NOTATION

$$
\begin{aligned}
P_r(n,k) &= \text{number of permutations with replacement} \\
P(n,k) &= \text{number of permutations without replacement} \\
C_r(n,k) &= \text{number of combinations with replacement} \\
\left.\begin{array}{c} C(n,k) \\ \binom{n}{k} \end{array}\right\} &= \text{number of combinations without replacement}
\end{aligned}
$$

## 2.4 Conditional probability and independence

### Conditional probability

Suppose you are meeting someone at an airport. The flight is likely to arrive on time; the probability of that is 0.8. Suddenly it is announced that the flight departed one hour behind the schedule. Now it has the probability of only 0.05 to arrive on time. New information affected the probability of meeting this flight on time. The new probability is called *conditional probability*, where the new information, that the flight departed late, is a *condition*.

*DEFINITION 2.15*

> **Conditional probability** of event $A$ given event $B$ is the probability that $A$ occurs when $B$ is *known to occur*.

NOTATION $\boldsymbol{P}\{A \mid B\}$ = conditional probability of $A$ given $B$

How does one compute the conditional probability? First, consider the case of equally likely outcomes. In view of the new information, occurrence of the condition $B$, only the outcomes contained in $B$ still have a non-zero chance to occur. Counting only such outcomes, the *unconditional probability* of $A$,

$$\boldsymbol{P}\{A\} = \frac{\text{number of outcomes in } A}{\text{number of outcomes in } \Omega},$$

is now replaced by the *conditional probability* of $A$ given $B$,

$$\boldsymbol{P}\{A \mid B\} = \frac{\text{number of outcomes in } A \cap B}{\text{number of outcomes in } B} = \frac{\boldsymbol{P}\{A \cap B\}}{\boldsymbol{P}\{B\}}.$$

This appears to be the general formula.

**Conditional probability**

$$\boxed{\boldsymbol{P}\{A \mid B\} = \frac{\boldsymbol{P}\{A \cap B\}}{\boldsymbol{P}\{B\}}} \qquad (2.7)$$

Rewriting (2.7) in a different way, we obtain the general formula for the probability of intersection.

**Intersection, general case**

$$\boldsymbol{P}\{A \cap B\} = \boldsymbol{P}\{B\}\,\boldsymbol{P}\{A \mid B\} \qquad (2.8)$$

### Independence

Now we can give an intuitively very clear definition of *independence*.

*DEFINITION 2.16*

Events $A$ and $B$ are **independent** if occurrence of $B$ does not affect the probability of $A$, i.e.,

$$\boldsymbol{P}\{A \mid B\} = \boldsymbol{P}\{A\}.$$

According to this definition, *conditional* probability equals *unconditional* probability in case of independent events. Substituting this into (2.8) yields

$$\boldsymbol{P}\{A \cap B\} = \boldsymbol{P}\{A\}\,\boldsymbol{P}\{B\}.$$

This is our old formula for independent events.

**Example 2.31.** Ninety percent of flights depart on time. Eighty percent of flights arrive on time. Seventy-five percent of flights depart on time and arrive on time.

(a) Eric is meeting Alyssa's flight, which departed on time. What is the probability that Alyssa will arrive on time?

(b) Eric has met Alyssa, and she arrived on time. What is the probability that her flight departed on time?

(c) Are the events, departing on time and arriving on time, independent?

Solution. Denote the events,

$$\begin{aligned} A &= \{\text{arriving on time}\}, \\ D &= \{\text{departing on time}\}. \end{aligned}$$

We have:

$$\boldsymbol{P}\{A\} = 0.8, \quad \boldsymbol{P}\{D\} = 0.9, \quad \boldsymbol{P}\{A \cap D\} = 0.75.$$

(a) $\boldsymbol{P}\{A \mid D\} = \dfrac{\boldsymbol{P}\{A \cap D\}}{\boldsymbol{P}\{D\}} = \dfrac{0.75}{0.9} = \underline{0.8333}$.

(b) $\boldsymbol{P}\{D \mid A\} = \dfrac{\boldsymbol{P}\{A \cap D\}}{\boldsymbol{P}\{A\}} = \dfrac{0.75}{0.8} = \underline{0.9375}$.

