4 Conditional Probability and Independent Events

Many times, we have to compute the probability of an event that depends on another event to some extent, so the probability of that other event has to be considered, too.

Definition 4.1. *Let $(S, \mathcal{K}, P)$ be a probability space and let $B \in \mathcal{K}$ be an event with $P(B) > 0$. Then for every $A \in \mathcal{K}$, the* ***conditional probability of*** *A* ***given*** *B (or the* ***probability of*** *A* ***conditioned by*** *B) is defined by*

$$P(A|B) \quad = \quad \frac{P(A \cap B)}{P(B)}. \tag{4.1}$$

Example 4.2. Ninety percent of flights depart on time. Eighty percent of flights arrive on time. Seventy-five percent of flights depart 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?

Solution. Denote the events
A: a flight arrives on time,
D: a flight departs on time.
Then we have:

$$P(A) = 0.8,\ P(D) = 0.9,\ P(A \cap D) = 0.75.$$

So,
a) $P(A|D) = \dfrac{P(A \cap D)}{P(D)} = \dfrac{0.75}{0.9} = 0.8333.$

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

■

An immediate consequence of Definition 4.1 is the following property:

Proposition 4.3. *Let $A, B \in \mathcal{K}$ with $P(A)P(B) \neq 0$. Then*

$$P(A \cap B) \quad = \quad P(A)P(B|A) \quad = \quad P(B)P(A|B). \tag{4.2}$$

This rule can be generalized to any number of events.

Proposition 4.4. (The Multiplication Rule)
Let $\{A_i\}_{i=\overline{1,n}} \subseteq \mathcal{K}$, with $P\left(A_1 \cap A_2 \cap \ldots \cap A_n\right) \neq 0$. Then

$$P(A_1 \cap \ldots \cap A_n) \quad = \quad P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \ldots P(A_n|A_1 \cap \ldots \cap A_{n-1}). \tag{4.3}$$

Proof. We start with the right hand side (RHS) of (4.3) and get to the left hand side (LHS). By (4.1), we have

$$RHS = P(A_1) \cdot \frac{P(A_1 \cap A_2)}{P(A_1)} \cdot \frac{P(A_1 \cap A_2 \cap A_3)}{P(A_1 \cap A_2)} \cdot \ldots \cdot \frac{P(A_1 \cap A_2 \ldots \cap A_n)}{P(A_1 \cap A_2 \ldots \cap A_{n-1})},$$

which, after cancellations, is $P(A_1 \cap \ldots \cap A_n)$, the LHS of (4.3).

$\square$

Proposition 4.5. *The probability of the complementary event formula still holds for conditional probabilities, i.e., if $A, B \in \mathcal{K}$ and $P(B) \neq 0$, then*

$$P(A|B) = 1 - P(\overline{A}|B). \tag{4.4}$$

Proof. We use the definition of conditional probability (4.1) and other known probability rules (from last time). We have

$$\begin{aligned} P(A|B) &= \frac{P(A \cap B)}{P(B)} = \frac{P\left(B \cap \overline{\overline{A}}\right)}{P(B)} \\ &= \frac{P\left(B \setminus \overline{A}\right)}{P(B)} = \frac{P(B) - P(\overline{A} \cap B))}{P(B)} \\ &= 1 - \frac{P(\overline{A} \cap B)}{P(B)} = 1 - P(\overline{A}|B). \end{aligned}$$

$\square$

Proposition 4.6. *For every $A, B \in \mathcal{K}$ with $0 < P(A) < 1$, we have*

$$P(B) = P(A)P(B|A) + P\left(\overline{A}\right) P\left(B|\overline{A}\right). \tag{4.5}$$

Proof. Since $\{A, \overline{A}\}$ form a partition of S, we have

$$B = B \cap S = B \cap \left(A \cup \overline{A}\right) = \underbrace{(B \cap A) \cup \left(B \cap \overline{A}\right)}_{\text{m.e.}}.$$

Note that $B \cap A$ and $B \cap \overline{A}$ are mutually exclusive, since A and $\overline{A}$ are. Then

$$P(B) = P(B \cap A) + P\left(B \cap \overline{A}\right).$$

Using (4.2) for both terms on the right hand side, we obtain (4.5).

$\square$

This result can also be generalized, for any partition of S.

