Bayesian Viral Load Test

Question from Fall 2022 Stanford Midterm

We are going to build a Bayesian Viral Load Test which updates a belief distribution regarding a patient's viral load. Though viral load is continuous, in our test we represent it by discretizing the quantity into whole numbers between 0 and 99, inclusive. The units of viral load are the number of viral instances per million samples.

If a person has a viral load of 9 (in other words, 9 viruses out of every 1 million samples) what is the probability that a random sample from the person is a virus?

$$\frac{9}{1,000,000}$$

We test 100,000 samples from one person for the virus. If the person's true viral load is 9, what is the probability that exactly 1 of our 100,000 samples is a virus? Use a computationally efficient approximation to compute your answer. Your approximation should respect that there is 0 probability of getting negative virus samples.

Let's define a random variable X, the number of samples that are viral given the true viral load is 9. The question is asking for $P(X = 1)$. We can think about this as a binomial process, where the number of trials n is the number of samples and the probability p is the probability that a sample is viral.

$$n = 100,000, p = \frac{9}{1,000,000}$$

Notice that n is very small and p is very large, so we can use the Poisson approximation to approximate our answer. We find $\lambda = np = 100,000 \cdot 9/1,000,000 = 0.9$, so $X \sim$\Poi$(\lambda = 0.9)$. The last step is to use the PMF of the Poisson distribution.

$$P(X = 1) = \frac{(0.9)^1 e^{-0.9}}{1!}$$

Based on what we know about a patient (their symptoms and personal history) we have encoded a prior belief in a list `prior` where `prior[i]` is the probability that the viral load equals i. `prior` is of length 100 and has keys 0 through 99.

Write an equation for the updated probability that the true viral load is i given that we observe a count of 1 virus sample out of 100,000 tested. Recall that $0 \leq i \leq 99$. You may use approximations.

We want to find

$$P(\text{viral load} = i|\text{observed count of}\frac{1}{100000})$$

We can apply Bayes Rule to get

$$= \frac{P(\text{observed count of}\frac{1}{100000}|\text{viral load} = i)P(\text{viral load} = i)}{P(\text{observed count of}\frac{1}{100000})}$$

We know that we can define a random variable $X \sim$ observed count out of 100,000$|$viral load $= i$, and we can model X as a Poisson approximation to a binomial with $n = 100000$ and $p = \frac{i}{1000000}$, with

$$\lambda = np = 100000 \cdot \frac{i}{1000000} = \frac{i}{10}$$

So X can be written as

$$X \sim Poi(\lambda = \frac{i}{10})$$

Now we can rewrite our Bayes Rule equation as

$$= \frac{P(X = 1)P(\text{viral load} = i)}{P(\text{observed count of}\frac{1}{100000})}$$

We can now use the Poisson PMF and our given \texttt{prior} to get:

$$= \frac{\frac{\frac{i}{10}e^{\frac{-i}{10}}}{1!} \cdot \texttt{prior[i]}}{P(\text{observed count of}\frac{1}{100000})}$$

We now need to expand our denominator. We can use the General Law of Total Probability to expand

$$P(\text{observed count of}\frac{1}{100000}) = \sum_{j=0}^{99} P(\text{observed count of}\frac{1}{100000}|\text{viral load} = i)P(\text{viral load} = i)$$

We can rewrite this as

$$= \sum_{j=0}^{99} \frac{\frac{j}{10}e^{\frac{-j}{10}}}{1!} \cdot \texttt{prior[j]}$$

$$= \sum_{j=0}^{99} \frac{j}{10}e^{\frac{-j}{10}} \cdot \texttt{prior[j]}$$

And finally, we can plug this in to get

$$\boxed{\frac{\frac{i}{10}e^{\frac{-i}{10}} \cdot \texttt{prior[i]}}{\sum_{j=0}^{99} \frac{j}{10}e^{\frac{-j}{10}} \cdot \texttt{prior[j]}}}.$$