(c) Events are not independent because

$$\boldsymbol{P}\{A \mid D\} \neq \boldsymbol{P}\{A\}, \quad \boldsymbol{P}\{D \mid A\} \neq \boldsymbol{P}\{D\}, \quad \boldsymbol{P}\{A \cap D\} \neq \boldsymbol{P}\{A\}\,\boldsymbol{P}\{D\}.$$

Actually, any one of these inequalities is sufficient to prove that $A$ and $D$ are dependent. Further, we see that $\boldsymbol{P}\{A \mid D\} > \boldsymbol{P}\{A\}$ and $\boldsymbol{P}\{D \mid A\} > \boldsymbol{P}\{D\}$. In other words, departing on time increases the probability of arriving on time, and vice versa. This perfectly agrees with our intuition. ♢

### Bayes Rule

The last example shows that two conditional probabilities, $\boldsymbol{P}\{A \mid B\}$ and $\boldsymbol{P}\{B \mid A\}$, are not the same, in general. Consider another example.

**Example 2.32** (Reliability of a test). There exists a test for a certain viral infection (including a virus attack on a computer network). It is 95% reliable for infected patients and 99% reliable for the healthy ones. That is, if a patient has the virus (event $V$), the test shows that (event $S$) with probability $\boldsymbol{P}\{S \mid V\} = 0.95$, and if the patient does not have the virus, the test shows that with probability $\boldsymbol{P}\left\{\overline{S} \mid \overline{V}\right\} = 0.99$.

Consider a patient whose test result is positive (i.e., the test shows that the patient has the virus). Knowing that sometimes the test is wrong, naturally, the patient is eager to know the probability that he or she indeed has the virus. However, this conditional probability, $\boldsymbol{P}\{V \mid S\}$, is not stated among the given characteristics of this test. ♢

This example is applicable to any testing procedure including software and hardware tests, pregnancy tests, paternity tests, alcohol tests, academic exams, etc. The problem is to connect the given $\boldsymbol{P}\{S \mid V\}$ and the quantity in question, $\boldsymbol{P}\{V \mid S\}$. This was done in the eighteenth century by English minister *Thomas Bayes* (1702–1761) in the following way.

Notice that $A \cap B = B \cap A$. Therefore, using (2.8), $\boldsymbol{P}\{B\}\,\boldsymbol{P}\{A \mid B\} = \boldsymbol{P}\{A\}\,\boldsymbol{P}\{B \mid A\}$.

Solve for $\boldsymbol{P}\{B \mid A\}$ to obtain

**Bayes Rule**

$$\boldsymbol{P}\{B \mid A\} = \frac{\boldsymbol{P}\{A \mid B\}\,\boldsymbol{P}\{B\}}{\boldsymbol{P}\{A\}} \tag{2.9}$$

**Example 2.33** (Situation on a midterm exam). On a midterm exam, students $X$, $Y$, and $Z$ forgot to sign their papers. Professor knows that they can write a good exam with probabilities 0.8, 0.7, and 0.5, respectively. After the grading, he notices that two unsigned exams are good and one is bad. Given this information, and assuming that students worked independently of each other, what is the probability that the bad exam belongs to student $Z$?

Solution. Denote good and bad exams by $G$ and $B$. Also, let $GGB$ denote two good and one bad exams, $XG$ denote the event "student $X$ wrote a good exam," etc. We need to find

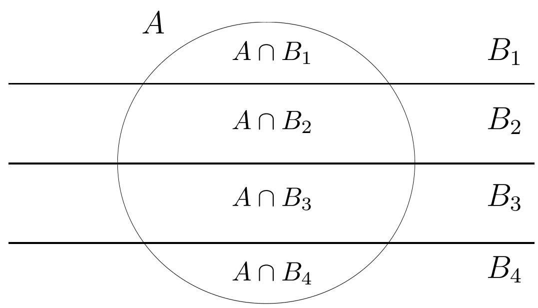


FIGURE 2.6: *Partition of the sample space $\Omega$ and the event $A$.*

$\boldsymbol{P}\{ZB \mid GGB\}$ given that $\boldsymbol{P}\{G \mid X\} = 0.8$, $\boldsymbol{P}\{G \mid Y\} = 0.7$, and $\boldsymbol{P}\{G \mid Z\} = 0.5$.

