Section 2.3. Baye's Theorem

Note. Baye's Theorem involves a partition of the sample space $S = \cup_{j=1}^{k} B_j$ and an event A. It relates $\Pr(B_i|A)$ to $\Pr(A|B_j)$ and $\Pr(B_j)$ for $j = 1, 2, \ldots, k$. This was illustrated in an example in Section 2.1, "The Definition of Conditional Probability" (the example based on the Harvey Mudd College Math Fun Facts example about false positives).

Example 2.3.2. In Examples 2.1.9 and 2.1.11 we considered a box of bolts. We had B_1 as the event that the box contains 60 long and 40 short bolts, B_2 as the event that the box contains 10 long and 20 short bolts, and A as the event that a randomly chosen bolt is long. We took $\Pr(B_1) = \Pr(B_2) = 1/2$, $\Pr(A|B_1) = 3/5$, $\Pr)A|B_2) = 1/3$, and computed $\Pr(A) = 2/15$. If we want to compute $\Pr(B_2|A)$ then we have by Definition 2.1.1, "conditional probability," that

$$\Pr(B_1|A) = \frac{\Pr(A \cap B_1)}{\Pr(A)} = \frac{\Pr(B_1)\Pr(A|B_1)}{\Pr(A)} = \frac{(1/2)(3/5)}{7/15} = \frac{9}{14}.$$

Similarly, $\Pr(B_2|A) = 5/14$. So if a bolt is chosen at random and it is a long bolt then this new information changes the probability of B_1 and B_2; notice that B_2 then becomes more likely as is reasonable since it implies the box probably has more long bolts.

Theorem 2.3.1 Baye's Theorem. Let the events $B_1, B_2, \ldots, B_k$ form a partition of sample space S such that $\Pr(B_j) \neq 0$ for $j \in \{1, 2, \ldots, k\}$ and let A be an event such that $\Pr(A) \neq 0$. Then for $i \in \{1, 2, \ldots, k\}$ we have

$$\Pr(B_i|A) = \frac{\Pr(B_i)\Pr(A|B_i)}{\sum_{j=1}^{k} \Pr(B_j)\Pr(A|B_j)}.$$

Proof. By Definition 2.1.1, "conditional probability,"

$$\Pr(B_i|A) = \frac{\Pr(B_i \cap A)}{\Pr(A)}.$$

By Theorem 2.1.1, $\Pr(B_i \cap A) = \Pr(B_i)\Pr(A|B_i)$. By Theorem 2.1.4, $\Pr(A) = \sum_{j=1}^{k} \Pr(B_j)\Pr(A|B_j)$, hence

$$\Pr(B_i|A) = \frac{\Pr(B_i)\Pr(A|B_i)}{\sum_{j=1}^{k} \Pr(B_j)\Pr(A|B_j)},$$

as claimed. ∎

Example 2.3.1 and 2.3.3. Suppose a medical test is 90% reliable in the following sense: If a person has the disease, there is a probability of 0.9 that the test will give a positive response whereas if a person does not have the disease there is a probability of only 0.1 that the test will give a positive response. Data indicate that your chances of having the disease (given that you are randomly chosen from the population) is 1 in 10,000. The Department of Public Health is giving the tests for free and you decide to take the test. If you receive a positive result from the text, what is the probability that you actually have the disease? Let B_1 denote the event that you have the disease, let B_2 denote the event that you do not have the disease (so that B_1 and B_2 partition the sample space). and let A be the event that you test positive. We have $\Pr(B_1) = 1/10,000 = 0.0001$, $\Pr(B_2) = 0.9999$, $\Pr(A|B_1) = 0.9$, and $\Pr(A|B_2) = 0.1$. The desired probability is $\Pr(B_1|A)$. By Baye's Theorem we have

$$\Pr(B_1|A) = \frac{\Pr(A|B_1)\Pr(B_1)}{\Pr(A|B_1)\Pr(B_1) + \Pr(A|B_2)\Pr(B_2)}$$

$$= \frac{(0.9)(0.0001)}{(0.9)(0.0001) + (0.1)(0.9999)} \approx 0.00090. \quad \square$$

