1 Conditional Probability

In this section, we are interested in answering this type of question: how the information “an event $B$ has occurred” affects the probability that “event $A$ occurs”.

Real-life examples:

- $A$ refers to the event that an individual having a particular disease. $B$ refers to the event that the blood test result is positive. Then the probability of $A$ occurs be affected by the information on whether $B$ occurs.

- An experiment consists of rolling a die once. Let $E = \{5, 6\}$ and $F = \{6\}$. If the die is fair, we have $P(F) = 1/6$. Now suppose that the die is rolled and we are told that the event $E$ has occurred. This leaves only two possible outcomes: 5 and 6. In the absence of any other information, we would still regard these outcomes to be equally likely, so the probability of $F$ becomes 1/2, making $P(F|E) = 1/2$.

- Voting. Three candidates A, B, and C are running for office. We decided that A and B have an equal chance of winning and C is only 1/2 as likely to win as A. Let $A$ be the event “A wins”, $B$ that “wins” and $C$ that “C wins.” Hence, we assigned probabilities

$$P(A) = 2/5, P(B) = 2/5, P(C) = 1/5.$$  

Suppose that before the election is held, A drops out of the race. It would be natural to assign new probabilities to the events $B$ and $C$ which are proportional to the original probabilities. Thus, we would have $P(B|A^c) = 2/3$, and $P(C|A^c) = 1/3$. 
General definition of conditional probability: For any two events $E$ and $F$ with $P(F) > 0$, the conditional probability of $E$ given that $F$ has occurred is defined by

$$P(E|F) = \frac{P(E \cap F)}{P(F)}.$$  

Idea behind definition: Given that $B$ has occurred, the relevant sample space becomes $B$ rather than $S$.

**Example 1** A couple has two children. What is the probability that both are girls given that the oldest child is a girl? What is the probability that both are girls given that at least one of them is a girl? Assume that the four possible outcomes – (younger is boy, older is girl), (younger is boy, older is girl), (younger is girl, older is girl), (younger is girl, older is boy) - are equally likely.
The multiplication rule: From the definition of conditional probability,

\[ P(E|F) = \frac{P(E \cap F)}{P(F)}, \]

we have that

(1.2) \[ P(E \cap F) = P(E|F)P(F). \]

This is a useful formula for computing \( P(E \cap F) \) when \( P(E|F) \) is easy to compute. Analogously, we have \( P(E \cap F) = P(F|E)P(E) \).

For three events \( E, F \) and \( G \), by thinking of \( E \cap F \) as a single event, say, \( H \), we can write

\[
P(E \cap F \cap G) = P(H \cap G) = P(G|H)P(H) = P(G|E \cap F)P(E \cap F) = P(G|E \cap F)P(F|E)P(E)
\]

Repeating the same argument for \( n \) events, we have

\[
P(E_1 \cap E_2 \cap \cdots \cap E_n) = P(E_n|E_{n-1}, \ldots, E_1)P(E_{n-1}|E_{n-2}, \ldots, E_1) \cdots P(E_2|E_1)P(E_1).
\]

**Example 2** An urn contains 5 white chips, 4 black chips and 3 red chips. Four chips are drawn sequentially and without replacement. What is the probability of obtaining the sequence (white, red, white, black)?
2 Real Life Applications

Example 3 The following question appeared in the ”Ask Marilyn” column in Parade Magazine in 1991 and evoked great controversy:

Suppose you’re on a game show, and you’re given the choice of three doors. Behind one door is a car, behind the others, goats. You pick a door, say number 1, and the host, who knows what’s behind the doors, opens another door, say number 3, which has a goat. He says to you, “Do you want to pick door number 2?” Is it to your advantage to switch your choice of doors? Why?

By Craig F. Whitaker
Columbia, MD.

The problem, often referred to as the Monte Hall Problem, arose in the 1970s game show Let’s Make a Deal hosted by Monte Hall.
Example 4  *Three Prisoner’s Paradox.*  Three prisoners, A, B and C are on death row. The governor decides to pardon one of the three and chooses at random the prisoner to pardon. He informs the warden of his choice but requests that the name be kept secret for a few days. The next day, A tries to get the warden to tell him who had been pardoned. The warden refuses. A then asks which of B or C will be executed. The warden thinks for a while, then tells A that B is to be executed.

Warden’s Reasoning: Clearly either B or C must be executed, so I give A no information about whether A will be pardoned.

A’s reasoning: Given that B will be executed, then either A or C will be pardoned. My chance of being pardoned has risen from 1/3 to 1/2.

