Bayes’ Rule – Application to Screening Tests for Disease

The definitive “gold standard” tests for detecting certain diseases are often invasive, lengthy, and/or expensive. For example, mammograms and colonoscopies are highly accurate, but are only cost-effective as preventive screening mechanisms for high-risk individuals, or those over a certain age. Pancreatic cancer, which is extremely aggressive, is often misdiagnosed as more common gastrointestinal conditions. In such cases, exploratory surgery is ultimately used as a last resort, only to reveal that the disease has already metastasized to other organs, and with a terminal prognosis, palliative care is the only option available. Other diseases, such as Alzheimer’s, and sudden death syndrome due to hypertrophic cardiomyopathy (enlarged heart), diagnostic tests do not exist at all, and the presence or absence of disease is determined by autopsy. The search for fast, easy, and inexpensive tests for such diseases, such as the presence of blood markers, is elusive and ongoing.

When such a screening test is developed, its merit is compared with the gold standard, for potential marketing to the general population. In this protocol, the disease status \((D^+\) or \(D^-\)) of a sample of patient volunteers is first determined via the gold standard. The new test is then administered, and typically renders either a positive \((T^+)\) or negative \((T^-)\) result, which may or may not agree with the gold standard.

- The sensitivity, or True Positive rate of the test, is defined as \(P(T^+ \mid D^+)\), the conditional probability that the test correctly shows positive, when the specific disease is present. However, errors can and do occur.
- The False Negative rate is defined as the complementary \(P(T^- \mid D^+)\), the conditional probability that the test incorrectly shows negative, when the specific disease is present. Hence \(P(T^- \mid D^+) = 1 - P(T^+ \mid D^+)\).
- Moreover, in this case, the test can be fooled into falsely reporting a positive result, in response to some other, perhaps benign, condition. For example, in a common over-the-counter test for colorectal cancer, a stool sample is mailed to a laboratory, where a pathologist checks it for the presence of blood cells. However, this may also occur if the patient has a ruptured hemorrhoid, rather than the specific disease colorectal cancer, for which the test was originally designed. This “false alarm” conditional probability \(P(T^+ \mid D^-)\) is the False Positive rate, i.e., a positive test result in the absence of the disease.
- The complement probability \(P(T^- \mid D^-) = 1 - P(T^+ \mid D^-)\), is the specificity, or True Negative rate of the test.

An ideal test has both high sensitivity and high specificity to the disease (i.e., low false positive and false negative rates). For example, a test that always reports a positive result, whether the disease is present or not, has 100% sensitivity, but 0% specificity.

These four conditional probabilities are not sufficient to determine a test’s efficacy for the general population, however. The purpose of such a test, is for an individual with unknown disease status to make a determination. Therefore, the posterior probabilities of main interest are the so-called predictive values \(PV^+ = P(D^+ \mid T^+)\) and \(PV^- = P(D^- \mid T^-)\) of a positive and negative test, respectively… both of which can be calculated via

Bayes’ Rule for two events: \(P(B \mid A) = \frac{P(A \mid B)P(B)}{P(A \mid B)P(B) + P(A \mid B^c)P(B^c)}\). In particular,

\[
PV^+ = P(D^+ \mid T^+) = \frac{P(T^+ \mid D^+)P(D^+)}{P(T^+ \mid D^+)P(D^+) + P(T^+ \mid D^-)P(D^-)}
\]

\[
PV^- = P(D^- \mid T^-) = \frac{P(T^- \mid D^-)P(D^-)}{P(T^- \mid D^-)P(D^-) + P(T^- \mid D^+)P(D^+)}.
\]
The prior probability $P(D^+)$ that appears in the numerator of $PV^+$ is called the prevalence of the disease in the population. (This is not to be confused with its incidence, the rate at which new cases develop.) If this prevalence is low, (i.e., the disease is relatively rare in the overall population), then its presence in the numerator of $PV^+$ tends to force it to be small. This is why physicians often use such a test as a way to rule out the presence of a disease if the test returns a negative result, and order a more sensitive test (i.e., one with a higher true positive rate) if it returns a positive result.

However, the prevalence is not well-approximated by the sample of patient volunteers from which the four conditional probabilities were derived, which typically yields a gross overestimate, and this is a problem. Accurate estimates of disease prevalence can be extremely difficult to obtain, and are usually a combination of published literature reviews, study data on federal websites, speculation by knowledgeable investigators conducting research in the field, etc. Over time, estimates of this prior probability may be updated as new data become available, resulting in newer posterior probability estimates. Note that this interpretation of “probability” is not consistent with the classical, frequentist definition of conducting repeated trials of a controlled experiment. For this reason, some investigators feel that, while this approach may be appropriate for certain applications, its subjective nature based on the strength of belief in the accuracy of the prior probability makes it inappropriate for other applications, especially in situations where policy recommendations are concerned. This has resulted in something of a clash between classical formalists and strict “Bayesian” practitioners. For most investigators however, Bayesian techniques are considered as tools, to be used in conjunction with other methods, with similar advantages and disadvantages. As seen here, in some situations, classical techniques are in fact inapplicable, leaving the Bayesian approach as the only alternative… as long as the results and interpretations are tempered with common sense.