Bayes' Theorem

G. Palafox

October 27, 2020

1 Introduction

The use of diagnostic tests is widespread in modern medicine. As Fletcher and Fletcher [2005] mention in their book, *establishing diagnoses is an imperfect process, resulting in a probability rather than a certainty of being right.* Since decisions are made based on the results of these tests, the correct interpretation of their outcome is important, and probability theory can be used to aid with this understanding.

2 SARS-CoV-2

Currently, a strain of coronavirus (SARS-CoV-2, colloquially known as Covid-19) is causing havoc around the world. Diagnostic tests are being used to fight the pandemic, mainly by requiring quarantine for people who test positive, or allowing less movement restrictions for those who test negative (i.e., not quarantining, permitting travel or entrance to places). Given this, it is essential to have a clear understanding of what test results mean. Bayes' theorem, which can be seen in Theorem 1, can help with these interpretations.

Theorem 1 (Bayes' theorem)**.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let $\{B_i\}_{i\in\mathbb{N}}$ be a partition of Ω such that $\mathbb{P}(B_i) > 0$ for each i, and let A be any event with positive probability. Then*

$$\mathbb{P}(B_i \mid A) = \frac{\mathbb{P}(A \mid B_i)\mathbb{P}(B_i)}{\sum_{j=1}^{\infty} \mathbb{P}(A \mid B_j)\mathbb{P}(B_j)}. \tag{1}$$

There are four possible outcomes with any diagnostic test: **true positive**, where a person with the disease tests positive, **false positive**, where a person does not have the disease yet tests positive, **false negative**, where a person with the disease tests negative, and **true negative**, where a person does not have the disease and tests negative. Additionally, with any test, we associate two measurements of how reliable it is: *specificity*, which is the percentage of true negatives out of healthy people, and *sensitivity*, which is the percentage of true positives among people with the disease. Let us denote by $\pm$test the events of having a positive or negative test, and $\pm$cov the events of having or not Covid-19. Using this we can express the sensitivity and specificity of a test as conditional probabilities, namely,

$$\mathbb{P}(+\text{test} \mid +\text{cov}) = \text{sensitivity}, \quad \mathbb{P}(-\text{test} \mid -\text{cov}) = \text{specificity}. \tag{2}$$

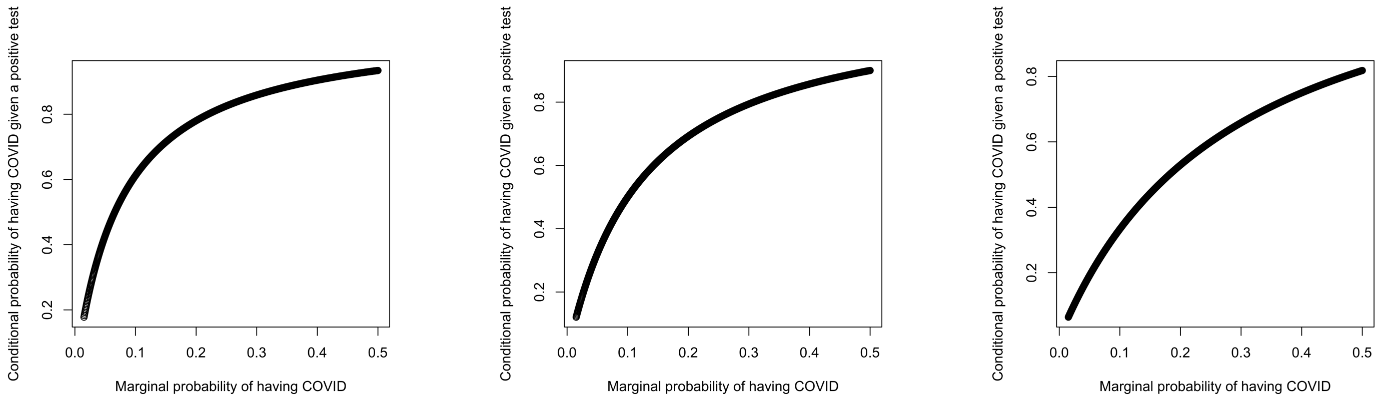
If specificity and sensitivity are know, all is left to know is the marginal probability $\mathbb{P}(+\text{cov})$, and Bayes' theorem can give the probability of having Covid-19 given that a test is positive as

$$\mathbb{P}(+\text{cov} \mid +\text{test}) = \frac{\mathbb{P}(+\text{test} \mid +\text{cov})\,\mathbb{P}(+\text{cov})}{\mathbb{P}(+\text{test} \mid +\text{cov})\,\mathbb{P}(+\text{cov}) + \mathbb{P}(+\text{test} \mid -\text{cov})\,\mathbb{P}(-\text{cov})}. \tag{3}$$

