Section 1.4. Conditional Probability and Independence

Note. The idea behind conditional probability is that the initial sample space $\mathcal{C}$ has been replaced with some subset $A \subset \mathcal{C}$. In practice, this could be due to some additional information about the outcome of an experiment. For example, you might choose a card at random from a deck of 52. The probability that you draw a king is 1/52. But if you are told that you have a face card (that is, a J, Q, or K) then the probability is $4/12 = 1/3$. When set $A \subset \mathcal{C}$ is the new sample space, that is the "given" information, then the probability of event $B \subset \mathcal{C}$ is denoted $P(B \mid A)$, read "the probability of B given A." We desire $P(A \mid A) = 1$ and $P(B \mid A) = P(A \cap B \mid a)$. We expect the ratio of the probabilities of events $A \cap B$ and A (relative to space A) to be the same as the ration of the probabilities of these events relative to space $\mathcal{C}$: $\dfrac{P(A \cap B \mid A)}{P(A \mid A)} = \dfrac{P(A \cap B)}{P(A)}$ (Hogg, McKean, and Craig inspire this with the comment about the "relative frequency point of view"; see page 24). This leads to the following definition.

Definition 1.4.1. Let B and A be events with $P(A) > 0$. Then the *conditional probability* of B given A as $P(B \mid A) = \dfrac{P(A \cap B)}{P(A)}$.

Theorem 1.4.A. Let $A, B, B_1, B_2, \ldots$ be events with $P(A) > 0$. Then

1. $P(B \mid A) \geq 0$.

2. $P(A \mid A) = 1$.

3. $P(\cup_{n=1}^{\infty} B_n \mid A) = \sum_{n=1}^{\infty} P(B_n \mid A)$ provided $B_1, B_2, \ldots$ are mutually exclusive.

Note. Hogg, McKean, and Craig comment on page 24: "It should be noted that this conditional probability set function, given A, is defined at this time only when $P(A) > 0$." But there are cases when $P(B \mid A)$ is meaningful when $P(A) = 0$. Consider the function $f(x) = 1$ on $[0, 1]$. For $A \subset [0, 1]$ define $P(A) = \int_A f(x)\, dx$. Then P is a probability set function on a σ-field of subsets of $[0, 1]$ which contains all open subsets of $[0, 1]$ (this σ-field includes all *Borel sets*), provided we use Lebesgue integration (in which case $P(B) = m(B)$ 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.

Note/Definition. If A and B are events where $P(A) > 0$ then $P(A \cap B) = P(A)P(B \mid A)$ by Definition 1.4.1. This is called the *multiplication rule* also. For these events A, B, C where $P(A \cap B) > 0$, we have

$$P(A\cap B\cap C) = P((A\cap B)\cap C) = P(A\cap B)P(C \mid A\cap B) = P(a)P(B \mid A)P(C \mid A\cap B).$$

Of course, by mathematical induction, this can be extended to any finite number of events.

Example. Exercise 1.4.6.

Theorem 1.4.B. Law of Total Probability.
Let $A_1, A_2, \ldots, A_k$ be events such that $P(A_i) > 0$ for $i = 1, 2, \ldots, k$ and are mutually exclusive and exhaustive (that is, $\mathcal{C} = \cup_{i=1}^{k} A_k$). Let B be another event such that $P(B) > 0$. Then

$$P(B) = \sum_{i=1}^{k} P(A_i)P(B \mid A_i).$$

Note. We now state and prove the most important result concerning conditional probability.

Theorem 1.4.1. Bayes' Theorem.
Let $A_1, A_2, \ldots, A_k$ be events such that $P(A_i) > 0$ for $i = 1, 2, \ldots, k$. Assume that $A_1, A_2, \ldots, A_k$ form a partition of the sample space $\mathcal{C}$. Let B be any event. Then for each $j = 1, 2, \ldots, k$ we have

$$P(A_j \mid B) = \frac{P(A_j)P(B \mid A_j)}{\sum_{i=1}^{k} P(A_i)P(B \mid A_i)}.$$

Example. Example 1.4.5.

Note/Definition. In Example 1.4.5, given event B we know that a red chip has been selected, so it is intuitive that $P(A_2 \mid B) > P(A_1 \mid B)$ since Bowl A_2 contains more red chips that Bowl A_1. The probabilities $P(A_1)$ and $P(A_2)$ are called *prior probabilities* since they are given simply based on the method of choice of Bowl A_1 or Bowl A_2. The probabilities $P(A_1 \mid B)$ an d$P(A_2 \mid B)$ are called *posterior probabilities* since they reflect how $P(A_1)$ and $P(A_2)$ change after the additional given information that event B has occurred.