Proposition 4.7. (The Total Probability Rule)
Let $\{A_i\}_{i\in I}$ be a partition of S and let $A \in \mathcal{K}$. Then

$$P(A) \;=\; \sum_{i\in I} P\left(A_i\right) P\left(A|A_i\right). \tag{4.6}$$

Proof. Just as before, we have

$$A = A \cap S = A \cap \left(\bigcup_{i\in I} A_i\right) = \bigcup_{i\in I} (\underbrace{A \cap A_i}_{\text{m.e.}}),$$

with $\{(A \cap A_i)\}_{i\in I}$ mutually exclusive and then

$$P(A) = \sum_{i\in I} P\left(A \cap A_i\right) \overset{(4.2)}{=} \sum_{i\in I} P(A_i)P(A|A_i).$$

$\square$

Example 4.8. A test for a certain viral infection is 95% reliable for infected patients (i.e. it gives a correct positive result) and 99% reliable for not infected ones (i.e. gives a correct negative result). It is known that 4% of the population is infected with that virus.
a) How reliable is the test in general (i.e. what is the probability that it shows a correct result)?
b) If a patient got a positive result, how likely is it that she truly is infected?

Solution. Denote the events
C: the test gives a correct result,
PR: the test gives a positive result,
V: a person has the virus (is infected).
What is given:

$$\begin{aligned}
P(C|V) &= P(PR|V) \;=\; 0.95,\\
P(C|\overline{V}) &= P(\overline{PR}|\overline{V}) \;=\; 0.99 \;\text{ and}\\
P(V) &= 0.04.
\end{aligned}$$

a) What we want is $P(C)$ (without any condition).
Notice that $\{V, \overline{V}\}$ form a partition of the sample space. By the Total Probability Rule (4.6), we have

$$\begin{aligned} P(C) &= P(C|V)P(V) + P(C|\overline{V})P(\overline{V}) \\ &= 0.95 \times 0.04 + 0.99 \times 0.96 \;=\; 0.9884. \end{aligned}$$

So, in general, the test is 98.84% reliable.

b) Here, we want $P(V|PR)$, which is given by

$$P(V|PR) = \frac{P(V \cap PR)}{P(PR)}.$$

The numerator is

$$P(V \cap PR) \overset{(4.2)}{=} P(V)P(PR|V) = 0.04 \times 0.95 = 0.038.$$

For the denominator, we use (4.6) again, with the same partition $\{V, \overline{V}\}$:

$$\begin{aligned} P(PR) &= P(PR|V)P(V) + P(PR|\overline{V})P(\overline{V}) \\ &\overset{(4.4)}{=} P(PR|V)P(V) + \left[1 - P(\overline{PR}|\overline{V})\right]P(\overline{V}) \\ &= 0.95 \times 0.04 + 0.01 \times 0.96 \;=\; 0.0476. \end{aligned}$$

Thus, the probability that the patient is indeed infected, is

$$P(V|PR) = \frac{0.038}{0.0476} = 0.7983.$$

■

Closely related to conditional probability is the notion of *independence* of events.

Definition 4.9. *Two events $A, B \in \mathcal{K}$ are said to be **independent** if*

$$P(A \cap B) = P(A)P(B). \tag{4.7}$$

*The events $\{A_n\}_{n\in\mathbb{N}} \subseteq \mathcal{K}$ are said to be **(mutually) independent** if*

$$P(A_{i_1} \cap \ldots \cap A_{i_k}) = P(A_{i_1}) \ldots P(A_{i_k}),$$

for any finite subset $\{i_1, \ldots, i_k\} \subset \mathbb{N}$.

Remark 4.10. If the events $A, B \in \mathcal{K}$ are independent, then $P(A|B) = P(A)$ and $P(B|A) = P(B)$. The converse is also true.

Example 4.11. Refer again to Example 4.2. Are the events "departing on time" and "arriving on time" independent?

Solution. No, because

$$0.75 = P(A \cap D) \neq P(A)P(D) = 0.8 \times 0.9 = 0.72.$$

Also notice that

$$P(A|D) = 0.8333 \neq 0.8 = P(A) \text{ and } P(D|A) = 0.9375 \neq 0.9 = P(D).$$

Further, we see that $P(A|D) > P(A)$ and $P(D|A) > P(D)$. In other words, departing on time increases the probability of arriving on time, and having arrived on time, it is more likely (probable) that the flight departed on time. ■

Proposition 4.12. *If $A = \emptyset$ (the impossible event, $P(A) = 0$) or $A = S$ (the certain event, P(A) = 1) and $B \in \mathcal{K}$ is any event, then A and B are independent.*

Proof. For the impossible event, we have

$$P(A \cap B) = P(\emptyset \cap B) = P(\emptyset) = 0 = P(A)P(B).$$

If A is the certain event, then

$$P(A \cap B) = P(S \cap B) = P(B) = P(A)P(B).$$

□

Proposition 4.13. *Let $A, B \in \mathcal{K}$ be independent events. Then A and $\overline{B}$ are also independent.*

Proof. We simply check the condition for independence:

$$\begin{aligned} P\left(A \cap \overline{B}\right) &= P(A \setminus B) \;=\; P(A) - P(A \cap B) \\ &= P(A) - P(A)P(B) \;=\; P(A)(1 - P(B)) \\ &= P(A)P\left(\overline{B}\right). \end{aligned}$$

$\square$

Remark 4.14.
1. A direct consequence of proposition 4.13 is that if $A, B \in \mathcal{K}$ are independent, then so are $\overline{A}, B$ and $\overline{A}, \overline{B}$.
2. More generally, if $A_1, A_2, ..., A_n \in \mathcal{K},\ n \in \mathbb{N}$ are independent, then so are $\overline{A}_1, \overline{A}_2, ..., \overline{A}_n$ and any combination of events and contrary events.

