Chapter 4. Conditional probability.

In many situations we have partial information about the outcome of an experiment and we may wish to update the probability measure to reflect this additional information. For example, suppose that a population of N people contains N_A females and N_B college graduates. For an individual selected at random from this population let A denote the event that the individual is a female and let B denote the event that the individual is a college graduate. Then, $\Pr(A) = \frac{N_A}{N}$ and $\Pr(B) = \frac{N_B}{N}$. Now suppose that it is known that the individual is a female. That is suppose that we know that the individual belongs to the subpopulation of females (event A). Using the subpopulation A as our reference space (new sample space) we note that the event that the individual is a college graduate given that the individual is a female is the intersection AB. Thus letting N_{AB} denote the number of people in this population who are female college graduates, we find that the conditional probability of the selected individual being a college graduate given that the selected individual is a female is

$$\Pr(B|A) = \frac{N_{AB}}{N_A} = \frac{\Pr(AB)}{\Pr(A)}.$$

Definition. *Given events A and B with $Pr(A) > 0$ the conditional probability of B given A is*

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

Theorem 4.1 (Multiplication rule). *Given events A and B with $Pr(A) > 0$,*

$$Pr(AB) = Pr(A)Pr(B|A).$$

Proof. Obvious from the definition of $\Pr(B|A)$. □

Note that if $\Pr(AB) > 0$, then $\Pr(A) > 0$ and $\Pr(B) > 0$ and we have $\Pr(AB) = \Pr(A)\Pr(B|A) = \Pr(B)\Pr(A|B)$.

Corollary 4.2. *Given events $A_1, \ldots, A_n$ with $Pr(A_1A_2\cdots A_n) > 0$,*

$$Pr(A_1A_2\cdots A_n) = Pr(A_1)Pr(A_2|A_1)Pr(A_3|A_1A_2)\cdots Pr(A_n|A_1A_2\cdots A_{n-1})$$

Corollary 4.3. *Given events $A_1, \ldots, A_n$ and B with $Pr(A_1\cdots A_nB) > 0$,*

$$Pr(A_1A_2\cdots A_n|B) = Pr(A_1|B)Pr(A_2|A_1B)Pr(A_3|A_1A_2B)\cdots Pr(A_n|A_1A_2\cdots A_{n-1}B)$$

Intuitively we say that the events A and B are independent (stochastically independent) when knowing that B has occurred has no effect on the probability of occurrence of A, *i.e.* when $\Pr(A) = \Pr(A|B)$. For mathematical convenience the formal definition of independence is in terms of a product so that it does not depend on the existence of conditional probabilities.

Definition. *The events A and B are said to be independent (stochastically independent) when $Pr(AB) = Pr(A)Pr(B)$.*

Example. If a fair die is tossed once and we let $A = \{2, 4, 6\}$ denote the event that an even value occurs and $B = \{1, 2, 3, 4\}$ the event that the value is four or less, then $\Pr(A) = \frac{1}{2}$, $\Pr(B) = \frac{2}{3}$, and $\Pr(AB) = \frac{1}{3}$. Thus, in this case, $\Pr(AB) = \Pr(A)\Pr(B)$ and A and B are independent.

Note that if two events are disjoint (mutually exclusive), then they cannot occur at the same time; thus if A and B are mutually exclusive, then they cannot be independent unless one or both is the null event.

Theorem 4.4. *If the events A and B are independent, then: the events A and B^c are independent; the events A^c and B are independent; and, the events A^c and B^c are independent.*

Proof. Let the independent events A and B be given. We will show that A and B^c are independent, the other results are proved analogously. Theorem 2.5 implies that $\Pr(AB^c) = \Pr(A) - \Pr(AB)$. Thus the independence of A and B implies that $\Pr(AB^c) = \Pr(A) - \Pr(A)\Pr(B) = \Pr(A)(1 - \Pr(B))$ which establishes the result. □

Definition. *The events $A_1, \ldots, A_n$ are said to be independent (mutually independent) when*
$Pr(A_iA_j) = Pr(A_i)Pr(A_j)$ for all pairs (i, j) with distinct elements
$Pr(A_iA_jA_k) = Pr(A_i)Pr(A_j)Pr(A_k)$ for all triples (i, j, k) with distinct elements
and so on for sets of four, five, ..., up to
$Pr(A_i \cdots A_n) = Pr(A_1) \cdots Pr(A_n)$.