Which reasoning is correct?
3 The Law of Total Probability

The law of total probability is useful when it is easier to compute conditional probabilities than to compute the probability of an event directly.

The law of total probability: Let $F_1, \cdots, F_n$ be disjoint events (i.e., $F_i \cap F_j = \emptyset$ for $i \neq j$) such that the sample space $S = \cup_{i=1}^n F_i$ and $P(F_i) > 0$ for $i = 1, \cdots, n$. Then, for any event $E$,

$$(3.3) \quad P(E) = \sum_{i=1}^n P(E|F_i)P(F_i).$$

Statement of law of total probability in words: The events $F_1, \cdots, F_n$ are disjoint events that partition the sample space $S$. The law of total probability states that to find the probability of $E$, we take a weighted average of the conditional probabilities of $P(E|F_i)$, weighted by $P(F_i)$.

Proof. Note that $S = \cup_{i=1}^n F_i$, we have

$$P(E) = P\{\cup_{i=1}^n (E \cap F_i)\} = \sum_{i=1}^n P(E \cap F_i) = \sum_{i=1}^n P(E|F_i)P(F_i)$$

The second step uses the fact that $F_1, \cdots, F_n$ are disjoint. The last step is due to the multiplication rule.
**Example 5** An urn contains three red balls and one black ball. Two balls are selected without replacement. What is the probability that a red ball is selected on the second draw?

**Example 6** We consider a model that has been used to study occupational mobility. Suppose that occupations are grouped into upper (U), middle (M) and lower (L) levels. $U_1$ will denote the event that a father’s occupation is upper-level; $U_2$ will denote the event that a son’s occupation is upper-level, etc. (the subscripts index generations). Glass and Hall (1954) compiled the following statistics on occupational mobility in England and Wales:

<table>
<thead>
<tr>
<th></th>
<th>$U_2$</th>
<th>$M_2$</th>
<th>$L_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$U_1$</td>
<td>0.45</td>
<td>0.48</td>
<td>0.07</td>
</tr>
<tr>
<td>$M_1$</td>
<td>0.05</td>
<td>0.70</td>
<td>0.25</td>
</tr>
<tr>
<td>$L_1$</td>
<td>0.01</td>
<td>0.50</td>
<td>0.49</td>
</tr>
</tbody>
</table>
Such a table, which is called a matrix of transition probabilities is to be read in the following way: If a father is in $U$, the probability that his son is in $U$ is .45, the probability that his son is in $M$ is .48 etc. The table thus gives conditional probabilities; for example $P(U_2|U_1) = 0.45$.

Question: Suppose that of the father’s generation, 10% are in $U$, 40% are in $M$ and 50% are in $L$. What is the probability that a son in the next generation is in $U$?
4 Bayes Formula

Continuing with the previous example, suppose we ask a different question: If a son has occupational status $U_2$, what is the probability that his father had occupational status $U_1$? Compared to the question asked in Example 2, this is an “inverse” problem: we are given an “effect” and are asked to find the probability of a particular “cause”. In situations like this, Bayes’ formula is useful.

The goal is to find $P(U_1|U_2)$. By definition,

$$P(U_1|U_2) = \frac{P(U_1 \cap U_2)}{P(U_2)} = \frac{P(U_2|U_1)P(U_1)}{P(U_2|U_1)P(U_1) + P(U_2|M_1)P(M_1) + P(U_2|L_1)P(L_1)}$$

Here, we used the multiplication rule to re-express the numerator and the law of total probability to re-express the denominator. The value of the numerator is

$$P(U_2|U_1)P(U_1) = 0.45 \times 0.10 = 0.045,$$

and we calculated the denominator in Example 2 to be 0.07, so we find that $P(U_1|U_2) = 0.045/0.07 = 0.64$. In other words, 64% of the sons who are in upper-level occupations have fathers who were in upper-level occupations.

