Applying Bayes' Theorem

David R. Duncan & Bonnie H. Litwiller, University of Northern Iowa <David.Duncan@uni.edu>

__David Duncan__ and __Bonnie Litwiller__ are both recently retired from long careers as Professors of Mathematics at the University of Northern Iowa. They taught numerous courses for both in-service and pre-service mathematics teachers, have spoken widely at NCTM and affiliate group meetings, and have published over 950 articles for mathematics teachers of varying grade levels.

Teachers are always looking for ways in which mathematical procedures and theorems can be applied to real-world situations. We shall describe one such situation, involving medicine and Bayes' Theorem.

Bayes' Theorem is used to estimate the likelihood of an unobserved event, given an observed consequence. It is developed as follows:

If E and F are two events,

P(E and F) can be expressed in either of two ways.

1. $P(EandF) = P(E) \cdot P(F|E)$, where $P(F|E)$ denotes the probability of F, given that E has occurred.
2. $P(EandF) = P(F) \cdot P(E|F)$

Then, $P(E) \cdot P(F|E) = P(F) \cdot P(E|F)$

Solving for $P(E|F)$ yields Theorem I:

$$P(E|F) = \frac{P(E) \cdot P(F|E)}{P(F)}$$

Theorem I is the essence of Bayes' Theorem.

To exemplify Theorem I, suppose that Basket 1 contains 5 red balls and 3 green balls and Basket 2 contains 4 red balls and 7 green balls. Suppose that a basket is selected randomly and then a ball is selected randomly from that basket. If the selected ball is observed to be red, what is the probability that it came from Basket 1?

Let: B1 be the event that Basket 1 is selected;
B2 be the event that Basket 2 is selected;
R be the event that a red ball is selected;
G be the event that a green ball is selected.

Then: P(B1) = $\frac{1}{2}$

P(B2) = $\frac{1}{2}$

P($(R|B1)$) = $\frac{5}{8}$

P($(R|B2)$) = $\frac{4}{11}$

P(R) = P(B1)P $(R|B1)$ +P(B2)P($R|B2$)

$$= \frac{1}{2}\left(\frac{5}{8}\right)+\frac{1}{2}\left(\frac{4}{11}\right)=\frac{5}{16}+\frac{2}{11}=\frac{87}{176}$$

By Bayes' Theorem,

$$P(B1|R) = \frac{P(B1)P(R|B1)}{P(R)} = \frac{\frac{5}{16}}{\frac{87}{176}} = \frac{5(11)}{87} = \frac{55}{87}$$

In practice, the denominator of Theorem I is often further refined. Let E_1, E_2, E_3,…E_n be a partition on the sample space, meaning that they are pairwise disjoint and that their union is the entire sample space. Then,

$P(F)=P(E_1)\cdot P(F|E_1)+P(E_2)\cdot(F|E_2)+\ldots+P(E_n)\cdot P(F|E_n)$. With this refinement, Theorem I can be rewritten as Theorem II:

$$P(E_i|F)=\frac{P(E_i)\cdot P(F|E_i)}{P(E_1)\cdot P(F|E_1)+\ldots+P(E_n)\cdot P(F|E_n)} \text{ for any } i \ \varepsilon\ 1,2,3,\ldots n$$

Let us now apply this Theorem II to a real-world situation. Suppose that a group of persons are being screened for the presence of a potentially dangerous virus. A preliminary test applied to a specific person suggests that she is carrying this virus. What is the probability that in fact she does carry this virus?

To analyze this situation, a set of symbols must be designated and a set of empirical observations must be stipulated.

For a random person, let V = the event that the person actually carries the virus, and (~V) = the event that the person does not carry the virus. Also let T = the event that the preliminary test suggests the presence of the virus, and (~T) = the event that the test does not suggest this presence.

Suppose it is known that:
P(V)=3% (3% of the persons carry the virus).
$P(T|V)$=90% (of those with the virus, 90% are correctly identified by the preliminary test.
$P(T|\sim V)$=5% (For those persons without the virus, 5% are incorrectly identified by the preliminary test as being carriers; this would be called a "false positive.")

With these assumptions, the probability that a preliminarily identified virus carrier actually carries the virus can be calculated:

$$P(V|T)=\frac{P(V)\cdot P(T|V)}{P(T)}$$

$$=\frac{P(V)\cdot P(T|V)}{P(V)\cdot P(T|V)+P(\sim V)\cdot P(T|\sim V)}$$

$$=\frac{(0.03)(0.9)}{(0.03)(0.9)+(0.97)(0.05)}$$

$$= \frac{0.027}{0.027 + 0.0485} = \frac{0.027}{0.0755} = 0.358 \text{ or } 35.8\%$$

In other words, a preliminary positive test is far from conclusive, in spite of 90% test accuracy for actual carriers and 95% test accuracy for actual non-carriers. The public health implications are obvious. The preliminary test alarms almost three times as many people as actually carry the virus. This could both needlessly distress the "false positives" and cause considerable additional costs to the health care establishment as further tests are conducted.

Another way to see what is happening in this example is through the following numerical data and resulting table. Suppose that 2000 individuals are randomly selected and tested for this virus. We would expect that 3% of them (i.e. 60 individuals) are carriers of the virus and the remaining1940 individuals are not carriers. Of the 60 carriers, we expect that 90% of them (i.e. 54 individuals) are correctly identified by the test as carriers, while the remaining 6 individuals are mis-identified as non-carriers. Similarly, of 1940 individuals who are not carriers, we expect that 95% of them (i.e. 1843 individuals) are correctly identified as non-carriers, while the remaining 97 are mis-identified as carriers.

	Test Positive	Test Negative	
Carriers	54	6	60
Non-Carriers	97	1843	1940
	151	1849	2000

Only $\frac{54}{151}$ or 35.8% of those who test positive actually are carriers.

The reader and students are encouraged to re-calculate $P(V|T)$ for other numerical assumptions, and to find other settings in which Bayes' Theorem can be applied.