By the *Bayes Rule*,

$$\boldsymbol{P}\{ZB \mid GGB\} = \frac{\boldsymbol{P}\{GGB \mid ZB\}\,\boldsymbol{P}\{ZB\}}{\boldsymbol{P}\{GGB\}}.$$

Given $ZB$, event $GGB$ occurs only when both $X$ and $Y$ write good exams. Thus, $\boldsymbol{P}\{GGB \mid ZB\} = (0.8)(0.7)$.

Event $GGB$ consists of three outcomes depending on the student who wrote the bad exam. Adding their probabilities, we get

$$\begin{aligned} &\boldsymbol{P}\{GGB\} \\ &= \boldsymbol{P}\{XG \cap YG \cap ZB\} + \boldsymbol{P}\{XG \cap YB \cap ZG\} + \boldsymbol{P}\{XB \cap YG \cap ZG\} \\ &= (0.8)(0.7)(0.5) + (0.8)(0.3)(0.5) + (0.2)(0.7)(0.5) = 0.47. \end{aligned}$$

Then

$$\boldsymbol{P}\{ZB \mid GGB\} = \frac{(0.8)(0.7)(0.5)}{0.47} = \underline{0.5957}.$$

◊

In the Bayes Rule (2.9), the denominator is often computed by the Law of Total Probability.

**Law of Total Probability**

This law relates the unconditional probability of an event $A$ with its conditional probabilities. It is used every time when it is easier to compute conditional probabilities of $A$ given additional information.

Consider some partition of the sample space $\Omega$ with mutually exclusive and exhaustive events $B_1, \ldots, B_k$. It means that

$$B_i \cap B_j = \varnothing \text{ for any } i \neq j \ \text{ and } \ B_1 \cup \ldots \cup B_k = \Omega.$$

These events also partition the event $A$,

$$A = (A \cap B_1) \cup \ldots \cup (A \cap B_k),$$

and this is also a union of mutually exclusive events (Figure 2.6). Hence,

$$\boldsymbol{P}\{A\} = \sum_{j=1}^{k} \boldsymbol{P}\{A \cap B_j\},$$

and we arrive to the following rule.

**Law of Total Probability**

$$\boldsymbol{P}\{A\} = \sum_{j=1}^{k} \boldsymbol{P}\{A \mid B_j\}\, \boldsymbol{P}\{B_j\}$$

In case of two events ($k = 2$),

$$\boldsymbol{P}\{A\} = \boldsymbol{P}\{A \mid B\}\, \boldsymbol{P}\{B\} + \boldsymbol{P}\{A \mid \overline{B}\}\, \boldsymbol{P}\{\overline{B}\} \tag{2.10}$$

Together with the Bayes Rule, it makes the following popular formula

**Bayes Rule for two events**

$$\boldsymbol{P}\{B \mid A\} = \frac{\boldsymbol{P}\{A \mid B\}\, \boldsymbol{P}\{B\}}{\boldsymbol{P}\{A \mid B\}\, \boldsymbol{P}\{B\} + \boldsymbol{P}\{A \mid \overline{B}\}\, \boldsymbol{P}\{\overline{B}\}}$$

**Example 2.34** (Reliability of a test, continued). Continue Example 2.32. Suppose that 4% of all the patients are infected with the virus, $\boldsymbol{P}\{V\} = 0.04$. Recall that $\boldsymbol{P}\{S \mid V\} = 0.95$ and $\boldsymbol{P}\{\overline{S} \mid \overline{V}\} = 0.99$. If the test shows positive results, the (conditional) probability that a patient has the virus equals

$$\begin{aligned} \boldsymbol{P}\{V \mid S\} &= \frac{\boldsymbol{P}\{S \mid V\}\, \boldsymbol{P}\{V\}}{\boldsymbol{P}\{S \mid V\}\, \boldsymbol{P}\{V\} + \boldsymbol{P}\{S \mid \overline{V}\}\, \boldsymbol{P}\{\overline{V}\}} \\ &= \frac{(0.95)(0.04)}{(0.95)(0.04) + (1 - 0.99)(1 - 0.04)} = \underline{0.7983}. \end{aligned}$$

◊

**Example 2.35** (Diagnostics of computer codes). A new computer program consists of two modules. The first module contains an error with probability 0.2. The second module is more complex; it has a probability of 0.4 to contain an error, independently of the first module. An error in the first module alone causes the program to crash with probability 0.5. For the second module, this probability is 0.8. If there are errors in both modules, the program crashes with probability 0.9. Suppose the program crashed. What is the probability of errors in both modules?

