagnosis basically a death sentence, in addition to the stigma that was attached to the disease. These made false positives and false negatives in HIV testing highly undesirable. A false positive is when a test returns positive while the truth is negative. That would for instance be that someone without HIV is wrongly diagnosed with HIV, wrongly telling that person they are going to die and casting the stigma on them. A false negative is when a test returns negative while the truth is positive. That is when someone with HIV undergoes an HIV test which wrongly comes back negative. The latter poses a threat to the blood supply if that person is about to donate blood. The HIV test that has been considered was an enzyme-linked immunosorbent assay, commonly known as an ELISA. The P(ELISA is positive Person tested has HIV)=0.93 and P(ELISA is negative Person tested has no HIV)=0.99. The prevalence of HIV in the overall population, which is estimated to be 1.48 out of every 1000 American adults so P(Person tested has HIV)=1.48/1000=0.00148.

What is the probability that someone (in the early 1980s) has HIV if ELISA tests positive?

Model	Prior	P(T+ |Model)	Prior x Likelihood	Posterior
HIV Positive	0.00148	0.93	0.001376	0.1219
HIV Negative	0.99852	0.01	0.009852	0.8781

The probability that someone has HIV given a positive ELISA test is approximately 12.2% P(T+|HIV-) was determined by 1 - P(T- |HIV-).
