Chapter 2. Conditional Probability

Section 2.1. The Definition of Conditional Probability

Note. In an experiment where two events A and B happen one after another, it is possible for the outcome of the first event, B, to affect the probability of the second event, A. For example, if two cards are drawn from a deck without replacement, the probability of getting two aces is the product of the probability that the first card i an ace, $4/52$, times the probability that the second card is an ace *given* that the first card is an ace, $3/51$. Notice that $\frac{4}{52}\frac{3}{51} = \binom{4}{2} \bigg/ \binom{52}{2}$. Notice that the probability of A given B requires a modified sample space of B and so equals $P(A \cap B)/P(B)$.

Definition 2.1.1. Suppose that we learn that an event B has occurred and that we wish to compute the probability of another event A taking into account that we know that B has occurred. The new probability of A is the *conditional probability* of the event A given that event B has occurred is denoted $\Pr(A|B)$. If $\Pr(B) \neq 0$ then we define

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

Note. DeGroot and Schervish say (page 56): "The conditional probability $\Pr(A|B)$ is not defined if $\Pr(B) = 0$." This need not be the case, though. Consider the function $f(x) = 1$ on $[0, 1]$. For $B \subset [0, 1]$ define $\Pr(B) = \int_B f(x)\,dx$. Then Pr is a probability measure on a σ-algebra of subsets of $[0, 1]$ which contains all open subsets of $[0, 1]$ (this σ-algebra includes all *Borel sets*), provided we use Lebesgue integration (in which case $\Pr(B) = m(B)$ where m denotes Lebesgue measure). For example, if a number in $[0, 1]$ is chosen using this probability measure, the probability that the number is between $1/2$ and 1 is $\int_{[1/2,1]} 1\,dx = 1/2$. Suppose a first and second number are chosen from $[0, 1]$. We want the probability that the sum of the two numbers is greater than or equal to 1 given that the first number is $1/2$; but this is simply the probability that the second numbers is in $[1/2, 1]$, which is $1/2$. However, notice that the probability that the first number is $1/2$ is $\int_{[1/2,1/2]} 1\,dx = 0$. So this is an example of a conditional probability where the probability of the first event is 0. So it isn't that such a conditional probability *cannot* be defined, but that it is not addressed in this level of a class. For more details on this idea, see my online notes on Measure Theory Based Probability, in particular the section 5.3. The General Concept of Conditional Probability and Expectation.

Example 2.1.3. Rolling Dice. Suppose that two dice were rolled and it is known that the sum T of the resulting two numbers was odd. We calculate the probability that T was less than 8. We have the sample space in Example 1.6.5. Let A be the event that $T < 8$ and let B be the event that T is odd. We have in the notation of Example 1.6.5 that

$$A \cap B = \{(1, 2), (1, 4), (1, 6), (2, 1), (2, 3), (2, 5), (3, 2), (3, 4), (4, 1), (4, 3), (5, 2), (6, 1)\}$$

and

$$B = \{(1,2),(1,4),(1,6),(2,1),(2,3),(2,5),(3,2),(3,4),(3,6),$$
$$(4,1),(4,3),(4,5),(5,2),(5,4),(5,6),(6,1),(6,3),(6,5)\}.$$

So $\Pr(A \cap B) = 12/36 = 1/3$ and $\Pr(B) = 18/36 = 1/2$. So

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)} = \frac{1/3}{1/2} = \frac{2}{3}. \qquad \square$$

Example 2.1.5. Rolling Dice Repeatedly. Suppose that two dice are to be rolled repeatedly and the sum T of the two numbers is to be observed for each roll. We calculate the probability that the value $T = 7$ will be observed before the value $T = 8$ is observed. We do so using conditional probability by computing the probability that $T = 7$, given that $T \in \{7, 8\}$. Let A be the event that $T = 7$ and let B be the event that the value of T is either 7 or 8. Notice $A \cap B = A$. By example 1.6.5 we have $\Pr(A) = \Pr(A \cap B) = P_7 = 6/36 = 1/6$ and $\Pr(B) = P_7 + P_8 = 6/36 + 5/36 = 11/36$. So the desired probability is:

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)} = \frac{1/6}{11/36} = \frac{6}{11}. \qquad \square$$

Note. From Definition 2.1.1, "conditional probability," we immediately have the following.

