FUN WITH BAYES

THOMAS S. SHORES

Some very elementary probability theory can expose surprising facts about testing outcomes. In this brief essay, I'll give a minimal, but (hopefully) mathematically complete, introduction to the probability theory that is needed to explaining these facts.

First, let's review some standard set notation. Loosely speaking, a *set* can be defined as a collection of objects, called the *members* of the set. This definition will suffice for us. Usually, sets will be designated by uppercase letters such as A, B, etc., and elements will be designated by lowercase letters such as a, b, etc. As usual, set A is a *subset* of set B if every element of A is an element of B, and a *proper* subset if it is a subset but not equal to B. Two sets A and B are said to be *equal* if they have exactly the same elements.

Some shorthand:

$\emptyset$ denotes the empty set, i.e., the set with no members.
$a \in A$ means "a is a member of the set A."
$A = B$ means "the set A is equal to the set B."
$A \subseteq B$ means "A is a subset of B."
$A \subset B$ means "A is a proper subset of B."

There are two ways in which we may define a set: we may *list* its elements, such as in the definition $A = \{0, 1, 2, 3\}$, or specify them by *rule* such as in the definition $A = \{x \mid x \text{ is an integer and } 0 \le x \le 3\}$. (Read this as "$A$ is the set of x such that x is an integer and $0 \le x \le 3$.") With this notation we can give formal definitions of set intersections and unions:

Definition. Let A and B be sets. Then the *intersection* of A and B is defined to be the set $A \cap B = \{x \mid x \in A \text{ and } x \in B\}$. The *union* of A and B is the set $A \cup B = \{x \mid x \in A \text{ or } x \in B\}$ (inclusive or, which means that $x \in A$ or $x \in B$ or both). The *difference* of A and B is the set $A - B = \{x \mid x \in A \text{ and } x \notin B\}$.

Date: May 5, 2019.

Example 1. Let $A = \{0, 1, 3, 5\}$ and $B = \{0, 1, 2, 4, 10\}$. Then

$$\begin{aligned} A \cup \emptyset &= A, \\ A \cap \emptyset &= \emptyset, \\ A \cup B &= \{0, 1, 2, 3, 4, 5, 10\}, \\ A \cap B &= \{0, 1\}, \\ A - B &= \{3, 5\}. \end{aligned}$$

□

Next, a very elememtary review of some basic probability concepts. We will only consider discrete sample spaces, i.e., spaces whose elements are countable (can be enumerated):

Definition. A *(discrete) sample space* is a countable set S, together with a probability measure P that assigns to any subset A of S a number $P(A)$ which satisfies the following conditions:

(1) $0 \leq P(A) \leq 1$.
(2) $P(S) = 1$ and $P(\emptyset) = 0$.
(3) If $A_1, A_2, A_2, \ldots$ is a sequence of pairwise disjoint subsets of S, i.e., $A_i \cap A_j = \emptyset$ if $i \neq j$, then

$$\begin{aligned} P(A_1 \cup A_2 \cup A_3 \cdots \cup A_j \cup \cdots) &= P(A_1) + P(A_2) + \ldots + P(A_n) + \ldots \\ &= \sum_{j=0}^{\infty} P(A_j). \end{aligned}$$

Definition. An *outcome* is any element of the sample space S and an *event* is any subset of outcomes of S.

We may assume that if a is any outcome in S, then $P(\{a\}) > 0$, for otherwise the presence or absence of a contributes no change in probability to any event.

There are a number of simple conclusions that can be inferred from the definitions. For example, let A and B be any events. Certainly $A \cap B$ and $A \cap \sim B$ are disjoint sets and $A \cup \sim A = S$, so property (3) implies that

$$P(A \cap B) + P(\sim A \cap B) = P((A \cap B) \cup (\sim A \cap B)) = P((A \cup \sim A) \cap B) = P(B).$$

Definition. Let A and B be events in the sample space S with $P(B) > 0$. Then

(1) The *complement* of event A is $\sim A = \{x \in S \mid x \notin A\}$.
(2) The *conditional probability* of event A, given that event B has occurred, is

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}.$$

(3) Events A and B are *independent* if $P(A \cap B) = P(A)\,P(B)$.

