False Negatives: Two Approaches

Riley Martin

Department of Mathematics and Statistics
Coastal Carolina University

1. Introduction

WHEN testing for infection of COVID-19, or other infectious diseases, it is necessary to consider the potential for errors in diagnosis (false positives/negatives). The potential consequences of inaccurate diagnoses are important to take into account, as those individuals may pose a danger to others within their communities. We will use two approaches to understanding these results using Bayesian methods.

Here is an explanation of some terminology that will be used:

- Sensitivity: Probability of correctly detecting the condition of those who actually have the condition (true positive probability, $P(T^+|D^+)$)
- Specificity: Probability of correctly giving a negative result to those who do not have the condition (true negative probability, $P(T^-|D^-)$)
- False negative: Probability of a negative test result, given that an individual has the condition. Denoted by: $P(D^+|T^-)$

2. Method One: LR⁻

IN order to calculate the probability of a false negative, we we will use the negative likelihood ratio (LR⁻) and Bayes' Rule. We define the negative likelihood ratio as:

$$\text{LR}^- = \frac{1-\text{sensitivity}}{\text{specificity}} = \frac{1-P(T^+|D^+)}{P(T^-|D^-)} = \frac{P(T^-|D^+)}{P(T^-|D^-)}$$

Using Bayes' Rule we can rewrite the sensitivity and specificity as follows:

$$1-\text{sensitivity} = P(T^-|D^+) = \frac{P(T^-)P(D^+|T^-)}{P(D^+)}$$

$$\text{specificity} = P(T^-|D^-) = \frac{P(T^-)P(D^-|T^-)}{P(D^-)}$$

Putting the previous results together, we get an equation that we can solve for the false negative probability:

$$\text{LR}^- = \frac{P(D^+|T^-)}{P(D^-|T^-)} \cdot \frac{P(D^-)}{P(D^+)} = \text{odds}_{D^+|T^-} \cdot \text{odds}_{D^-}$$

We are able to fix LR⁻ using known sensitivity and specificity values for COVID-19 PCR tests. Using prior knowledge of a patient's risk of exposure, we can input different values for the odds of being disease negative. Solving the equation will give us the odds of being disease positive given a negative test result, and thus the chance of a false negative.

3. Method Two: Bayesian Framework

USING Bayesian statistics, we are able to find a distribution of values for the false negative probability, known as the posterior distribution. The posterior distribution is based on a likelihood function and prior distribution. We will define our likelihood function based on data for n patients who have tested negative for COVID, with similar risks of exposure and symptoms. Therefore, $X \sim \text{Binomial}(n, p)$ where $p = P(D^+|T^-)$.

Since p is unknown, we will assume a distribution of possible values that is determined by the clinician based on prior information. The prior distribution for the probability is

