EECS-70: Discrete Mathematics and Probability Theory **Spring 2020**

Topic Specific Review: Bayesian Inference

Discussion, *Rohan*

Note: *LaTeX template courtesy of UC Berkeley EECS dept.*
Disclaimer: *This latex is adapted from a template that was republished Berkeley EECS materials republished on a cmu server. The material itself is based on some past homeworks from Berkeely EECS 70.*

1.1 Medical Testing

1.1.1 a

Suppose, 1 in 1000 citizens have a novel sickness, and there is a test you can use for the illness (for a price of course!) The makers advertise that the test is 99 percent reliable. They explain that healthy patients come back with a negative test exactly 99 percent of the time. Truly sick patients results come back with a positive test 99 percent of the time. A randomly selected citizen came back positive. What is the probability this randomly selected citizen is actually infected.

1.1.1.1 Solution

By Bayes Rule

$$P(sick|+) = \frac{P(+|sick) * P(sick)}{P(+)}$$

$$= \frac{P(+|sick) * P(sick)}{P(+|sick) * P(sick) + P(+|healthy) * P(healthy)}$$

From the problem:

$$P(sick) = 1/1000$$

$$P(+|sick) = .99$$

$$P(+|healthy) = .01$$

$$P(healthy) = 1 - \frac{1}{1000} = .999$$

Plugging in and evaluating...

$$P(sick|+) = \frac{.99 * 1/1000}{.99 * 1/1000 + .01 * .999} \approx .09$$

So much for the reliable tests! One way to interpret this result, is that it was originally so unlikely to have the disease, that even this strong evidence can't quite sway your beliefs.

1.1.2 b

The citizen was woefully unsatisfied with how unconvincing his "positive" test results were in part a, given the cost of treatment. So s/he takes the test a second time, and it comes back positive. What is the probability that the citizen now actually has the disease?

1.1.2.1 Solution

One way to look at this is we now think there is a .09 chance he is sick (as our "prior belief") before the test. Repeating the same analysis with:

$$P(sick) = .09$$

$$P(+|sick) = .99$$

$$P(+|healthy) = .01$$

$$P(healthy) = .91$$

$$P_{new}(sick|+) = \frac{.09 * .99}{.09 * .99 + .91 * .01} \approx .91$$

We still aren't completely sure, but now we should probably get this citizen to treatment.

1.1.3 c

(Open Ended, don't answer numerically): Suppose you felt very ill and demonstrated symptoms of this novel disease and decided to get tested for the disease, and it came back positive. How sure can you be that you have the disease? Is it the same probability as your answer to a?

1.1.3.1 Solution

Now, we don't have a good estimate for the probability because we don't know what portion of people without the disease demonstrate symptoms, and conversely what portion of the population with the disease have symptoms. Most likely, showing these symptoms is a pretty big sign you should get treated for the disease, and our analysis in a doesn't account for this. This is the spirit of Bayesian inference: using all evidence to update our prior beliefs.

1.1.4 d

(Challenge) In healthy citizens the amount of a certain protein is normally distributed $N(\mu_0 = 100, \sigma^2 = 100)$ but in infected individuals the distribution is better modeled by $N(\mu_1 = 160, \sigma^2 = 100)$. Now a randomly selected citizen is administered a new, more sophisticated (ideal) test which measures an exact protein level of 170, what is the probability the citizen has the illness, if 1/1000 citizens have the illness?
Hint: Use, without proof, the fact the probability density distribution of a guassian is given by.

$$f_x(x) = 1/sqrt(2 * \pi * \sigma^2) * e^{-\frac{(x-\mu)^2}{2*\sigma^2}}$$

Hint: Let X denote the patients protein level, and consider:

$$P(x < X < x + \epsilon|sick)$$

And then consider the limiting behavior as ϵ becomes small. .

1.1.4.1 Solution

Bayes rule

$$P(sick|x < X < x + \epsilon) = \frac{P(x < X < x + \epsilon|sick) * P(sick)}{P(x < X < x + \epsilon)}$$

$$= \frac{P(x < X < x + \epsilon|sick) * P(sick)}{P(x < X < x + \epsilon|sick) * P(sick) + P(x < X < x + \epsilon|healthy) * P(healthy)}$$

For small epsilon this approximately equals

$$= \frac{f_{sick}(x) * P(sick) * \epsilon}{f_{healthy}(x) * P(healthy) * \epsilon + f_{sick}(x) * P(sick) * \epsilon}$$

Note that $f_{sick}(x)$ is the probability density function of X assuming the patient is sick. We can then cancel epsilon from the top and bottom.

$$= \frac{f_{sick}(x) * P(sick)}{f_{healthy}(x) * P(healthy) + f_{sick}(x) * P(sick)}$$

$$= \frac{e^{-\frac{1}{2}(\frac{(x-\mu_1)}{\sigma})^2}}{e^{-\frac{1}{2}(\frac{(x-\mu_1)}{\sigma})^2} + e^{-\frac{1}{2}(\frac{(x-\mu_0)}{\sigma})^2}}$$

Plugging in values:

$$= \frac{e^{-1/2}}{e^{-1/2} + e^{-49/2}}$$

$$= \frac{1}{1 + e^{-24}}$$

$$\approx 0.999999...$$

$$\approx 1$$

We can be so sure, that a "seven sigma event" (aka, an event that is seven standard deviations (square roots of variance)) won't happen that some online calculators will round it off to one. However, as we've mentioned this value only holds if our modelling assumptions are valid: perhaps, there are some outliers healthy people with elevated levels, that weren't accounted for in the modeling phase, so use of this probability should be accompanied with such an understanding.