The intuitive meaning of conditional probability is reasonably clear: Given that an event B has occurred, then the probability of event B occurring should be 1 and the only other possible events are subsets of B. The definition of conditional probability guarantees both. Moreover, it offers an explanation of the meaning of independent events: If A and $B \neq \emptyset$ are independent, then we have from definitions of independence and conditional probability that

$$P(A)\,P(B) = P(A \cap B) = P(A\,|B)\,P(B)\,.$$

Now cancel $P(B)$ from both sides to obtain that $P(A\,|B) = P(A)$, i.e., the probability of event A occuring is independent of whether or not B has occurred.

Example 2. Let $S = \{1,2,3,4,5,6,7,8,9,10\}$, $A = \{2,3,5,7\}$, $B = \{2,4,6,8,10\}$ and assume that each outcome is equally likely. Think of the event A as that of selecting a prime number from the sample space S, event B as that of selecting an even number from S and event $A \cap B = \{2\}$ as that of selecting a prime number that is even. Individual outcomes are equally likely and each set is a disjoint union of subsets consisting of single outcomes, so we see from (3) of the definition of sample space that $P(A) = 4/10 = 2/5$, $P(B) = 5/10 = 1/2$ and $P(A \cap B) = 1/10$. From the definition of conditional probability, we see that $P(A \mid B) = P(A \cap B)\,/P(B) = (1/10)\,/\,(1/2) = 1/5$. And of course, this makes perfect sense since, given that you have selected even numbers from S, of which there are 5, the probability of selecting a prime is $1/5$, for exactly one of those outcomes in B will be prime. Note that in this case $P(A)\,P(B) = (2/5)\,(1/2) = 1/5 \neq 1/10$, so these events are not independent. Here's an exercise for you: Perform all of the above calculations with $\sim B$ in place of B. □

Now we're ready for an important medical application of the above elementary concepts, but first we need some terminology: Let S be a (finite) population of people (or animals, for testing) and think of each outcome as the event of selecting at random an individual from that population. Assume that each outcome is equally likely. Suppose that we are testing for the presence of a specific disease in individuals from the population, by way of a specific test T which returns a positive or negative indication for the disease. Consider the following events:

- $A = \{x \in S \mid x \text{ has the disease}\}$
- $B = \{x \in S \mid x \text{ tests positive for the disease}\}$

Next, suppose that data collected from various sources provides us with reasonable estimates of the following quantities:

- The *sensitivity* $P(B \mid A)$ of test T, which is the probability that test T returns a positive indication, given that the tested individual has the disease.
- The *specificity* $P(\sim B \mid \sim A)$ of test T, which is the probability that test T returns a negative indication, given that the tested individual does not have the disease.
- The *likelihood* $P(A)$ that an individual randomly selected from the population has the disease.

The question we want to answer is basically this: How good is the test? In other words, given that an individual tests postive with test T, how likely is it that the individual actually has the disease or doesn't have the disease? Specifically, the numbers we are after are $P(A \mid B)$ (likelihood of a true positive) and $P(\sim A \mid B)$ (likelihood of a false positive) , respectively. It suffices to get one of these numbers since a simple application of part (3) of the definition of sample space, which we leave as an exercise, shows that $P(A \mid B) + P(\sim A \mid B) = 1$. Naturally, we want the first number to be close to 1 and the second to be close to 0.

This is where things get a little surprising in some cases. So how do we go from event $B \mid A$ to event $A \mid B$? We'll assume that events A and B can actually happen, i.e., $P(A) > 0$ and $P(B) > 0$. Note that we can calculate $P(A \cap B)$ in two different ways using the definition of conditional probability, namely,

$$P(B)\, P(A \mid B) = P(A \cap B) = P(A)\, P(B \mid A)\,.$$

Solve for $P(A \mid B)$ and use the fact that $P(B \mid \sim A) + P(\sim B \mid \sim A) = 1$ (see above) to obtain

$$\begin{aligned}
P(A \mid B) &= \frac{P(A)\, P(B \mid A)}{P(B)} \\
&= \frac{P(A)\, P(B \mid A)}{P(B \cap A) + P(B \cap \sim A)} \\
&= \frac{P(A)\, P(B \mid A)}{P(A)\, P(B \mid A) + P(\sim A)\, P(B \mid \sim A)} \\
&= \frac{P(A)\, P(B \mid A)}{P(A)\, P(B \mid A) + P(\sim A)\,(1 - P(\sim B \mid \sim A))}.
\end{aligned}$$

This is our key formula. Just for the record, the first equality is important in its own right:

Theorem. *(Bayes' Theorem) Given events A and B from the sample space S,*

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

Now, let's put the formula to work on an actual example.

