Bayes' Rule and Covid-19 Tests

January 28, 2021

The Economist published "Test match" in their Nov. 12, 2020, edition, p. 73. This was in the Science & technology section under the headline "Dealing with the pandemic."

Below is a chart duplicating one in the article. I easily matched their calculations in the accompanying Excel® spreadsheet. Change either or both accuracy measures to update the chart.
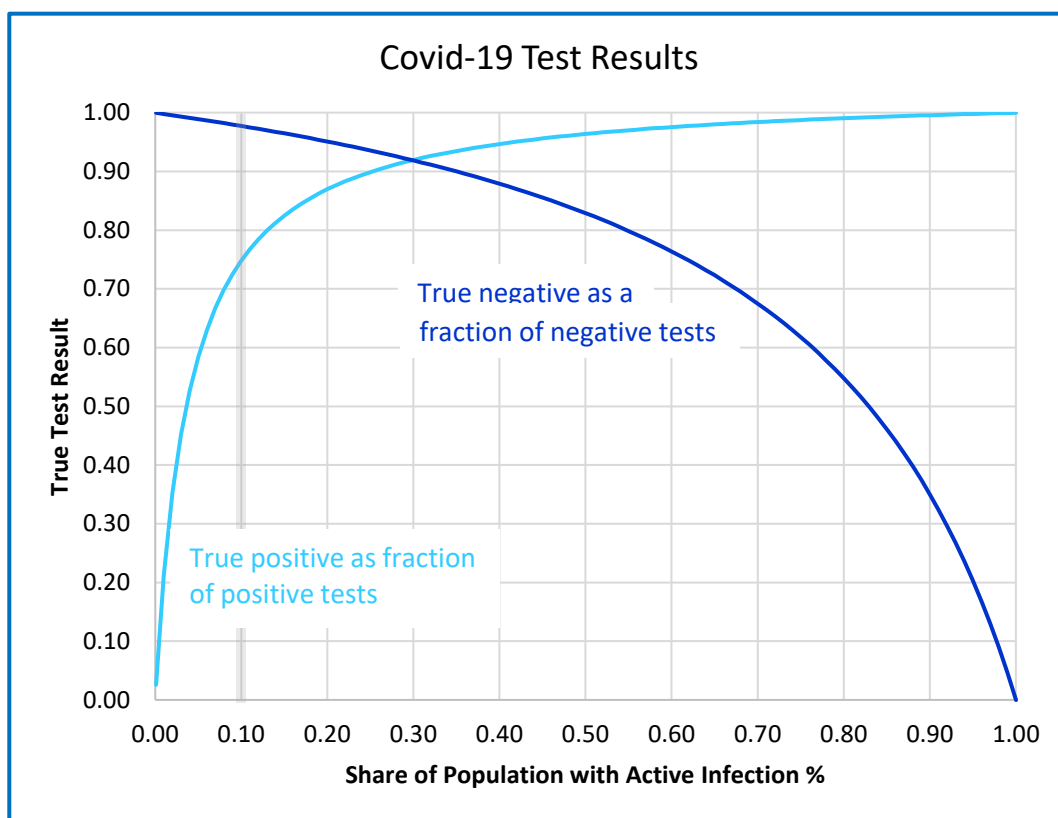

Figure. Curves are for a test with 80% sensitivity and 97% specificity. These represent the World Health Organization's guideline for minimum "acceptable" accuracy for rapid tests. With 10% of the population infected (gray line, a level chosen for illustration convenience): A "Positive" test indication has a 74% chance of being correct, and a "Negative" test indication has a 98% chance of being correct.

The key parameters are:

- **Sensitivity**: the ability of a test to correctly identify patients with a disease.
- **Specificity**: the ability of a test to correctly identify people without the disease.

Bayes' Rule

"Bayesian analysis" appears with increasing frequency. As of this writing, Google finds over 42M references to "Bayes" and about 73M for "Bayesian." Bayesian analysis is the basis for **value of information problems** (VOI), common in decision analysis.

Bayes' rule provides a way to revise a probability estimate based upon new, imperfect information. The Covid-19 pandemic offers opportunities to learn, practice, and apply the calculation.

Bayes' Formula

Here is a typical textbook expression for the Bayes' rule formula (though I've added quotation marks around *B*, explained soon):

$$P(A \mid "B") = \frac{P(A) \times P("B" \mid A)}{P("B")}$$

where *A* represents an outcome of the *event of interest* (something that affects outcome value), such as Product Demand.

"*B*" represents an outcome of an *information event* (a symptom of *A*), such as a Market Demand study.

The | (pipe character) is read "given that" in expressing a conditional probability.

The • (dot character) represents "and" in Boolean logic.

Because the *joint probability* numerator in the right-side is typically available, I prefer to remember and apply Bayes' formula in this simpler form:

$$P(A \mid "B") = \frac{P(A \bullet "B")}{P("B")}$$

Quotation Marks

I add quotation marks around *information event* outcomes to distinguish these from *event of interest* outcomes. This is especially helpful when using the same base symbols. As examples:

- In oil exploration, we might use:

  O = there actually is Oil, that will be found if drilled (event of interest outcome)

  "O" = "looks like Oil" from a geophysical survey (information event outcome)

- A manufacturer might forecast:

  o Market Demand as High, Medium, or Low (event of interest outcomes)

  o An information-gathering survey to improve the Market Demand forecast could classify survey results in a Market Interest parameter having categories "High," "Medium," and "Low" (information event outcomes).

In all cases, to be useful, there must be *correlation* between the information event(s) and the event of interest.

Organizing Judgments

Structures for organizing the parameters in Bayesian calculations include:

- Venn diagrams
- Probability trees (especially useful if there are multiple information events)
- Joint probability tables
- Multi-dimensioned "data clouds," from which we can obtain approximate expected values, distributions, and probabilities by partitioning.

Covid Tests

A frequent problem in using Bayes' rule is putting the numbers in the right places. Changing the Bayes' formula into one meaningful to the problem greatly reduces the chance of a calculation error.

For this problem, I replaced *A* with *I* (Infected actual) and "*B*" with "*I*" ("Infection indicated").

The Bayes' formula becomes:

$$P(I\,|\,"I") = \frac{P(I \bullet "I")}{P("I")}$$

After testing a person for the virus, what is the probability that the test result is correct? In the simple case— ignoring other symptoms and health conditions—there are four possibilities:

- Test is positive and the person is infected   True test result $P(I\bullet"I")$

- Test is positive but the persons is not infected   False test result $P(\text{not } I\bullet"I")$

- Test is negative but the person is infected   False test result $P(I\bullet"\text{not } I")$

- Test is negative and the person is not infected   True test result $P(\text{not } I\bullet"\text{not } I")$

The chart on the first page shows the first and last of these as functions of the assessed Share of Population with Active Infection.

Please see the accompanying Excel spreadsheet for the table used for the chart and check the example calculations. I hope that you will inspect and find the spreadsheet instructive. You may easily update the calculations and chart for different Sensitivity and Specificity.

∎