Example 2.3.5. Identifying Genotypes. Consider a gene that has two alleles, A and a. Suppose the gene exhibits itself in two phenotypes: genotypes AA and Aa display the "dominant trait" and genotype aa displays the "recessive trait." Suppose the genotypes AA, Aa, and aa occur in a population with probabilities (or "frequencies") of 1/4, 1/2, and 1/4, respectively. We select an individual at random and let E be the event that the individual has the dominant trat (of course, this is just $1/4 + 1/2 = 3/4$). We introduce the possible genotypes of the parents, denoted $B_1, B_2, \ldots, B_6$ as given below, and then compute $\Pr(E|B_i)$ for $i = 1, 2, \ldots, 6$. We have:

	(AA, AA)	(AA, Aa)	(AA, aa)	(Aa, Aa)	(Aa, aa)	(aa, aa)
Parental genotype	B_1	B_2	B_3	B_4	B_5	B_6
Probability of B_i	1/16	1/4	1/8	1/4	1/4	1/16
$\Pr(E\|B_i)$	1	1	1	3/4	1/2	0

By Baye's Theorem, $\Pr(B_i|E) = \dfrac{\Pr(B_i)\Pr(E|B_i)}{\sum_{j=1}^6 \Pr(B_j)\Pr(E|B_j)}$. Now

$$\sum_{j=1}^{6} \Pr(B_j)\Pr(E|B_j) = \frac{1}{16} + \frac{1}{4} + \frac{1}{8} + \frac{1}{4}\frac{3}{4} + \frac{1}{4}\frac{1}{2} + 0 = \frac{12}{16} = \frac{3}{4},$$

so $\dfrac{Pr}{(}B_1|E) = \dfrac{(1/16)(1)}{3/4} = \dfrac{1}{12}$, $\Pr(B_2|E) = \dfrac{(1/4)(1)}{3/4} = \dfrac{1}{3}$, $\Pr(B_3|E) = \dfrac{(1/8)(1)}{3/4} = \dfrac{1}{6}$, $\Pr(B_4|E) = \dfrac{(1/4)(3/4)}{3/4} = \dfrac{1}{4}$, $\Pr(B_5|E) = \dfrac{(1/4)(1/2)}{3/4} = \dfrac{1}{6}$, and $\Pr(B_6|E) = 0$.

Note. The "conditional version of Baye's Theorem" states that for events B_1, B_2, ..., B_k which form a partition of sample space S such that $\Pr(B_j) \neq 0$ for $j \in \{1, 2, \ldots, k\}$ and for events A and C where $\Pr(C) \neq 0$ and $\Pr(B_j \cap C) \neq 0$ for $j \in \{1, 2, \ldots, k\}$ we have

$$\Pr(B_i|A \cap C) = \frac{\Pr(B_i|C)\Pr(B|B_i \cap C)}{\sum_{j=1}^k \Pr(B_j|C)\Pr(A|B_j \cap C)}.$$

Definition/Note. In the above examples, a probability like $\Pr(B_i)$ is called the *prior probability* because it is given before additional information is available. A probability like $\Pr(B_i|A)$ or $\Pr(B_i|E)$ is called a *posterior probability* because it is computed based on the additional given event.

Example. Computation of Posterior Probabilities is More Than One Stage. Suppose a box contains one fair coin and one coin with a head on each side. A coin is selected at random from the box, it is tossed and a head is obtained. We now want to know the probabilities that the coin is the fair coin. Let B_1 be the event that the coin is fair, let B_2 be the event that the coin has two heads, and let H_1 be the event that a head is obtained when the coin is tossed. We have $\Pr(B_1) = \Pr(B_2) = 1/2$, $\Pr(H_1|B_1) = 1/2$, $\Pr(H_1|B_2) = 1$, and we want to find $\Pr(B_1|H_1)$. By Baye's Theorem

$$\Pr(B_1|H_1) = \frac{\Pr(B_1)\Pr(H_1|B_1)}{\Pr(B_1)\Pr(H_1|B_1) + \Pr(B_1)\Pr(H_1|B_2)} = \frac{(1/2)(1/2)}{(1/2)(1/2) + (1/2)(1)} = \frac{1}{3}.$$