Example 1.4.A. 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 $P(A) = 1/10,000 = 0.0001$, $P(B|A) = 0.99$, $P(A^c) = 9,999/10,000 = 0.9999$, and $P(B|A^c) = 0.01$ ($P(B|A^c)$ is the probability of a "false positive"), and we want to find $P(A|B)$. By Bayes' Theorem, Theorem 1.4.1,

$$P(A|B) = \frac{P(A)P(B|A)}{P(B|A)P(A) + P(B|A^c)P(A^c)} = \frac{(0.0001)(0.99)}{(0.99)(0.0001) + (0.01)(0.9999)}$$

$$= \frac{(0.0001)(0.99)}{.010098} \approx 0.0098.$$

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 $P(A)$ (notice that if $P(A)$ increases then it has a strong effect on $P(A|B)$ 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.

Note. Thomas Bayes was born on the outskirts of London and graduated in 1719 from the University of Edinburgh where he studied logic and theology. He became a minister in a Presbyterian chapel near London. Bayes published his theory of probability in "Essay Towards Solving a Problem in the Doctrine of Chances," *Philosophical Transactions of the Royal Society of London* in 1764. He was elected a Fellow of the Royal Society in 1742, even though he never published in math in his lifetime under his own name, although he did work in the foundations of calculus (on "the theory of fluxions" and some work on series). This biograpical information (and the following image) is from the MacTutor History of Mathematics Archive biography of Thomas Bayes.

Thomas Bayes (1702–1761)

Note. If the occurrence of event B has no effect on event A, and conversely, then the events are "independent." More formally, we have the following.

Definition 1.4.2. Let A and B be two events. Then A and B are *independent* if $P(A \cap B) = P(A)P(B)$.

Theorem 1.4.C. Suppose A and B are independent events. The the following three pairs of events are independent: A^c and B, A and B^c, and A^c and B^c.

Definition. Events $A_1, A_2, \ldots, A_n$ are *mutually independent* if for every collection of k events of these $2 \le k \le n$ and for every permutation $d_1, d_2, \ldots, d_k$ of $1, 2, \ldots, k$, we have

$$P(A_{d_1} \cap A_{d_2} \cap \cdots \cap A_{d_k}) = P(A_{d_1})P(A_{d_2}) \cdots P(A_{d_k}).$$

In particular, if $A_1, A_2, \ldots, A_n$ are mutually independent then $P(A_1 \cap A_2 \cap \cdots \cap$

$A_n) = P(A_1)P(A_2)\cdots P(A_n)$.

Note. As with two sets, combinations of mutually independent events and their complements are mutually independent.

Example. Exercise 1.4.18.

Exercise 1.4.30. Monty Hall Problem.

Suppose there are three curtains. Behind one curtain there is a nice prize, while behind the other two there are worthless prizes. A contestant selects one curtain at random, and then Monty Hall (of *Let's Make a Deal* fame) opens one of the other two curtains to reveal a worthless prize. Hall then expresses a willingness to trade the curtain that the contestant has chosen for the other curtain that has not been opened. Should the contestant switch curtains or stick with the one that she has? To answer the question, determine the probability that she wins the prize if she switches.

Solution. Unlike many solutions to be found online and in the literature, we explicitly give the sample space and describe each possible outcome in terms of events in the sample space. We take indices i, j, k to be $1, 2, 3$ and we number the curtains $\#1, \#2, \#3$. Three things happen in the game. First, the nice prize is behind some curtain; if the nice prize is behind curtain i, denote this as P_i. Second, the contestant chooses a curtain; we denote the choice of curtain $\#j$ as C_j. Finally, Monty Hall reveals what is behind curtain k, which we denote of R_k. The sample space is then $\{(P_i, C_j, R_k) \mid i, j, k = 1, 2, 3\}$. The event that the contestant selects the curtain containing the nice prize is the event $\{(P_i, C_i, R_k) \mid i, k = 1, 2, 3\}$. We assume that the prize is randomly assigned to one of the curtains and that the contestant randomly chooses a curtain. The event that the nice prize is behind curtain $\#i$ is $\{(P_i, C_j, R_k) \mid j, k = 1, 2, 3\}$ and the event that the contestant chooses curtain $\#j$ is $\{(P_i, C_j, R_k) \mid 1, k = 1, 2, 3\}$, and both of the events have probability 1/3. We also assume the Monty Hall neither reveals the location of the nice prize nor reveals what is behind the chosen curtain; so $((P_1, C_j, R_i) = 0$ for $i, j = 1, 2, 3$ and $P((P_i, C_j, R_j)) = 0$ for $i, j = 1, 2, 3$. Finally, we assume that when the contestant has chosen the curtain with the nice prize then Monty Hall reveals one of the other two curtains with probability 1/2 (with the flip of a coin, for example). We then have the following 27 possibilities:

