

CENG 222 - Chapter 2: Probability

Burak Metehan Tunçel - March 2022

1 Events and Their Probabilities

The concept of *probability* perfectly agrees with our intuition. In everyday life, probability of an event is understood as a *chance* that this event will happen.

1.1 Outcomes, Events, and the Sample Space

Definition 1

A collection of all (*observable*) **outcomes**, or elementary results, of an experiment, is called a **sample space**. If an outcome is not observable, it should not be in sample space.

Note: Outcome is the result of certain random experiment.

Definition 2

Any set of outcomes is an **event**. Thus, events are *subsets of the sample space*. A sample space of N possible outcomes yields 2^N possible events.

Note: Events are basically saying “either this or this will happen”.

Example 1

A tossed die can produce one of 6 possible outcomes: 1 dot through 6 dots. *Each outcome is an event*. There are other events: observing an even number of dots, an odd number of dots, a number of dots less than 3, etc.

Example 2

Consider a football game between the Dallas Cowboys and the New York Giants. The sample space consists of 3 outcomes,

$$\Omega = \{\text{Cowboys win, Giants win, they tie}\}$$

Combining these outcomes in all possible ways, we obtain the following $2^3 = 8$ events: Cowboys win, lose, tie, get at least a tie, get at most a tie, no tie, get some result, and get no result.

The event “*some result*” is the entire sample space Ω , and by common sense, it should have *probability 1*. The event “*no result*” is empty, it does not contain any outcomes, so its *probability is 0*.

NOTATION

Ω = sample space

\emptyset = empty event

$P\{E\}$ = probability of event E

1.2 Set Operations

Events are *sets of outcomes*.

- A **union** of events A, B, C, \dots is an event *consisting of all the outcomes in all these events*. It occurs if any of A, B, C, \dots occurs, and therefore, corresponds to the word “**or**”: A or B or C or
- An **intersection** of events A, B, C, \dots is an event consisting of outcomes that are *common* in all these events. It occurs if each A, B, C, \dots occurs, and therefore, corresponds to the word “**and**”: A and B and C and
- A **complement** of an event A is an event that occurs every time *when A does not occur*. It consists of outcomes excluded from A , and therefore, corresponds to the word “**not**”: not A .
- A **difference** of events A and B consists of all outcomes included *in A but excluded from B* . It occurs *when A occurs and B does not*, and corresponds to “**but not**”: A but not B .