Example 3. The following estimates are drawn from various sources of data. It is estimated that for women in the United States under the age of 40, the mammogram test for breast cancer has a sensitivity of 71.3% and a specificity of 85.9%. It is also estimated that breast cancer for these women occurs in approximately 115 out of 100,000 individuals. Now suppose that you are a woman under the age of 40 and your mammogram tests positive for cancer. How concerned should you be?

In this example the population S is the set of all American women under the age of 40. The event A is the case of an individual from S having breast cancer and B is the event that the mammogram tests positive for breast cancer. Thus, we are given that the sensitivity of the test is $P(B \mid A) = 0.713$, specificity is $P(\sim B \mid \sim A) = 0.859$ and the likelihood of a randomly selected woman under the age of 40 having breast cancer is $P(A) = 115/100,000 = 0.00115$. Plug these numbers into the equation above to obtain that the likelihood that a positive test result correctly predicts breast cancer is

$$\begin{aligned} P(A \mid B) &= \frac{P(A)\,P(B \mid A)}{P(A)\,P(B \mid A) + P(\sim A)\,(1 - P(\sim B \mid \sim A))}. \\ &= \frac{0.00115 \cdot 0.713}{0.00115 \cdot 0.713 + 0.99885\,(1 - 0.859)} \\ &\approx 0.005788. \end{aligned}$$

In other words, that individual has a 0.5788% likelihood of having cancer. □

This is a rather surprising result – not to say that there should be no concern about the outcome of this test, but the possibility of a false positive is very high, namely approximately 99.4%.

Here's one more surprising example. It's taken from Daniel Kahneman's book *Thinking, fast and slow* [1]. I've varied the numbers just a bit.

Example 4. A hit and run occurred at night. Conditions were not ideal, but a witness identified the offending vehicle, which happened to be a cab, as a blue colored cab. This information was used to identify a suspect. It so happens that in this city, there is only one cab company whose cabs are blue, and they comprise about 10% of the cabs in the city. To ensure reliability, the prosecutors subjected this witness to a battery of tests with random images in comparable conditions, and they found that she was 80% accurate in correctly identifying blue vehicles as blue and 85% accurate in identifying non-blue vehicles as not blue. Thus, they argued that her testimony was very reliable. What do you think?

We can cast this example like the medical test of earlier discussion as follows: Let A be the event that a randomly selected cab is blue and B the event that the witness positively identifies a randomly selected cab as blue. Witness identification is the "test". What we are given is that the sensitivity of the test is $P(B \mid A) = 0.8$, the specificity of the test is $P(\sim B \mid \sim A) = 0.85$ and the likelihood of a cab being blue is $P(A) = 0.1$. Apply the test of the previous example to obtain that the likelihood that a cab that the witness identified as blue was actually blue is

$$\begin{aligned} P(A \mid B) &= \frac{P(A)\,P(B \mid A)}{P(A)\,P(B \mid A) + P(\sim A)\,(1 - P(\sim B \mid \sim A))}. \\ &= \frac{0.1 \cdot 0.8}{0.1 \cdot 0.8 + 0.9\,(1 - 0.85)} \\ &\approx 0.37209. \end{aligned}$$

In other words, that individual has at best a 38% likelihood of correctly identifying a blue cab. Looks like we could have a problem if our case depended heavily on this witness. □

The problem in both of these examples is that affected groups (diseased, blue) are substantially outnumbered by the unaffected groups, so the second term (false positives) dominates the denominator.

References

[1] Kahneman, Daniel, *Thinking, fast and slow*, London: Penguin Books (2011).