P_1

	R_1	R_2	R_3
C_1	0	1/18	1/18
C_2	0	0	1/9
C_3	0	1/9	0

P_2

	R_1	R_2	R_3
C_1	0	0	1/9
C_2	1/18	0	1/18
C_3	1/9	0	0

P_3

	R_1	R_2	R_3
C_1	0	1/9	0
C_2	1/9	0	0
C_3	1/18	1/18	0

So the probability that the contestant wins without changing after the revelation is simply the probability of the event $W_1 = \{(P_i, C_i, R_k) \mid i, k = 1, 2, 3\}$ is

$$P(W_1) = \left(\frac{1}{18} + \frac{1}{18}\right) + \left(\frac{1}{18} + \frac{1}{18}\right) + \left(\frac{1}{18} + \frac{1}{18}\right) = \frac{1}{3};$$

the elements of the event with nonzero probability are (P_1, C_1, R_2), (P_1, C_1, R_3), (P_2, C_2, R_1), (P_2, C_2, R_3), (P_3, C_3, R_1), and (P_3, C_3, R_2). The probability that the contestant loses without changing after the revelation is therefore 2/3 (or

$$\left(\frac{1}{9} + \frac{1}{9}\right) + \left(\frac{1}{9} + \frac{1}{9}\right) + \left(\frac{1}{9} + \frac{1}{9}\right) = \frac{2}{3};$$

the elements of the event with nonzero probability are (P_1, C_2, R_3), (P_1, C_3, R_2), (P_2, C_1, R_3), (P_2, C_3, R_1), (P_3, C_1, R_2), and (P_3, C_2, R_1)). With the obvious assumptions that, if the contestant switches curtains, then she does not switch to the revealed curtain (which has been revealed to contain a worthless prize), then the contestant wins the nice prize by switching if and only if the contest originally had NOT chosen the curtain hiding the nice prize. So the event that represents the contestant winning after switching is $W_2 = \{(P_i, C_j, R_k) \mid u, j, k = 1, 2, 3 \text{ and } i \neq j\}$ which has probability

$$P(W_2) = \left(\frac{1}{9} + \frac{1}{9}\right) + \left(\frac{1}{9} + \frac{1}{9}\right) + \left(\frac{1}{9} + \frac{1}{9}\right) = \frac{2}{3}.$$

Therefore, a contestant doubles her probability of winning by switching after the reveal.

Note. To intuitively understand the surprising result of the Monty Hall Problem, notice that the reveal does not give the contestant any information about what she *has* chosen and so the probability that the contestant has originally chosen the curtain with the nice prize does not change following the reveal. But the reveal does give the contestant information about what she *has not* chosen and this new information changes the probability that the unrevealed and unchosen curtain contains the the nice prize (the probability changes from 1/3 to 1/3).

Note. For a possibly more convincing intuitive argument, suppose that there were 100 instead of 3 curtains. If the contestant chooses curtain #35, say, and Monty Hall reveals what is behind the 98 curtains marked #1–#34, #36–#73, and

#75–#100, showing that each of these contain a worthless prize. The contestant would think it unlikely that she had chosen the curtain with the nice prize (with probability 1/100) and more likely that the nice prize is behind one of the other curtains (with probability 99/100). Learning what is behind 98 of the 99 other curtains is very suggestive that something is special about curtain #74; in fact, there is a 99/100 probability (after the reveal) that this curtain contains the nice prize.

Note. The Monty Hall Problem has a well-documented, and sometimes humorous, background. It spread widely through popular culture in 1990 when Marilyn vos Savant discussed it in her "Ask Marilyn" column in *Parade Magazine* (a widely circulated insert in the Sunday newspaper of many newspapers throughout the U.S.). Many Ph.D.s wrote her explaining whey her argument is incorrect. The solution she gave is, roughly, the (correct) one given above. Details on the history of the problem and some related problems can be found on the Monty Hall Problem Wikipedea page. For more academic references, see the following two papers which appeared in publications of the Mathematical Association of America:

1. Ed Barbeau, Fallacies, Flaws, and Flimflam, *The College Mathematics Journal* **24**(2), 149–154 (March 1993).

2. Leonard Gillman, The Car and the Goats, *The American Mathematical Monthly* bf 99(1), 3–7 (January 1992).

Revised: 10/14/2022