NOTATION

$A \cup B$ = union

$A \cap B$ = intersection

\bar{A} or A^c = complement

$A \setminus B$ = difference

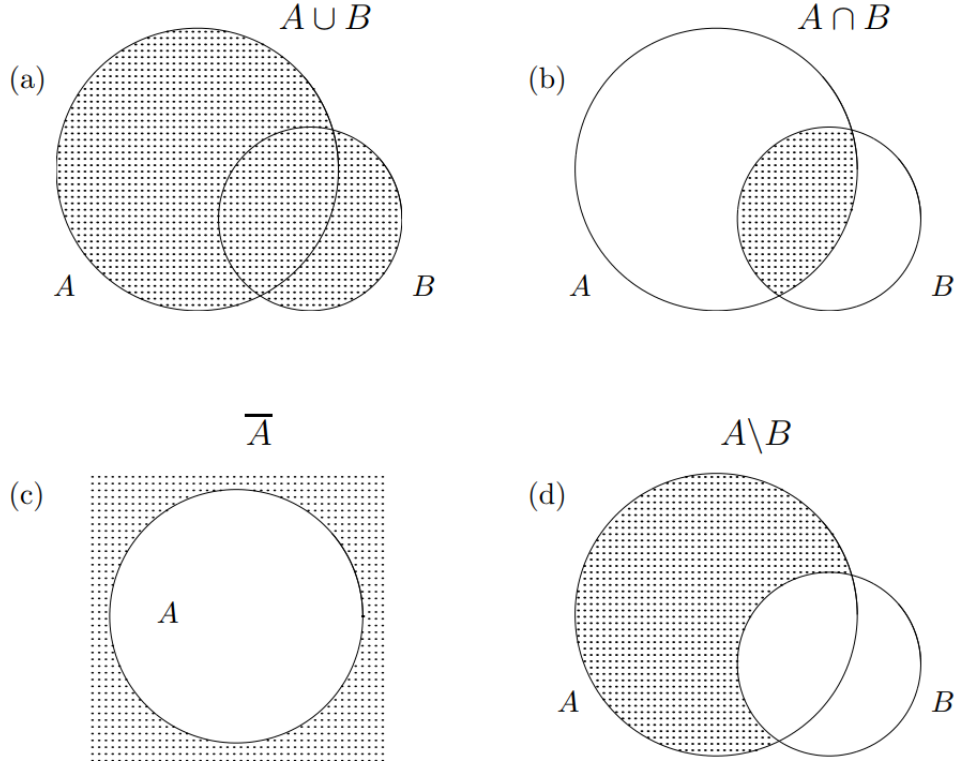


Figure 1: Venn diagrams for union, intersection, complement, and difference of events.

Definition 3

Events A and B are **disjoint** if their *intersection is empty*,

$$A \cap B = \emptyset$$

Events A_1, A_2, A_3, \dots are **mutually exclusive** or **pairwise disjoint** if any two of these events are disjoint, i.e.,

$$A_i \cap A_j = \emptyset \text{ for any } i \neq j$$

If $E_1 \cap E_2 = \emptyset$, E_1 and E_2 are distinct, disjoint or mutually exclusive, events. One cannot observe both events at the same time, they do not have common outcome.

On the other side, if $E_1 \cap E_2 \neq \emptyset$, E_1 and E_2 are not distinct events. One can observe *both events at the same time*.

Definition 4

Events A, B, C, \dots are **exhaustive** if their union equals the whole sample space, i.e.,

$$A \cup B \cup C \cup \dots = \Omega$$

- **Mutually exclusive** events will *never occur at the same time*. Occurrence of any one of them eliminates the possibility for all the others to occur.
- **Exhaustive events** *cover the entire Ω* , so that “*there is nothing left*.” In other words, among any collection of exhaustive events, at least one occurs for sure. Also, *they do not have to be disjoint or single outcome events. Only union giving sample space enough for being exhaustive events.*

It is often easier to compute probability of an intersection than probability of a union. Taking complements converts unions into intersections.

$$\overline{E_1 \cup \dots \cup E_n} = \overline{E_1} \cap \dots \cap \overline{E_n}$$

$$\overline{E_1 \cap \dots \cap E_n} = \overline{E_1} \cup \dots \cup \overline{E_n}$$

Proof. Since the union $E_1 \cup \dots \cup E_n$ represents the event “*at least one event occurs*,” its complement has the form

$$\begin{aligned}\overline{E_1 \cup \dots \cup E_n} &= \{\text{none of them occurs}\} \\ &= \{E_1 \text{ does not occur} \cap \dots \cap E_n \text{ does not occur}\} \\ &= \overline{E_1} \cap \dots \cap \overline{E_n}\end{aligned}$$

The other equality can be proved by using the same way. □

Rephrasing, a complement to “*nothing*” is “*something*,” and “*not everything*” means “*at least one is missing*”

2 Rules of Probability

Mathematically, probability is introduced through several axioms.

2.1 Axioms of Probability

Not included.

2.2 Computing Probabilities of Events

2.2.1 Extreme Cases

A sample space Ω consists of all possible outcomes, therefore, it occurs for sure. On the contrary, an empty event \emptyset never occurs. So,

$$P\{\Omega\} = 1 \qquad \text{and} \qquad P\{\emptyset\} = 0$$

2.2.2 Union

Consider an event that consists of some finite or countable collection of *mutually exclusive outcomes*,

$$E = \{\omega_1, \omega_2, \omega_3, \dots\}$$

Summing probabilities of these outcomes, we obtain the probability of the entire event,

$$P\{E\} = \sum_{\omega_k \in E} P\{\omega_k\} = P\{\omega_1\} + P\{\omega_2\} + P\{\omega_3\} + \dots$$

It is crucial to notice that *only mutually exclusive events* (those with empty intersections) satisfy the **sigma-additivity**. If events intersect, their probabilities cannot be simply added.