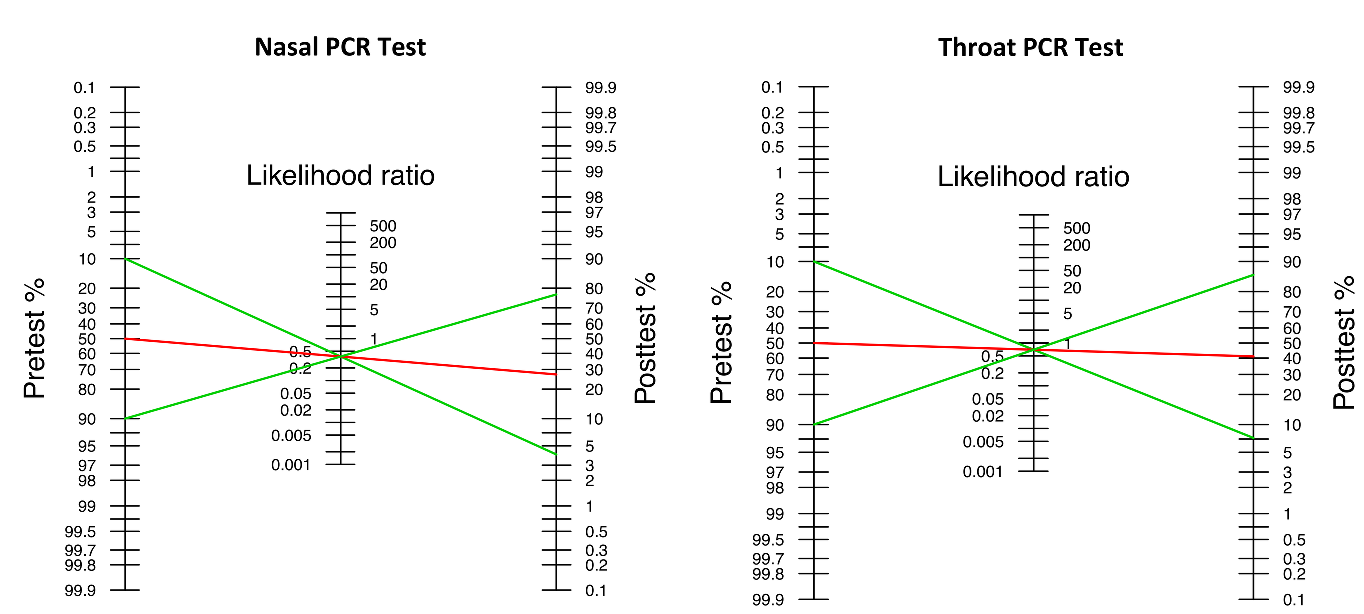
$$p \sim \text{Beta}(\alpha, \beta)$$

where alpha and beta are chosen based on the patient's situation and knowledge of the false negative rate in general. Based on these two distributions, we arrive at our posterior distribution:

$$\begin{aligned} P(p|X=x) &\propto p^x(1-p)^{n-x}\, p^{\alpha-1}(1-p)^{\beta-1} \\ &\propto p^{x+\alpha-1}(1-p)^{n-x+\beta-1} \end{aligned}$$

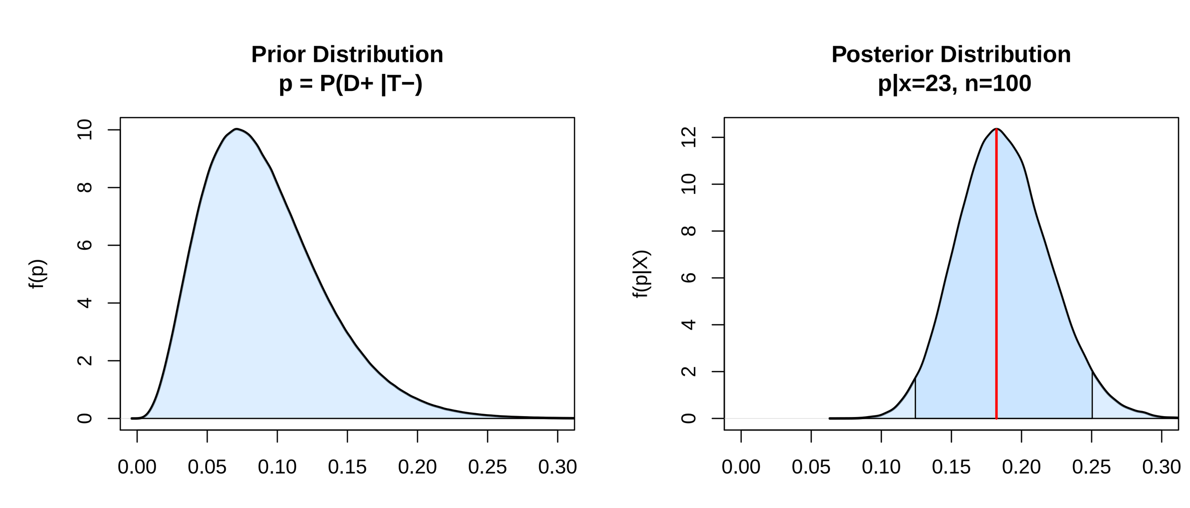
Thus, $p|X \sim \text{Beta}(x+\alpha, n-x+\beta)$. Using this distribution we can calculate the MAP estimator to get a point estimate of the probability of false negative. We can also calculate a high density credible interval (HDCI), to get a range of the most probable values.