Definition. *The events $A_1, \ldots, A_n$ are said to be pairwise independent when*
$Pr(A_iA_j) = Pr(A_i)Pr(A_j)$ for all pairs (i, j) with distinct elements.

Example. Let $\Pr(\omega) = 1/8$ for $\omega \in \Omega = \{1, 2, 3, 4, 5, 6, 7, 8\}$, let $A = \{1, 2, 3, 4\}$, $B = \{1, 2, 5, 6\}$, and $C = \{1, 3, 5, 7\}$. Then $\Pr(A) = \Pr(B) = \Pr(C) = \frac{1}{2}$, $\Pr(AB) = \Pr((AC) = \Pr(BC) = \frac{1}{4} = (\frac{1}{2})^2$, and $\Pr(ABC) = \frac{1}{8} = (\frac{1}{2})^3$. Thus, in this example, the events A, B, and C are independent (mutually independent).

Example. Let $\Pr(\omega) = 1/8$ for $\omega \in \Omega = \{1, 2, 3, 4, 5, 6, 7, 8\}$, let $A = \{1, 2, 3, 4\}$, $B = \{1, 2, 5, 6\}$, and $C = \{1, 2, 7, 8\}$. Then $\Pr(A) = \Pr(B) = \Pr(C) = \frac{1}{2}$ and $\Pr(AB) = \Pr((AC) = \Pr(BC) = \frac{1}{4} = (\frac{1}{2})^2$. But $\Pr(ABC) = \frac{1}{4} \neq (\frac{1}{2})^3$. Thus, in this example, the events A, B, and C are pairwise independent but not mutually independent.

Recall that if $\Pr(AB) > 0$, then A and B are independent if, and only if $\Pr(A|B) = \Pr(A)$ and $\Pr(B|A) = \Pr(B)$. A similar result holds for a collection of events.

Theorem 4.5. *Given events* $A_1, \ldots, A_n$ *with* $Pr(A_1 A_2 \cdots A_n) > 0$.
The events $A_1, \ldots, A_n$ *are independent if, and only if*

$$Pr(A_{i_1} \cdots A_{i_a} | A_{j_1} \cdots A_{j_b}) = Pr(A_{i_1} \cdots A_{i_a})$$

for all nonempty disjoint subsets $\{i_1, \ldots, i_a\}$ *and* $\{j_1, \ldots, j_b\}$ *of* $\{1, \ldots, n\}$.

Definition. *Given events* A, B, *and* C *with* $Pr(ABC) > 0$, *the events* A *and* B *are said to be conditionally independent given the event* C *when* $Pr(AB|C) = Pr(A|C)Pr(B|C)$.

Theorem 4.6 (The law of total probability). *If the events* $B_1, \ldots, B_n$ *form a partition of* Ω, *i.e. if* $B_i B_j = \emptyset$ *for all* $i \neq j$ *and* $\Omega = B_1 \cup \cdots \cup B_n$, *then, for any event* A,

$$Pr(A) = \sum_{i=1}^{n} Pr(AB_i).$$

Corollary 4.7. *If the events* $B_1, \ldots, B_n$ *form a partition of* Ω, *and* $Pr(B_i) > 0$ *for* $i = 1, \ldots, n$, *then*

$$Pr(A) = \sum_{i=1}^{n} Pr(A|B_i)Pr(B_i).$$

Corollary 4.8 (Bayes' theorem). *If the events* $B_1, \ldots, B_n$ *form a partition of* Ω, *and* $Pr(B_i) > 0$ *for* $i = 1, \ldots, n$, *then for any event* A *with* $Pr(A) > 0$, *we have*

$$Pr(B_1|A) = \frac{Pr(A|B_1)Pr(B_1)}{\sum_{i=1}^{n} Pr(A|B_i)Pr(B_i)}.$$

Bayes' theorem is particularly useful for a situation where the occurrence of event A follows the occurrence of one of the events B_i in time and we are interested in the conditional probability that a particular B_i, say B_1, has occurred given that event A has occurred.