Solution. Denote the events,

$$A = \{\text{errors in module I}\}, \quad B = \{\text{errors in module II}\}, \quad C = \{\text{crash}\}.$$

Further,

$$\begin{aligned}\{\text{errors in module I alone}\} &= A\backslash B = A\backslash(A\cap B)\\ \{\text{errors in module II alone}\} &= B\backslash A = B\backslash(A\cap B).\end{aligned}$$

It is given that $\boldsymbol{P}\{A\} = 0.2$, $\boldsymbol{P}\{B\} = 0.4$, $\boldsymbol{P}\{A\cap B\} = (0.2)(0.4) = 0.08$, by independence, $\boldsymbol{P}\{C \mid A\backslash B\} = 0.5$, $\boldsymbol{P}\{C \mid B\backslash A\} = 0.8$, and $\boldsymbol{P}\{C \mid A\cap B\} = 0.9$.

We need to compute $\boldsymbol{P}\{A\cap B \mid C\}$. Since $A$ is a union of disjoint events $A\backslash B$ and $A\cap B$, we compute

$$\boldsymbol{P}\{A\backslash B\} = \boldsymbol{P}\{A\} - \boldsymbol{P}\{A\cap B\} = 0.2 - 0.08 = 0.12.$$

Similarly,

$$\boldsymbol{P}\{B\backslash A\} = 0.4 - 0.08 = 0.32.$$

Events $(A\backslash B)$, $(B\backslash A)$, $A\cap B$, and $\overline{(A\cup B)}$ form a partition of $\Omega$, because they are mutually exclusive and exhaustive. The last of them is the event of no errors in the entire program. Given this event, the probability of a crash is 0. Notice that $A$, $B$, and $(A\cap B)$ are neither mutually exclusive nor exhaustive, so they cannot be used for the Bayes Rule. Now organize the data.

| Location of errors | Probability of a crash |
|---|---|
| $\boldsymbol{P}\{A\backslash B\} = 0.12$ | $\boldsymbol{P}\{C \mid A\backslash B\} = 0.5$ |
| $\boldsymbol{P}\{B\backslash A\} = 0.32$ | $\boldsymbol{P}\{C \mid B\backslash A\} = 0.8$ |
| $\boldsymbol{P}\{A\cap B\} = 0.08$ | $\boldsymbol{P}\{C \mid A\cap B\} = 0.9$ |
| $\boldsymbol{P}\{\overline{A\cup B}\} = 0.48$ | $\boldsymbol{P}\{C \mid \overline{A\cup B}\} = 0$ |

Combining the Bayes Rule and the Law of Total Probability,

$$\boldsymbol{P}\{A\cap B \mid C\} = \frac{\boldsymbol{P}\{C \mid A\cap B\}\,\boldsymbol{P}\{A\cap B\}}{\boldsymbol{P}\{C\}},$$

where

$$\begin{aligned}\boldsymbol{P}\{C\} &= \boldsymbol{P}\{C \mid A\backslash B\}\,\boldsymbol{P}\{A\backslash B\} + \boldsymbol{P}\{C \mid B\backslash A\}\,\boldsymbol{P}\{B\backslash A\}\\ &\quad + \boldsymbol{P}\{C \mid A\cap B\}\,\boldsymbol{P}\{A\cap B\} + \boldsymbol{P}\{C \mid \overline{A\cup B}\}\,\boldsymbol{P}\{\overline{A\cup B}\}.\end{aligned}$$

Then

$$\boldsymbol{P}\{A\cap B \mid C\} = \frac{(0.9)(0.08)}{(0.5)(0.12)+(0.8)(0.32)+(0.9)(0.08)+0} = \underline{0.1856}.$$

◊

## Summary and conclusions

Probability of any event is a number between 0 and 1. The empty event has probability 0, and the sample space has probability 1. There are rules for computing probabilities of