So the prior probability is $\Pr(B_1) = 1/2$ and the posterior probability after the first toss is $\Pr(B_1|H_1) = 1/3$. Now suppose the coin is tossed again and a head is obtained again (the event which we denote as H_2). We want to compute the new ("second stage") posterior probability $\Pr(B_1|H_1 \cap H_2)$. We assume conditional independence given B_1 so that $\Pr(B_1|H_1 \cap H_2) = (1/2)(1/2) = 1/4$. Then by Baye's Theorem,

$$\begin{aligned}\Pr(B_1|H_1 \cap H_2) &= \frac{\Pr(B_1)\Pr(H_1 \cap H_2|B_1)}{\Pr(B_1)\Pr(H_1 \cap H_2|B_1) + \Pr(B_2)\Pr(H_1 \cap H_2|B_2)} \\ &= \frac{(1/2)(1/4)}{(1/2)(1/4) + (1/2)(1)} = \frac{1}{5}.\end{aligned}$$

We can also compute this using the conditional version of Baye's Theorem. We know $\Pr(B_1|H_1) = 1/3$ so that $\Pr(B_2|H_1) = 2/3$. Also, $\Pr(H_2|B_1 \cap H_1) = \Pr(H_2|B_1) = 1/2$ by Theorem 2.2.4 (since H_1 and H_2 are conditionally independent given B_1) and, of course, $\Pr(H_2|B_2 \cap H_1) = \Pr(H_2|B_2) = 1$. So

$$\begin{aligned}\Pr(B_1|H_1 \cap H_2) &= \frac{\Pr(B_1|H_1)\Pr(H_2|B_1 \cap H_1)}{\Pr(B_1|H_1)\Pr(H_2|B_1 \cap H_1) + \Pr(B_2|H_1)\Pr(H_2|B_2 \cap H_1)} \\ &= \frac{(1/3)(1/2)}{(1/3)(1/2) + (2/3)(1)} = \frac{1}{5}\end{aligned}$$

as expected. □

Example 2.3.6. A machine produces defective parts in one of two proportions, either $p = 0.01$ or $p = 0.4$. Suppose that the prior probability that $p = 0.01$ is 0.9 (so the prior probability the $p = 0.4$ is 0.1. After sampling six parts at random, suppose that we observe two defectives. We want the posterior probability that $p = 0.01$. Let B_1 be the event that $p = 0.01$, let B_2 be the event that $p = 0.4$, and let A be the event that two defectives occur in a sample of size six. So we want to find $\Pr(B_1|A)$. We have $\Pr(B_1) = 0.9$, $\Pr(B_1) = 0.1$, $\Pr(A|B_1) = \binom{6}{2}(0.01)^2(0.99)^4 \approx 1.44 \times 10^{-3}$, and $\Pr(A|B_2) = \binom{6}{2}(0.4)^2(0.6)^4 \approx 0.311$. Then by Baye's Theorem

$$\Pr(B_1|A) = \frac{\Pr(B_1)\Pr(A|B_1)}{\Pr(B_1)\Pr(A|B_1) + \Pr(B_2)\Pr(A|B_2)}$$

$$\approx \frac{(0.9)(1.44 \times 10^{-3})}{(0.9)(1.44 \times 10^{-3}) + (0.1)(0.311)} \approx 0.04.$$

So the prior probability $\Pr(B_1) = 0.9$ is followed by a posterior probability of $\Pr(B_1|A) = 0.04$. This is because under event B_1 (i.e., $p = 0.01$), the probability of getting two defectives in a sample of size six is so small (namely, $\Pr(A|B_1) \approx 0.00144$).

Revised: 8/2/2019