Theorem 2.1.1. Multiplication Rule for Conditional Probabilities. Let A and B be events. If $\Pr(B) \neq 0$ then $\Pr(A \cap B) = \Pr(B)\Pr(A|B)$. If $\Pr(A) \neq 0$ then $\Pr(A \cap B) = \Pr(A)\Pr(B|A)$. If $\Pr \neq 0 \neq \Pr(B)$ then

$$\Pr(A|B) = \frac{\Pr(A)\Pr(B|A)}{\Pr(B)} \text{ and } \Pr(B|A) = \text{Pr(B) Pr(A—B)}\Pr(A).$$

Example. An application of Theorem 2.1.1 can be found in the Math Fun Facts webpage of Harvey Mudd College. "Suppose that you are worried that you might have a rare disease. You decide to get tested, and suppose that the testing methods for this disease are correct 99 percent of the time (in other words, if you have the disease, it shows that you do with 99 percent probability, and if you don't have the disease, it shows that you do not with 99 percent probability). Suppose this disease is actually quite rare, occurring randomly in the general population in only one of every 10,000 people. If your test results come back positive, what are your chances that you actually have the disease?" We let A be the event that one has the disease, and let B be the event that one tests positive for the disease. Then we are given $\Pr(A) = 1/10,000 = 0.0001$, $\Pr(B|A) = 0.99$, $\Pr(A^c) = 9,999/10,000 = 0.9999$, and $\Pr(B|A^c) = 0.01$ ($\Pr(B|A^c)$ is the probability of a "false positive"), and we want to find $\Pr(A|B)$. We will see below (in Theorem 2.1.4) that we can calculate $\Pr(B)$ as

$$\Pr(B) = \Pr(B|A)\Pr(A) + \Pr(B|A^c)\Pr(A^c)$$

$$= (0.99)(0.0001) + (0.01)(0.9999) = 0.010098.$$

We then have by Theorem 2.1.1 that the desired probability is

$$\Pr(A|B) = \frac{\Pr(A)\Pr(B|A)}{\Pr(B)} = \frac{(0.0001)(0.99)}{.010098} \approx 0.0098.$$

So given that you tested positive, the probability that you actually have the disease is less then 1%. Numbers such as these are an argument against widespread drug testing, for example. There are some problems with this particular example, though. The statement of the problem includes the claim that "you are worried that you might have a rare disease." So it appears that "you" have some additional information leading you to this suspicion; so it does not sound like you were chosen at random from the population, thus affecting the value of $\Pr(A)$ (notice that if $\Pr(A)$ increases then it has a strong effect on $\Pr(A|B)$ in this case). Notice that in drug testing a population (such as job applicants), presumably the probability of drug use is small, individuals are chosen at random to be tested, and then the numbers above are realistic indicating the probability of a large number of false positives.

Theorem 2.1.2. Multiplication Rule for Conditional Probabilities. Suppose that $A_1, A_2, \ldots, A_n$ are events such that $\Pr(A_1 \cap A_2 \cap \cdots \cap A_{n-1}) \neq 0$. Then

$$\Pr(A_1 \cap A_2 \cap \cdots \cap A_n) = \Pr(A_1)\Pr(A_2|A_1)\Pr(A_3|A_1 \cap A_2)\Pr(A_4|A_1 \cap A_2 \cap A_3) \cdots \Pr(A_n|A_1 \cap A_2 \cap \cdots \cap A_{n-1}).$$

Theorem 2.1.3. Suppose $A_1, A_2, \ldots, A_n, B$ are events such that $\Pr(B) \neq 0$ and $\Pr(A_1 \cap A_2 \cap \cdots \cap A_{n-1}|B) \neq 0$. Then

$$\frac{\Pr(A_1 \cap A_2 \cap \cdots \cap A_n|B)}{\Pr(B)} = \Pr(A_1|B)\Pr(A_2|A_1 \cap B)\Pr(A_3|A_1 \cap A_2 \cap B) \cdots \Pr(A_n|A_1 \cap A_2 \cap \cdots \cap A_{n-1} \cap B).$$

Definition 2.1.2. Let S denote the sample space of some experiment and consider k events $B_1, B_2, \ldots B_k$ in S such that $B_1, B_2, \ldots, B_k$ are disjoint and $\cup_{i=1}^{k} B_i = S$. These events *partition* S. We can similarly partition any event that is a subset of S.