In the sum $P\{A\} + P\{B\}$, all the common outcomes are counted twice. Each outcome should be counted only once! To correct the formula, subtract probabilities of common outcomes, which is $P\{A \cap B\}$.

$$\begin{aligned}P\{A \cup B\} &= P\{A\} + P\{B\} - P\{A \cap B\} \\ P\{A \cup B\} &= P\{A\} + P\{B\} \qquad \text{For mutually exclusive events}\end{aligned}$$

Generalization of this formula is not straightforward. For 3 events,

$$P\{A \cup B \cup C\} = P\{A\} + P\{B\} + P\{C\} - P\{A \cap B\} - P\{A \cap C\} - P\{B \cap C\} + P\{A \cap B \cap C\}$$

2.2.3 Complement

Recall that events A and \overline{A} are *exhaustive*, hence $A \cup \overline{A} = \Omega$. Also, they are *disjoint*, hence

$$P\{A\} + P\{\overline{A}\} = P\{A \cup \overline{A}\} = P\{\Omega\} = 1$$

Solving this for $P\{A\}$, we obtain a rule that perfectly agrees with the common sense,

$$P\{\overline{A}\} = 1 - P\{A\}$$

2.2.4 Intersection of Independent Events

Definition 5

Events E_1, \dots, E_n are **independent** if they occur *independently of each other*, i.e., occurrence of one event does not affect the probabilities of other.

The following basic formula can serve as the *criterion of independence*.

$$P\{E_1 \cap \dots \cap E_n\} = P\{E_1\} \times \dots \times P\{E_n\}$$

2.3 Applications in Reliability

Formulas of the previous section are widely used in *reliability*, when one computes the probability for a system of several components to be functional.

Example 3

There is a 1% probability for a hard drive to crash. Therefore, it has two backups, each having a 2% probability to crash, and all three components are independent of each other. The stored information is lost only in an unfortunate situation when all three devices crash. What is the probability that the information is saved?

Solution: Organize the data. Denote the events, say,

$$H = \{\text{hard drive crashes}\}, \quad B_1 = \{\text{first backup crashes}\}, \quad B_2 = \{\text{second backup crashes}\}$$

It is given that H , B_1 , and B_2 are independent,

$$P\{H\} = 0.01, \quad P\{B_1\} = P\{B_2\} = 0.02$$

Applying rules for the complement and for the intersection of independent events,

$$\begin{aligned} P\{\text{saved}\} &= 1 - P\{\text{lost}\} = 1 - P\{H \cap B_1 \cap B_2\} \\ &= 1 - P\{H\} \times P\{B_1\} \times P\{B_2\} \\ &= 1 - (0.01)(0.02)(0.02) = 0.999996 \end{aligned}$$

When the system's components are connected in *parallel*, it is sufficient for at least one component to work in order for the whole system to function. Reliability of such a system is computed as in example above. Backups can always be considered as devices connected in parallel.

At the other end, consider a system whose components are connected in *sequel*. Failure of one component inevitably causes the whole system to fail. Such a system is more “vulnerable.” In order to function with a high probability, it needs each component to be reliable, as in the next example.

Example 4

Suppose that a shuttle's launch depends on three key devices that operate independently of each other and malfunction with probabilities 0.01, 0.02, and 0.02, respectively. If any of the key devices malfunctions, the launch will be postponed. Compute the probability for the shuttle to be launched on time, according to its schedule.

