MthStat 465, Spring 2005, Lecture Number 10
Bayes' Theorem. Random Variables.

Bayes' Theorem allows us to relate $\Pr(A|B)$ to $\Pr(B|A)$. Random variables are functions whose domain is the sample space, but not all functions on the sample space are random variables.

1. Bayes' Theorem

Suppose that $0 < \Pr(A) < 1$ and $\Pr(B) > 0$. Then

$$\begin{aligned}\Pr(A|B) &= \frac{\Pr(A \cap B)}{\Pr(B)} \\ &= \frac{\Pr(B|A)\Pr(A)}{\Pr(B)} \\ &= \frac{\Pr(B|A)\Pr(A)}{\Pr(B \cap A) + \Pr(B \cap A^c)} \\ &= \frac{\Pr(B|A)\Pr(A)}{\Pr(B|A)\Pr(A) + \Pr(B|A^c)\Pr(A^c)}\end{aligned}$$

Therefore,

$$\Pr(A|B) = \frac{\Pr(B|A)\Pr(A)}{\Pr(B|A)\Pr(A) + \Pr(B|A^c)\Pr(A^c)}. \tag{1}$$

Equation (1) is called **Bayes' Formula**. See pages 46 to 50 of *The Cartoon Guide to Statistics* for an application to medical testing.

There are two key ideas in Bayes' Formula. One is that $\Pr(A\cap B) = \Pr(B|A)\Pr(A)$, which is called **conditioning**, and the other is that $\Pr(B) = \Pr(B\cap A)+\Pr(B\cap A^c)$, which is called **partitioning**.

2. Random Variables

Suppose we have a sample space S and a sigma algebra Σ of subsets of S. A function R with domain S and range contained in the real numbers is called a **random variable** if for each interval of the form $(a, b]$ the set $\{s \in S : a < R(s) \leq b\}$ is an element of Σ, that is, $\{s \in S : a < R(s) \leq b\}$ is an event. The set $\{s \in S : a < R(s) \leq b\}$ is usually abbreviated $\{R \in (a, b]\}$.

It is important to be aware that not every function with domain S and range in the real numbers may be a random variable, as there is a dependence on the sigma algebra, Σ. However, if Σ consists of all subsets of S, then every real valued function on S is a random variable. If S is a countable set, we can always arrange to have Σ be the set of all subsets of S.

Example:

$$\begin{aligned}S &= \{1, 2, 3, 4, 5\}, \\ \Sigma &= \{S, \emptyset, \{1, 2\}, \{3, 4, 5\}\}\end{aligned}$$

Define $R_1(s) = s^2$. R_1 is not a random variable, since $\{s : 20 < R(s) \leq 25\} = \{5\}$, and this set is not an event.

Define $R_2(s) = 1$ if $s = 1$ or $s = 2$, and $R_2(s) = 0$ if $s = 3$, $s = 4$ or $s = 5$. R_2 is a random variable. There are 4 cases to check. Let I be any interval of the form $(a, b]$.

Case 1:: Neither 0 nor 1 is in I. Then $\{s \in S : R(s) \in I\} = \emptyset$.

Case 2:: 0 is not in I, but 1 is in I. Then $\{s \in S : R(s) \in I\} = \{1, 2\}$.
Case 3:: 1 is not in I, but 0 is in I. Then $\{s \in S : R(s) \in I\} = \{3, 4, 5\}$.
Case 4:: Both 0 and 1 are in I. Then $\{s \in S : R(s) \in I\} = S$.

To each random variable R we may assign two functions whose domain is the real numbers and whose range is contained in $[0, 1]$. The first is the **probability mass function** of R, denoted p_R. For each real number t we define $p_R(t) = \Pr(\{s \in S : R(s) = t\})$. Note that if t is not in the range of R then $p_R(t) = 0$.

The second function is the **probability distribution function** of R, denoted by $F_R(t)$. For each real number t we define $F_R(t) = \Pr(\{s \in S : R(s) \leq t\})$. Note that $F_R(t) = 0$ if t is smaller than any range value of R and $F_R(t) = 1$ if t is greater than any range value of R.