Note that if A and B are events and $0 < \Pr(B) < 1$, then the events AB and AB^c form a partition of Ω. Thus $\Pr(A) = \Pr(AB) + \Pr(AB^c)$ and Bayes' theorem reduces to

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

Example. Consider a box containing 100 balls of which 20 are labeled A, 30 are labeled B, and 50 are labeled C, and three other boxes labeled A, B, and C such that: box A contains 8 red and 2 green balls; box B contains 7 red and 3 green balls; and, box C contains 6 red and 4 green balls. Now suppose that a ball is chosen at random from the box containing 100 balls, the letter (A, B, C) on the ball is noted, and then a ball is chosen at random from the 10 balls in the box with the appropriate letter label. Obviously, the conditional probabilities of choosing a red ball given the letter label are: $\Pr(R|A) = .8$, $\Pr(R|B) = .7$, and $\Pr(R|C) = .6$. It is also obvious that the probabilities of selecting the label (A, B, C) are: $\Pr(A) = .2$, $\Pr(B) = .3$, and $\Pr(C) = .5$. The values of conditional probabilities of the form $\Pr(A|R)$, the conditional probability that the ball was selected from box A given that it was red, are less obvious. However, these conditional probabilities are readily computed using Bayes' Theorem. Thus

$$\Pr(R) = \Pr(R|A)\Pr(A) + \Pr(R|B)\Pr(B) + \Pr(R|C)\Pr(C) = .16 + .21 + .30 = .67$$

$$\Pr(A|R) = \frac{\Pr(R|A)\Pr(A)}{\Pr(R)} = \frac{16}{67} \approx .24$$

$$\Pr(B|R) = \frac{\Pr(R|B)\Pr(B)}{\Pr(R)} = \frac{21}{67} \approx .31$$

$$\Pr(C|R) = \frac{\Pr(R|C)\Pr(C)}{\Pr(R)} = \frac{30}{67} \approx .45$$

It is interesting to compare the unconditional probabilities of drawing from boxes A, B, and C, $\Pr(A) = .2$, $\Pr(B) = .3$, and $\Pr(C) = .5$, to the corresponding conditional probabilities given that the ball drawn is known to be red, $\Pr(A|R) = \frac{16}{67} \approx .24$, $\Pr(B|R) = \frac{21}{67} \approx .31$, and $\Pr(C|R) = \frac{30}{67} \approx .45$. The initial probabilities (before we obtain the additional information that the ball drawn was red) are known as prior probabilities and the updated probabilities (conditional on the added information) are known as posterior probabilities.

Example. Suppose balls (objects) are selected at random with replacement from a population of N balls, of which N_1 are red (R), N_2 are green (G), and $N_3 = N - N_1 - N_2$ are black (B), sequentially until either a red ball is selected or a green ball is selected. In this context the elementary outcomes can be represented by finite sequences of the form R, G, BR, BG, BBR, BBG, $\ldots$, *i.e.* sequences of the form $B \ldots BR$ or $B \ldots BG$. What is the probability that a red ball will be selected before a green ball is selected? Reasoning as in the geometric distribution example of Section 3.3, it is clear that

$$\Pr(\text{red before green}) = \sum_{x=0}^{\infty} \left(\frac{N_3}{N}\right)^x \left(\frac{N_1}{N}\right) = \frac{N_1}{N - N_3} = \frac{N_1}{N_1 + N_2},$$

and analogously

$$\Pr(\text{green before red}) = \frac{N_2}{N_1 + N_2}.$$

There is an interesting connection between these probabilities and certain conditional probabilities defined in terms of the selection of a single ball from this population. Note that if one ball is selected at random, then $\Pr(R) = \frac{N_1}{N}$, $\Pr(G) = \frac{N_2}{N}$, and $\Pr(R \text{ or } G) = \frac{N_1+N_2}{N}$.