Solution: In this case,

$$\begin{aligned} P\{\text{on time}\} &= P\{\text{all devices function}\} \\ &= P\{\overline{H} \cap \overline{B_1} \cap \overline{B_2}\} && \text{(independence)} \\ &= P\{\overline{H}\}P\{\overline{B_1}\}P\{\overline{B_2}\} && \text{(complement rule)} \\ &= (1 - 0.01)(1 - 0.02)(1 - 0.02) \\ &= 0.9508 \end{aligned}$$

3 Combinatorics

3.1 Equally Likely Outcomes

A simple situation for computing probabilities is the case of *equally likely outcomes*. That is, when the sample space Ω consists of n possible outcomes, $\omega_1, \dots, \omega_n$, each having the same probability. Since

$$\sum_1^n P\{\omega_k\} = P\{\Omega\} = 1$$

we have in this case $P\{\omega_k\} = 1/n$ for all k . Further, a probability of any event E consisting of t outcomes, equals

$$P\{E\} = \sum_{\omega_k \in E} \left(\frac{1}{n}\right) = t \left(\frac{1}{n}\right) = \frac{\text{number of outcomes in } E}{\text{number of outcomes in } \Omega}$$

3.2 Permutations and Combinations

- **Sampling with replacement:** You can choose an object *more than once*. Any object can be selected with $1/n$ probability at *any time*.
- **Sampling without replacement:** Sampled item is removed. Set of possibilities *reduces by 1* after every selection.
- **Distinguishable Objects:** Sampling of exactly the same objects in a *different order yields a different outcome*.
- **Indistinguishable Objects:** The *order is not important*. Differently arranged orders does not yield a different outcome.

3.2.1 Permutations with Replacement

Possible selections of k *distinguishable* objects from a set of n are called **permutations**.

Permutations with replacement

$$P_r(n, k) = \overbrace{n \cdot n \cdot \dots \cdot n}^{k \text{ terms}} = n^k$$

3.2.2 Permutations without Replacement

During selection, the number of possible selections reduces by 1 each time an object is sampled.

Permutations without replacement

$$P(n, k) = \overbrace{n(n-1)(n-2) \cdot \dots \cdot (n-k+1)}^{k \text{ terms}} = \frac{n!}{(n-k)!}$$

3.2.3 Combinations without Replacement

Possible selections of k *indistinguishable* objects from a set of n are called **combinations**. *The only difference from $P(n, k)$ is disregarding the order*

Combinations without replacement

$$C(n, k) = \binom{n}{k} = \frac{P(n, k)}{P(k, k)} = \frac{n!}{k!(n-k)!}$$

3.2.4 Computational Shortcuts

$$\begin{aligned} C(n, k) &= \binom{n}{k} = \frac{n \cdot (n-1) \cdot \dots \cdot (n-k+1)}{k \cdot (k-1) \cdot \dots \cdot 1} \\ C(n, k) &= C(n, n-k) \text{ for any } k \text{ and } n \\ C(n, 0) &= 1 \\ C(n, 1) &= n \end{aligned}$$

3.2.5 Combinations with Replacement

Combinations with replacement

$$C_r(n, k) = \binom{k+n-1}{k} = \frac{(k+n-1)!}{k!(n-1)!}$$

NOTATION

$P_r(n, k)$ = number of permutations with replacement

$P(n, k)$ = number of permutations without replacement

$C_r(n, k)$ = number of combinations with replacement

$C(n, k)$ = number of combinations without replacement

$\binom{n}{k}$ = number of combinations without replacement

4 Conditional Probability and Independence

4.1 Conditional Probability

Definition 6

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

NOTATION

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

Conditional probability basically is “What is the probability of observing event A given that event B is observed?”

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 ,

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

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

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

This appears to be the general formula

Conditional Probability

$$P\{A \mid B\} = \frac{P\{A \cap B\}}{P\{B\}}$$
$$P\{B \mid A\} = \frac{P\{A \cap B\}}{P\{A\}}$$

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

Intersection, general case

$$P\{A \cap B\} = P\{B\}P\{A \mid B\}$$
$$P\{A \cap B\} = P\{A\}P\{B \mid A\}$$

4.2 Independence

Definition 7

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

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

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

So, if $P\{A \mid B\} = P\{A\}$ and $P\{B \mid A\} = P\{B\}$, it shows that A and B are **independent event**, the probability of observing $A \mid B$ is not affected by observing other event!