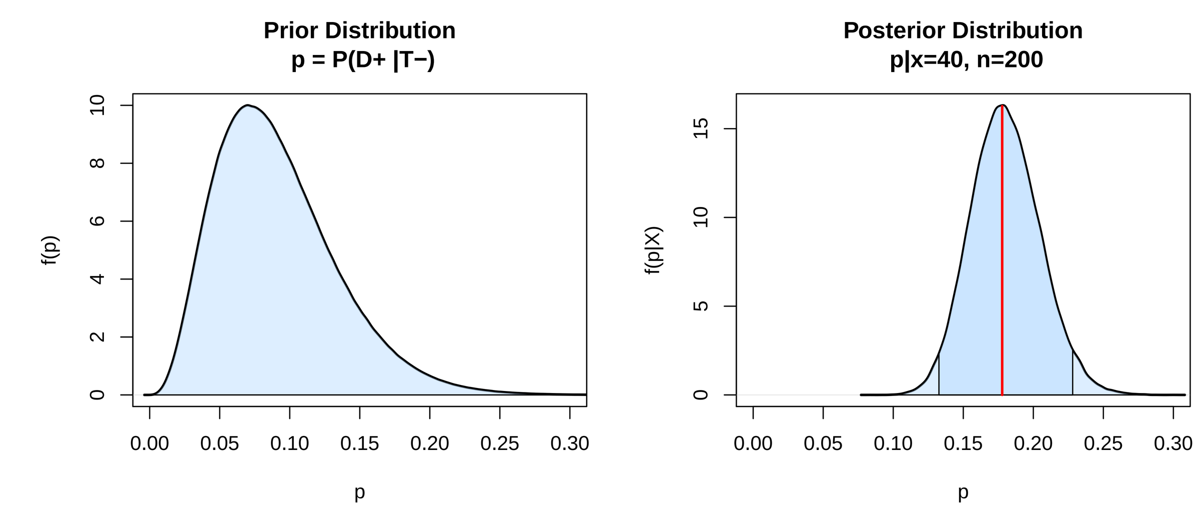
4. Results: Method One



For the nasal PCR test, we see that based on a pre-test probability of 10%, the post-test probability of a false negative is equal to 4%. For a pre-test probability of 90%, the probability of a false negative is 77%.

5. Results: Method Two





In both scenarios, the prior distribution used is Beta(4,40). This reflects the belief in low probabilities of false negatives.

6. Discussion

BASED on the first method, that utilized the negative likelihood ratio, we found that those individuals who have a lower pre-test probability of having COVID-19 (10%) and receive a negative PCR test result, will have a lower post-test probability (4%). This indicates that individuals who have been involved in lower risk activities and are not presenting any symptoms, can have increased confidence in a negative test result. However, as the pre-test probability increases, we see an increase in post-test probabilities as well. This method is limited in that it requires the assumption of fixed values for sensitivity and specificity, even when these are not certain.

To address this limitation, the use of Bayesian methods allow the clinician to incorporate distributions of values for both data and prior knowledge. In the first scenario, the data is from a sample of size n = 100, while the second scenario is from a sample of size n = 200. As, n increases, the likelihood function has greater influence on the posterior distribution, while the prior distribution becomes less important. This is evidenced by the mean for the posterior distribution: $\mu = \frac{x+\alpha}{\alpha+n+\beta}$. In the following table, we see MAP estimates and interval limits for the false negative probabilities.

Scenario	MAP Estimate	Lower HDI	Upper HDI
x=23 n=100	0.19	0.1241	0.2505
x=40 n=200	0.18	0.1325	0.2284

References

[1] Bolstad, W. M., & Curran, J. M. (2016). Introduction to Bayesian Statistics. Wiley.

[2] Chan G. M. (2020). Bayes' theorem, COVID19, and screening tests. The American journal of emergency medicine, 38(10), 2011–2013. https://doi.org/10.1016/j.ajem.2020.06.054

[3] R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

[4] Hans Landsheer (2021). UncertainInterval: Uncertain Interval Methods for Three-Way Cut-Point Determination in Test Results. R package version 0.7.0. https://CRAN.R-project.org/package=UncertainInterval

[5] Mike Meredith and John Kruschke (2020). HDInterval: Highest (Posterior) Density Intervals. R package version 0.2.2. https://CRAN.R-project.org/package=HDInterval

[6] Makowski, D., Ben-Shachar, M., & Lüdecke, D. (2019). bayestestR: Describing Effects and their Uncertainty, Existence and Significance within the Bayesian Framework. Journal of Open Source Software, 4(40), 1541. doi:10.21105/joss.01541