Bayes’ Formula:

Let $F_1, \cdots, F_n$ be disjoint events such that the sample space $S = \bigcup_{i=1}^{n}$ and $P(F_i) > 0$ for $i = 1, \cdots, n$. Then, for any event $E$,

$$P(F_j|E) = \frac{P(E \cap F_j)}{P(E)} = \frac{P(E|F_j)P(F_j)}{\sum_{i=1}^{n} P(E|F_i)P(F_i)}$$
Example 7 Consider the problem of screening for breast cancer. A doctor discovers a lump in a woman’s breast during a routine physical exam. The lump could be a cancer. Without performing any further tests, the probability that the lump is a cancer is 0.01. A mammogram is a test that, on average, is correctly able to establish whether a lump is benign or cancerous 90% of the time. What is the probability that the lump is cancerous if the test result from a mammogram is positive?
Example 8 It rains in Raleigh area 30% of the time in October. If it is raining, then Tommy will wear his hat with probability 0.8. If it is not raining, then Tommy will wear his hat with probability 0.2. If Tommy is seen wearing his hat, what is the probability that it is raining?
Example 9 (please study after class) Digitalis therapy is often beneficial to patients who have suffered congestive heart failure, but there is the risk of digitalis intoxication, a serious side effect, that is moreover difficult to diagnose. To improve the chances of a correct diagnosis, the concentration of digitalis in the blood can be measured. Let $T^+ = \text{patient has high blood concentration of digitalis (positive test)}$ $T^- = \text{patient has low blood concentration of digitalis (negative test)}$ $D^+ = \text{patient will suffer digitalis intoxication if treated with digitalis therapy (disease present)}$ $D^- = \text{patient will not suffer digitalis intoxication if treated with digitalis therapy (disease absent)}$. The probability of a randomly selected patient having each possible combination of $D,T$ are the following:

\[
\begin{align*}
P(T^+ \cap D^+) &= 0.185 \\
P(T^+ \cap D^-) &= 0.104 \\
P(T^- \cap D^+) &= 0.133 \\
P(T^- \cap D^-) &= 0.578 \\
\end{align*}
\]

The probability that a random selected patient has $D^+$ (so that digitalis therapy should not be used) is $0.185 + 0.133 = 0.318$.

What is the probability that a randomly selected patient has $D^+$ given that the randomly selected patient has a positive test, $T^+$?

The proportion of patients with $T^+$ who have $D^+$ is

\[
P(D^+ | T^+) = \frac{P(T^+ \cap D^+)}{P(T^+)} = \frac{0.185}{0.185 + 0.104} = 0.640.
\]

What if the test is negative?

The proportion of patients with $T^-$ who have $D^+$ is

\[
P(D^+ | T^-) = \frac{P(T^- \cap D^+)}{P(T^-)} = \frac{0.133}{0.133 + 0.578} = 0.187.
\]
5 Independent Events

In general, $P(E|F)$, the conditional probability of $E$ given $F$, is not equal to $P(E)$, the unconditional probability of $E$. In other words, knowing that $F$ has occurred generally changes the chances of $E$’s occurrence. In the special cases where $P(E|F)$ does in fact equal $P(E)$, we say that $E$ is independent of $F$. That is, $E$ is independent of $F$ if knowledge that $F$ has occurred does not change the probability that $E$ occurs.

Since $P(E|F) = P(E \cap F)/P(F)$, we see that $E$ is independent of $F$ if and only if $P(E \cap F) = P(E)P(F)$. We thus have the following definition:

Two events $E$ and $F$ are said to be independent if

(5.5) \[ P(E \cap F) = P(E)P(F). \]

Two events $E$ and $F$ that are not independent are said to be dependent. Some examples:

- A card is selected at random from an ordinary deck of 52 playing cards. Let $E$ be the event that the selected card is an ace and $F$ be the event that it is a spade. Are $E$ and $F$ independent?

- Suppose that we toss two fair dice (green and red). Let $E$ be the event that the sum of the two dice is 6 and $F$ be the event that the green die equals 4. Are $E$ and $F$ independent?

If event $E$ is independent of event $F$, then $E$ is also independent of $F^c$.

Proof. Assume that $E$ is independent of $F$. Since $E = (E \cap F) \cup (E \cap F^c)$ and $E \cap F$ and $E \cap F^c$ are disjoint, we have that

\[ P(E) = P(E \cap F) + P(E \cap F^c) = P(E)P(F) + P(E \cap F^c), \]

or equivalently

\[ P(E \cap F^c) = P(E)[1 - P(F)] = P(E)P(F^c). \]
By similar reasoning, it follows that if \( E \) is independent of \( F \), then (i) \( E^c \) is independent of \( F \) and (ii) \( E^c \) is independent of \( F^c \).

**Independence of more than two events**

We define events \( E_1, \ldots, E_n \) to be mutually independent if

\[
P(E_{i_1} \cap E_{i_2} \cap \cdots \cap E_{i_r}) = P(E_{i_1})P(E_{i_2}) \cdots P(E_{i_r})
\]

for every subset \( E_{i_1}, \ldots, E_{i_r} \). If \( E_1, \ldots, E_n \) are mutually independent, then knowing that some subset of the events \( E_{i_1}, \ldots, E_{i_r} \) has occurred does not affect the probability that an event \( E_j \) has occurred where \( j \neq i_1, \ldots, i_r \).