According to this definition, *conditional* probability equals *unconditional* probability in case of independent events. Substituting this into formula of the probability of intersection:

$$P\{A \cap B\} = P\{A\}P\{B\}$$

Note: If the equation $P\{A \mid B\} = P\{A\}$ does not hold, the events are not independent.

Example 5

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) You are meeting a flight that departed on time. What is the probability that it will arrive on time?
- (b) You have met a flight, and it arrived on time. What is the probability that it departed on time?
- (c) Are the events, departing on time and arriving on time, independent?

Solution: Denote the events,

$A = \{\text{arriving on time}\},$

$D = \{\text{departing on time}\}$

We have:

$$P\{A\} = 0.8$$

$$P\{D\} = 0.9$$

$$P\{A \cap D\} = 0.75$$

- (a) $P\{A \mid D\} = \frac{P\{A \cap D\}}{P\{D\}} = \frac{0.75}{0.9} = 0.8333$
- (b) $P\{D \mid A\} = \frac{P\{A \cap D\}}{P\{A\}} = \frac{0.75}{0.8} = 0.9375$
- (c) Events are not independent because

$$P\{A \mid D\} \neq P\{A\}$$

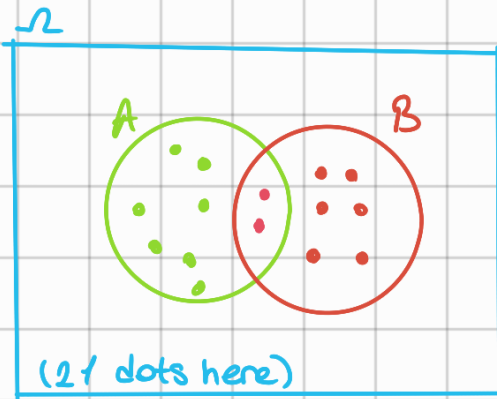
$$P\{D \mid A\} \neq P\{D\}$$

$$P\{A \cap D\} \neq P\{A\}P\{D\}$$

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

The following page belongs to *Berfinnur Oktay*.

→ We can see that this is true by the following example:



→ We have 36 points total in the sample space (all of them are equally probable)

→ A has 9 dots
→ B has 8 dots } $A \cap B = 2 \text{ dots}$

→ Is A and B are independent events?

If $P(A \cap B) = P(A) \cdot P(B)$ then A and B are independent events.

$$\frac{2}{36} \stackrel{?}{=} \frac{9}{36} \cdot \frac{8}{36} \rightarrow \frac{2}{36} = \frac{2}{36} \checkmark \text{ Yes, they are independent.}$$

→ What is the probability of observing A given that B is observed?

If B is observed, one of the 8 points in B is observed. So our sample space is 8. To observe A at the same time, one of the 2 points in the intersection must have been observed. Thus,

$$\frac{\text{desired}}{\text{sample}} = \frac{2}{8} = 0.25 = P(A) \text{ since } P(A) = \frac{9}{36} = \frac{1}{4} = 0.25$$

→ What is the probability of observing B given that A is observed?
Same logic as above.

$$\frac{\text{desired out.}}{\text{sample}} = \frac{\# \text{ in } A \cap B}{\# \text{ in } A} = \frac{2}{9} = P(B) \text{ since } P(B) = \frac{8}{36} = \frac{2}{9}$$

* So independent events does not affect each others probability of happening!

4.3 Bayes Rule

Two conditional probabilities, $P\{A | B\}$ and $P\{B | A\}$, are not the same, in general.

Example 6: 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 $P\{S | V\} = 0.95$, and if the patient does not have the virus, the test shows that with probability $P\{\bar{S} | \bar{V}\} = 0.99$

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, $P\{V | S\}$, is not stated among the given characteristics of this test.