Theorem 2.1.4. Law of Total Probability. Suppose the events $B_1, B_2, \ldots, B_k$ form a partition of sample space S and $\Pr(B_j) \neq 0$ for $j = 1, 2, \ldots, k$. Then, for every event A in S, $\Pr(A) = \sum_{j=1}^{k} \Pr(B_j)\Pr(A|B_j)$.

Note. Similar to Theorem 2.1.4, we can express a conditional probability using a partition. As is to be shown in Exercise 2.1.17, if $B_1, B_2, \ldots, B_k$ is a partition of sample space S and $\Pr(B_j) \neq 0$ for $j = 1, 2, \ldots, k$. Then for events A and C where $\Pr(C) \neq 0$ we have

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

Examples 2.1.9 and 2.1.11. There is one box of bolts that contains some long and some short bolts. A manager is unable to open the box at present, so she asks her employees what is the composition of the box. One employer says that it contains 60 long bolts and 40 short bolts. Another says that it contains 10 long bolts and 20 short bolts. Unable to reconcile these opinions, the manager decides that each of the employees is correct with probability 1/2. Let B_1 be the event that the box contains 60 long and 40 short bolts, and let B_2 be the event that the box contains 1- long and 20 short bolts. Let A be the event that a randomly chosen bolt is long. Since B_1 and B_2 partition the sample space, then by Theorem 2.1.4 $\Pr(A) = \Pr(B_i)\Pr(A|B_1) + \Pr(B_2)\Pr(A|B_2)$. We have $\Pr(B_1) = \Pr(B_2) = 1/2$, $\Pr(A|B_1) = 3/5$, and $\Pr(A|B_2) = 1/3$, so that

$$\Pr(A) = (1/2)(3/5) + (1/2)(1/3) = 3/10 + 1/6 = 14/30 = 7/15. \qquad \square$$

Note. We state a final definition which is "worded somewhat vaguely" (see page 63).

Definition 2.1.3. If desired, any *experiment* can be *augmented* to include the potential or hypothetical observation of as much additional information as we would find useful to help us calculate any probabilities that we desire.

Example. The Game of Craps. In craps, a player rolls two dice, and the sum of the two numbers is observed. If the sum on the first roll is 7 or 11, the player wins the game immediately. If the sum on the first roll is 2, 3, or 12, the player loses the game immediately. If the sum on the first roll is 4, 5, 6, 8, 9, or 10, then the two dice are rolled again and again until the sum is either 7 or the original value. If the original value is obtained a second time before 7 if obtained, then the player wins. If the sum 7 is obtained before the original value is obtained a second time, then the player loses. Let W be the event that the player wins. Let B_i be the event that the first roll is i (where $i \in \{2, 3, \ldots, 12\}$). Then we have $\Pr(W|B_2) = \Pr(W|B_3) = \Pr(W|B_{12}) = 0$ and $\Pr(W|B_2) = \Pr(W|B_{11}) = 1$. Now for $i \in \{4, 5, 6, 8, 9, 10\}$,

$\Pr(W|B_i)$ is the probability of rolling a sum of i before rolling a sum of 7. We saw in Example 2.1.5 (for $i = 8$) that $\Pr(W|B_i) = \dfrac{\Pr(B_i)}{\Pr(B_i \cup B_7)}$. Using the relevant probabilities from Example 1.6.5 we find

$$\Pr(W|B_4) = \Pr(W|B_{10}) = \frac{3/36}{3/36 + 6/36} = \frac{1}{3},$$

$$\Pr(W|B_5) = \Pr(W|B_9) = \frac{4/36}{4/36 + 6/36} = \frac{2}{5},$$

$$\Pr(W|B_6) = \Pr(W|B_8) = \frac{5/36}{5/36 + 6/36} = \frac{5}{11}.$$

By Theorem 2.1.4,

$$\Pr(W) = \sum_{i=2}^{12} \Pr(B_i)\Pr(W|B_i)$$

$$= 0 + 0 + \frac{3}{36}\frac{1}{3} + \frac{4}{36}\frac{2}{5} + \frac{5}{36}\frac{5}{11} + \frac{6}{36} + \frac{5}{36}\frac{5}{11} + \frac{4}{36}\frac{2}{5}\frac{3}{36}\frac{1}{3} + \frac{2}{36} + 0 = \frac{2928}{5940} \approx 0.493.$$

□

Revised: 7/30/2019