However, two issues arise. The first one is that in the case of Covid-19, sensitivity and specificity are largely unknown for the widely used PCR test [West et al., 2020]. Watson and Whiting [2020] found a sensitivity ranging from 71% to 98%, and use a specificity of 95%. The second issue is the non-trivial calculation of $\mathbb{P}(+\text{cov})$. Ranjan [2020] and Lewis [2020] calculate it as the number of confirmed cases divided by the total population. Schnipper and Sax [2020] consider half the ratio of positive tests in a given population as an estimate of $\mathbb{P}(+\text{cov})$. Others [Ming Chan, 2020, Good et al., 2020] vary $\mathbb{P}(+\text{cov})$, since it can naturally vary depending on who you are testing (random people, people with symptoms, hospital workers, etc.). Personally, the author feels the latter approach gives more insight, since it is adapted easier to different scenarios.

2.1 The case of Nuevo León

Consider the case of the Mexican state of Nuevo León. As of October 24, 2020, Nuevo Leon had 77807 confirmed Covid cases in a 5.4 million population [Gobierno del Estado de Nuevo León, 2020a, Secretaría de Economía y Trabajo de Nuevo



(a) Sensitivity of 71%, specificity of 95%.

(b) Sensitivity of 90%, specificity of 90%.

(c) Sensitivity of 90%, specificity of 80%.

Figure 1: Marginal vs. conditional probability, varying specificity and sensitivity of tests.

León, 2020]. If $\mathbb{P}(+\text{cov})$ is calculated as confirmed cases over population, that would give $\mathbb{P}(+\text{cov}) = 0.0144$. Assuming a sensitivity of 71% and specificity of 95%, this would give

$$\mathbb{P}(+\text{cov} \mid +\text{test}) = \frac{(.71)(0.0144)}{(.71)(0.0144) + (.05)(.9856)} = 0.17. \tag{4}$$

It may seem counter-intuitive that a positive test gives you only a 17% of having the disease, but this is a consequence of the seemingly low probability of being infected. Consider, on the other hand, that around 40% of tests performed in Nuevo Leon turn out positive [Gobierno del Estado de Nuevo León, 2020b]. If we use half of this value as $\mathbb{P}(+\text{cov})$, Bayes' theorem would give $\mathbb{P}(+\text{cov} \mid +\text{test}) = 0.78$. As it is seen in Figure 1, how widespread the disease is (measured by the marginal probability of being infected) impacts greatly on the interpretation of the test. Specificity and sensitivity, which are not well known for the Covid-19 test, also affect greatly[1].

3 HIV

Unlike the tests for the new coronavirus, tests for HIV have a well-established specificity and sensitivity of over 99% [CDC, 1998]. Of course, the conditional probability when interpreting a positive result still depends on the prevalence of HIV in the community, i.e., in the marginal probability of being HIV positive. However, this allow us to vary only that parameter, fixing the specificity and sensitivity of the test. For reference, considering $\mathbb{P}(+\text{hiv}) = 1.2/327.5 = 0.003$ (cases in the United States over population of the United States, both at the end of 2018 [United States Census Bureau, 2020, CDC, 2020]), would give $\mathbb{P}(+\text{hiv} \mid +\text{test}) = 0.78$, using a sensitivity and specificity of 99%. Varying this marginal probability, with sensitivity and specificity fixed, changes the conditional probability as seen in Figure 2.

References

CDC. Current trends update: Serologic testing for antibody to human immunodeficiency virus. *MMWR*, January 1998. https://www.cdc.gov/mmwr/preview/mmwrhtml/00051681.htm.