Notice that $A \cap B = B \cap A$. Therefore, $P\{B\}P\{A | B\} = P\{A\}P\{B | A\}$, by using the general case intersection formula,. Solve for $P\{B | A\}$ to obtain

Bayes Rule

$$P\{B | A\} = \frac{P\{A | B\}P\{B\}}{P\{A\}}$$

$$P\{A | B\} = \frac{P\{B | A\}P\{A\}}{P\{B\}}$$

Example 7

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

By the *Bayes Rule*,

$$P\{ZB | GGB\} = \frac{P\{GGB | ZB\}P\{ZB\}}{P\{GGB\}}$$

Given ZB , event GGB occurs only when both X and Y write good exams. Thus, $P\{GGB | ZB\} = (0.8)(0.7)$. Event GGB consists of three outcomes depending on the student who wrote the bad exam. Adding their probabilities, we get

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

Then

$$P\{ZB | GGB\} = \frac{(0.8)(0.7)(0.5)}{0.47} = 0.5957$$

Sometimes, it can be hard to determine the denominator in the *Bayes Rule*. In the Bayes Rule, the denominator is often computed by the **Law of Total Probability**.

4.4 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 Ω with *mutually exclusive* and *exhaustive* events B_1, \dots, B_k . It means that

$$B_i \cap B_j = \emptyset \text{ for any } i \neq j \text{ and } B_1 \cup \dots \cup B_k = \Omega$$

These events also partition the event A

$$A = (A \cap B_1) \cup \dots \cup (A \cap B_k)$$

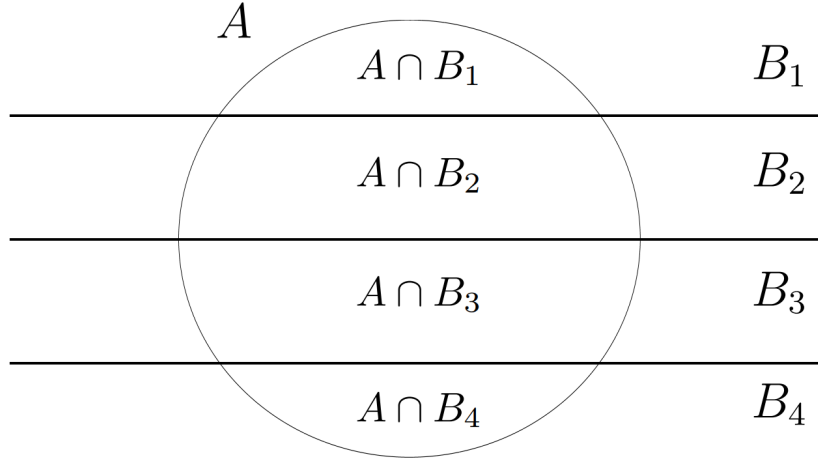


Figure 2: Partition of the sample space Ω and the event A

and this is also a union of mutually exclusive events. Hence,

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

and we arrive to the following rule.

Law of Total Probability

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

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

$$P\{A\} = P\{A \mid B\}P\{B\} + P\{A \mid \bar{B}\}P\{\bar{B}\}$$

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

Bayes Rule for two events

$$P\{B \mid A\} = \frac{P\{A \mid B\}P\{B\}}{P\{A \mid B\}P\{B\} + P\{A \mid \bar{B}\}P\{\bar{B}\}}$$

$$P\{A \mid B\} = \frac{P\{B \mid A\}P\{A\}}{P\{B \mid A\}P\{A\} + P\{B \mid \bar{A}\}P\{\bar{A}\}}$$

Example 8: Reliability of a test, continued

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

$$\begin{aligned} P\{V \mid S\} &= \frac{P\{S \mid V\}P\{V\}}{P\{S \mid V\}P\{V\} + P\{S \mid \bar{V}\}P\{\bar{V}\}} \\ &= \frac{(0.95)(0.04)}{(0.95)(0.04) + (1 - 0.99)(1 - 0.04)} = 0.7983 \end{aligned}$$

Check the example 2.35 from book.