Probability and statistics have been areas of strong academic and intellectual interest for me, and I have developed a well-rounded understanding of their core concepts and applications.

My learning and exploration span fundamental topics such as probability distributions and probability density functions (PDFs), random variables, expectation and variance, conditional probability, and statistical inference techniques.

I am particularly interested in how these concepts work together to model uncertainty, analyze data, and support logical decision-making in real-world scenarios.

Among the many areas within this field, one of my favorites and most intellectually engaging topics is Bayesian reasoning, especially Bayes' theorem, which provides a powerful framework for updating beliefs using new evidence.

To demonstrate my interest and understanding, I will present a few examples centered around Bayes' theorem.

The probability of event *A*, given that event *B* has subsequently occurred, is

$$P(A|B) = \frac{P(A) \cdot P(B|A)}{[P(A) \cdot P(B|A)] + [P(\bar{A}) \cdot P(B|\bar{A})]}$$

Example

In a XYZ Country, 51% of the adults are males. One adult is randomly selected for a survey involving credit card usage.

a. Find the prior probability that the selected person is a male.

b. It is later learned that the selected survey subject was smoking a cigar. Also, 9.5% of males smoke cigars, whereas 1.7% of females smoke cigars (based on data from the Substance Abuse and Mental Health Services Administration). Use this additional information to find the probability that the selected subject is a male.

Solution

Let's use the following notation:

M = male

F = female (or not male)

C = cigar smoker

NC = not a cigar smoker.

a. Before using the information given in part b, we know only that 51% of the adults in XYZ County are males, so the probability of randomly selecting an adult and getting a male is given by *P*(M) = 0.51.

b. Based on the additional given information, we have the following:

P (M) = 0.51	because 51% of the adults are males
P (F) = 0.49	because 49% of the adults are females (not males)
P (C|M) = 0.095	because 9.5% of the males smoke cigars (That is, the probability of getting someone who smokes cigars, given that the person is a male, is 0.095.)
P (C| F) = 0.017	because 1.7% of the females smoke cigars (That is, the probability of getting someone who smokes cigars, given that the person is a female, is 0.017.)

Let's now apply Bayes' theorem by using the preceding formula with M in place of A, and C in place of B. We get the following result:

$$P(M \mid C) = \frac{P(M)\cdot P(C|M)}{[P(M)\cdot P(C|M)]+[P(F)\cdot P(C|F)]}$$

$$= \frac{0.51\cdot 0.095}{[0.51\cdot 0.095]+[0.49\cdot 0.017]}$$

$$= 0.85329341$$

= 0.853 (rounded)

Conclusion:

Before we knew that the survey subject smoked a cigar, there is a 0.51 probability that the survey subject is male (because 51% of the adults in XYZ County are males). However, after learning that the subject smoked a cigar, we revised the probability to 0.853. There is a 0.853 probability that the cigar smoking respondent is a male. This makes sense, because the likelihood of a male increases dramatically with the additional information that the subject smokes cigars (because so many more males smoke cigars than females).

Intuitive Bayes Theorem

The preceding solution illustrates the application of Bayes' theorem with its calculation using the formula. Unfortunately, that calculation is complicated enough to create an abundance of opportunities for errors and/or incorrect substitution of the involved probability values. Fortunately, here is another approach that is much more intuitive and easier:

Assuming some convenient value for the total of all items involved, then constructing a table of rows and columns with the individual cell frequencies based on the known probabilities.

For the preceding example, simply assuming some value for the adult population of XYZ County, such as 100,000, then using the given information to construct a table, such as the one shown below.

Finding the number of males who smoke cigars: If 51% of the 100,000 adults are males, then there are 51,000 males. If 9.5% of the males smoke cigars, then the number of cigar smoking males is 9.5% of 51,000, or $0.095 \times 51{,}000 = 4845$. See the entry of 4845 in the table. The other males who do *not* smoke cigars must be $51{,}000 - 4845 = 46{,}155$. See the value of 46,155 in the table.

Finding the number of females who smoke cigars: Using similar reasoning, 49% of the 100,000 adults are females, so the number of females is 49,000. Given that 1.7% of the females smoke cigars, the number of cigar smoking females is $0.017 \times 49{,}000 = 833$. The number of females who do *not* smoke cigars is $49{,}000 - 833 = 48{,}167$. See the entries of 833 and 48,167 in the table.

	C (Cigar Smoker)	NC (Not a Cigar Smoker)	**Total**
M (male)	4845	46,155	**51,000**
F (female)	833	48,167	**49,000**
Total	**5678**	**94,322**	**100,000**

The above table involves relatively simple arithmetic. Simply partitioning the assumed population into the different cell categories by finding suitable percentages.

Now we can easily address the key question as follows: To find the probability of getting a male subject, given that the subject smokes cigars, we simply use the conditional probability. To find the probability of getting a male given that the subject smokes, we restrict the table to the column of cigar smokers, then find the probability of getting a male in that column. Among the 5678 cigar smokers, there are 4845 males, so the probability we seek is 4845/5678 = 0.85329341. That is, P (M | C) = 4845/5678 = 0.85329341 = 0.853 (rounded).