CDC. HIV basics. Basic statistics, 2020. https://www.cdc.gov/hiv/basics/statistics.html.

R. H. Fletcher and S. W. Fletcher. *Clinical Epidemiology: The Essentials*. Fourth edition, 2005.

Gobierno del Estado de Nuevo León. Casos de COVID en Nuevo León, 2020a.

Gobierno del Estado de Nuevo León. Monitoreo de indicadores estatales de salud para la reapertura económica semana 42 (11 de octubre – 17 de octubre), 2020b. https://www.nl.gob.mx/presentacion-indicadoresymedidasdemitigacion-covid19-22-10-2020.

C. B. Good, I. Hernandez, and K. Smith. Interpreting COVID-19 test results: a Bayesian approach. *Journal of General Internal Medicine*, 35(8):2490–2491, jun 2020. doi: 10.1007/s11606-020-05918-8. URL https://doi.org/10.1007/s11606-020-05918-8.

[1]The notebook with the code creating these graphics, as well as this report, can be found in the Github Repository: https://github.com/palafox794/AppliedProbabilityModels/tree/master/Assignment8

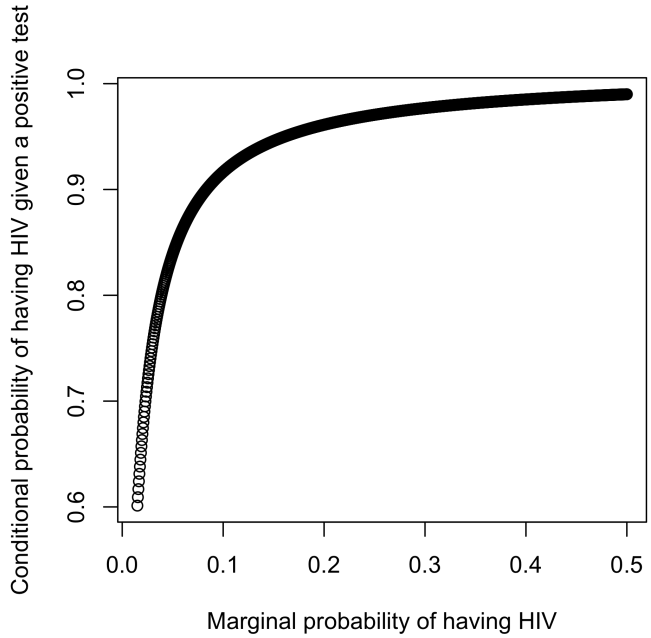


Figure 2: Marginal vs. conditional probability, for a test with specificity and sensitivity of 99%.

M. A. Lewis. Bayes theorem and Covid-19 testing, 2020. https://www.significancemagazine.com/science/660-bayes-theorem-and-covid-19-testing.

G. Ming Chan. Bayes' theorem, COVID19, and screening tests. *The American Journal of Emergency Medicine*, jun 2020. doi: 10.1016/j.ajem.2020.06.054. URL https://doi.org/10.1016/j.ajem.2020.06.054.

A. Ranjan. Covid-19, Bayes' theorem and taking probabilistic decisions, 2020. https://towardsdatascience.com/covid-19-bayes-theorem-and-taking-data-driven-decisions-part-1-b61e2c2b3bea.

J. L. Schnipper and P. E. Sax. Covid-19 test accuracy supplement: the math of Bayes' theorem, 2020. https://www.statnews.com/2020/08/20/covid-19-test-accuracy-supplement-the-math-of-bayes-theorem/.

Secretaría de Economía y Trabajo de Nuevo León. Datos Nuevo León, 2020. http://datos.nl.gob.mx/.

United States Census Bureau. U.S. and world population clock, 2020. https://www.census.gov/popclock/.

J. Watson and P. F. Whiting. Interpreting a covid-19 test result. *The BMJ*, May 2020. doi: 10.1136/bmj.m1808. URL https://www.bmj.com/content/bmj/369/bmj.m1808.full.pdf.

C. P. West, V. M. Montori, and P. Sampathkumar. Covid-19 testing. *Mayo Clinic Proceedings*, 95(6):1127–1129, Jun 2020. ISSN 00256196. doi: 10.1016/j.mayocp.2020.04.004.