Chapter 2. Classical Probabilistic Models

In probability theory, one can notice that some experiments follow the same "patterns", so they are said to be in the same "class of experiments". Therefore, for each such class, we design a **probabilistic model**, which depends on certain parameters. For each model, we then find the corresponding general computational formulas, which then are applied to each experiment from that class, giving specific values to each parameter.

Sometimes, the easiest setup for describing a probabilistic model is to consider one (or more) box(es) containing a number (known or unknown) of balls, having a certain color distribution. The experiment consists of extracting one (or more) ball(s) from the box(es) (with or without putting it back) and noting its (their) color.

There is one important distinction that must be made! For an experiment, we can have

− **sampling *with* replacement**, meaning that once an object (a ball) is selected (extracted), it is replaced (returned to the box), so it can be selected again,

or

− **sampling *without* replacement**, which means that once an object is selected, it is NOT replaced, so it **cannot** be selected again.

If nothing else is specified, then the sampling is considered to be done *with* replacement.

1 Binomial Model

This model is used when the trials of an experiment satisfy three conditions, namely

(i) they are independent,

(ii) each trial has only two possible outcomes, which we refer to as "success" (A) and "failure" ($\overline{A}$) (i.e. the sample space for each trial is $S = A \cup \overline{A}$),

(iii) the probability of success $p = P(A)$ is the same for each trial (we denote by $q = 1-p = P(\overline{A})$ the probability of failure).

Trials of an experiment satisfying (i) − (iii) are known as **Bernoulli trials**.

<u>**Model:**</u> Given n Bernoulli trials with probability of success p, find the probability $P(n; k)$ of exactly k ($0 \le k \le n$) successes occurring.

Remark 1.1. For the Binomial model, the parameters are n (number of trials) and p (probability of success). These are the numbers that describe the model. The number k is **not** a parameter of the model. It varies from 0 to n (all possible numbers of successes in n trials), depending on which probability we are interested in computing.

Proposition 1.2. *The probability $P(n;k)$ in a Binomial model is given by*

$$P(n;k) \;=\; C_n^k p^k (1-p)^{n-k} \;=\; C_n^k p^k q^{n-k}, \;\; k = 0, 1, \ldots, n. \tag{1.8}$$

Remark 1.3.
1. The probability $P(n;k)$ is the coefficient of x^k in the Binomial expansion

$$(px+q)^n \;=\; \sum_{k=0}^{n} C_n^k p^k q^{n-k} x^k \;=\; \sum_{k=0}^{n} P(n;k) x^k,$$

hence the name of this model.
2. As a consequence (let $x = 1$ above),

$$\sum_{k=0}^{n} P(n;k) \;=\; 1.$$

This also follows from the fact that the events $\{k \text{ successes occur}\}_{k=0}^{n}$ form a partition of S.

Example 1.4. A die is rolled 5 times. Find the probability of the events
a) A : getting three 6's,
b) B : getting at least two even numbers.

Solution. Here a trial is a roll of the die. Therefore, $n = 5$.
a) For the first part, "success" means rolling a 6. Hence, $p = \dfrac{1}{6}$. This is a Binomial model with parameters $n = 5, p = 1/6$ and we have

$$P(A) = P(5;3) = C_5^3 \left(\frac{1}{6}\right)^3 \left(\frac{5}{6}\right)^2 \approx 0.0322.$$

b) For part two, "success" means getting an even number, so $p = \dfrac{1}{2}$. This a Binomial model with $n = 5$ and $p = 1/2$. To obtain at least 2 successes (out of 5 trials), means to obtain 2, 3, 4 or 5 successes. These events are mutually exclusive (one and only one at a time can happen), thus,

$$P(B) = P(5;2) + P(5;3) + P(5;4) + P(5;5).$$

However, in this case it is easier to compute the probability of the contrary event, which is "at most 1 success", since there are fewer cases (0 or 1). Thus,

$$\begin{aligned} P(B) &= 1 - P(\overline{B}) = 1 - \Big(P(5;0) + P(5;1)\Big) \\ &= 1 - \left(C_5^0 \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^5 + C_5^1 \left(\frac{1}{2}\right)^1 \left(\frac{1}{2}\right)^4\right) \approx 0.8125. \end{aligned}$$

■