Thus, the conditional probability of selecting a red ball given that the ball selected is red or green is $\Pr(R|R \text{ or } G) = \frac{N_1}{N_1+N_2}$. Similarly, the conditional probability of selecting a green ball given that the ball selected is red or green is $\Pr(G|R \text{ or } G) = \frac{N_2}{N_1+N_2}$.

Example. Craps. Craps is a popular dice game. In this game a pair of fair dice is thrown (tossed) and the sum of the numbers on the dice is computed; this action is repeated until the player either wins or loses. The outcome of the game is determined on the first throw when the player throws: *seven or eleven* “a natural” in which case the player wins or *two, three or twelve* “craps” in which case the player loses (craps out). If any other sum is thrown *four, five, six, eight, nine, or ten*, then the number thrown becomes the player’s “point” and play continues until the player makes his point or throws a seven. If the player makes his point he wins and if he throws a seven he loses (craps out).

The probabilities of the various sums on a single throw, which were computed in an example in Section 3.1, are: $\Pr(2) = \Pr(12) = \frac{1}{36}$, $\Pr(3) = \Pr(11) = \frac{2}{36}$, $\Pr(4) = \Pr(10) = \frac{3}{36}$, $\Pr(5) = \Pr(9) = \frac{4}{36}$, $\Pr(6) = \Pr(8) = \frac{5}{36}$, and $\Pr(7) = \frac{6}{36}$. The probability of winning on the first throw is

$$p_0 = \Pr(7 \text{ or } 11) = \tfrac{8}{36}.$$

The other ways to win correspond to throwing a 4, 5, 6, 8, 9, or 10 on the first throw and then making this point in a sequence of throws. Consider first the case when the player throws a 4 on the first throw. It is easy to see that the event that the player makes his point by throwing a 4 before a 7 is independent of the outcome of the first toss so that the probability of winning with a 4 is

$$p_4 = \Pr(4 \text{ on the first throw})\Pr(\text{making the point } 4)$$

Appealing to the preceding example the probability of making the point 4 is equal to the conditional probability of throwing a 4 on a single throw given that the single throw results in a 4 or a 7. Hence p_4 is given by the product

$$p_4 = \Pr(4)\Pr(4|4 \text{ or } 7) = \left(\tfrac{3}{36}\right)\left(\tfrac{3}{3+6}\right) = \tfrac{9}{36\cdot 9} = \tfrac{1}{36}$$

Applying this argument to the other possible point values gives:

$$p_5 = \Pr(5)\Pr(5|5 \text{ or } 7) = \left(\tfrac{4}{36}\right)\left(\tfrac{4}{4+6}\right) = \tfrac{16}{36\cdot 10} = \tfrac{2}{45}$$

$$p_6 = \Pr(6)\Pr(6|6 \text{ or } 7) = \left(\tfrac{5}{36}\right)\left(\tfrac{5}{5+6}\right) = \tfrac{25}{36\cdot 11} = \tfrac{25}{396}$$

$$p_8 = \Pr(8)\Pr(8|8 \text{ or } 7) = \left(\tfrac{5}{36}\right)\left(\tfrac{5}{5+6}\right) = \tfrac{25}{36\cdot 11} = \tfrac{25}{396}$$

$$p_9 = \Pr(9)\Pr(9|9 \text{ or } 7) = \left(\tfrac{4}{36}\right)\left(\tfrac{4}{4+6}\right) = \tfrac{16}{36\cdot 10} = \tfrac{2}{45}$$

$$p_{10} = \Pr(10)\Pr(10|10 \text{ or } 7) = \left(\tfrac{3}{36}\right)\left(\tfrac{3}{3+6}\right) = \tfrac{9}{36\cdot 9} = \tfrac{1}{36}$$

Thus the probability of winning is $p_0 + p_4 + p_5 + p_6 + p_8 + p_9 + p_{10} = \frac{244}{495} \approx .4930$.