**Example 10** An executive on a business trip must rent a car in each of three different cities. Let \( A \) denote the event that the executive is offered a free upgrade in the first city, and \( B \) and \( C \) represent the analogous event for the second and third cities. Suppose that \( P(A) = .3 \), \( P(B) = .4 \), and \( P(C) = .5 \) and that \( A, B \) and \( C \) are independent events.

- What is the probability that the executive is offered a free upgrade in all three cities?
- What is the probability that the executive is offered a free upgrade in at least one of the three cities?
Consider three events $E, F$ and $G$. Does pairwise independence of the events (i.e., $E$ is independent of $F$, $E$ is independent of $G$ and $F$ is independent of $G$) guarantee mutual independence? The answer is no.

**Group project (each group hand in one solution):**

- Find an example that two events are disjoint but dependent.
- Find an example that three events are pairwise independent but not mutually independent.

*Hints: For the first problem, consider $E$ and $E^c$. For the second problem, if your group cannot find a good example, use the following one. A fair coin is tossed twice. Let $E$ denote the event of heads on the first toss, $F$ denote the event of heads on the second toss, and $G$ denote the event that exactly one head is thrown. Verify that $E, F$ and $G$ are pairwise independent but that $P(G|E \cap F) \neq P(G)$.**
A lesson from real life: be careful about assuming events are independent without justification.

Example 11 *People vs. Collins.* In 1964, a woman shopping in LA had her purse snatched by a young, blond female wearing a ponytail. The thief fled on foot but was seen shortly thereafter getting into a yellow automobile driven by an African American male who had a mustache and a beard. A police investigation subsequently turned up a suspect, one Janet Collins, who was blond, wore a ponytail and associated with an African American male who drove a yellow car and had a mustache. An arrest was made.

Not having any tangible evidence, and no reliable witnesses, the prosecutor sought to build his case on the unlikelihood of Ms. Collins and her companion sharing these characteristics and not being the guilty parties. First, the bits of evidence that were available were assigned probabilities. It was estimated, for example, that the probability of a female wearing a ponytail in Los Angeles was 1/10. The following are the probabilities quoted for the six “facts” agreed on by the victim and eyewitnesses:

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yellow automobile</td>
<td>1/10</td>
</tr>
<tr>
<td>Man with a mustache</td>
<td>1/4</td>
</tr>
<tr>
<td>Woman with a ponytail</td>
<td>1/10</td>
</tr>
<tr>
<td>Woman with blond hair</td>
<td>1/3</td>
</tr>
<tr>
<td>Black man with a beard</td>
<td>1/10</td>
</tr>
<tr>
<td>Interracial couple in car</td>
<td>1/1000</td>
</tr>
</tbody>
</table>

The prosecutor multiplied these numbers together and claimed that the product $\frac{1}{10} \times \frac{1}{4} \times \frac{1}{10} \times \frac{1}{3} \times \frac{1}{10} \times \frac{1}{1000}$, or 1 in 12 million was the probability of the intersection - that is, the probability that a random couple would fit this description. A probability of 1 in 12 million is so small, he argued, that the only reasonable decision is to find the defendants guilty. The jury agreed, and handed down a verdict of second-degree robbery. Later though, the Supreme Court of California disagreed. Ruling on an appeal, the higher court reversed the decision, claiming that the probability argument was incorrect and misleading. What do you think?
Repeated Independent Trials

A common setup in real life is that the overall probability experiment consists of a sequence of identical, independent sub-experiments.

Example 12 Consider a system of five identical components connected in series. Denote a component that fails by $F$ and one does not fail by $S$.

- A **series system** fails if any of the individual components fails.
- A **parallel system** fails if all individual components fail.

Denote by $A$ the event that a series system fails and $B$ the event that a parallel system fails. Let $p$ be the probability that one component fails, $i = 1, \ldots, 5$. Assume that different components fail independently with one another. What is $P(A)$ and $P(B)$?
Example 13 During the 1978 baseball season, Pete Rose of the Cincinnati Reds set a National League record by hitting safely in 44 consecutive games. Assume that Rose is a .300 hitter and that he comes to bat four times each game. If Rose’s at bats are assumed to be identical, independent experiments, what is the probability that Rose would get a hit in all 44 games of a set of 44 games?