Applied Probability for Computer Science

Federico Segala

Table of Contents

1. Overview of Elementary Probability

This chapter is focused on providing a concise overview of some fundamental concepts in probability theory that will be essential for understanding more advanced topics that will be presented in the following chapters.

1.1. Sample Space and Events

This section is focused on providing a clear understanding of the concepts of **sample space** and **events**, which are foundational elements in probability theory. It will first introduce the definitions and then provide some examples to better understand these concepts.

> **Definition 1.1 (Sample Space and Events)**
>
> A collection of all elementary results, or **outcomes** of an experiment, is called a **sample space**. Any set of *outcomes* from the sample space is called an **event**. In other words, we can state that an event can be viewed as an **arbitrary subset** of the sample space. ♣

Following we introduce some common *notation* that will be used throughout the notes for referring to specific concepts:

- Ω, S are typically used to denote the **sample space**.
- $\emptyset$ is typically used to denote the empty set or event
- A, B, E and other *capital letters* are used for events
- ω, s are going to be used for **individual outcomes**
- we will use the notation $\mathbb{P}[E]$ or $P[E]$ to denote the **probability** of an event

> 💬 **Remark**
>
> One important aspect to consider about the empty set $\emptyset$ is that it belongs to any sample space, i.e., $\emptyset \in \Omega \; \forall \Omega$.

> 💬 **Remark**
>
> Regarding events and individual outcomes it is important to remember that $E \subset \Omega$ and that, to distinguish between an event with a single outcome and the outcome itself we have respectively $\omega \in \Omega$ and $\{\omega\} \subset \Omega$.

> **Example: Simple Die Toss**
>
> Suppose a tossed die can produce one of 6 possible outcomes: 1 dot, through 6 dots. Each outcome is an **event**. Anyway there are other possible events, such as 'observing an even number of dots', a number of dots which is less than 4, an so on.

1.1.1. Some examples of Sample Spaces

Exercise. Two identical birds are initially on two nearby trees (A and B respectively). At random intervals, a sudden noise frightens one of the birds, making it fly to the other tree. At

each event, each bird has the same probability of being the one frightened and changing tree. Find the **sample space** if *only the number of birds on tree A* is considered.

Solution. Since the problem asks for the sample space in terms of number of birds on a specific tree, i.e., A. Tree A can have either 0, 1, or 2 birds; Therefore the sample space is $\Omega = \{0, 1, 2\}$.

> 💬 **Remark**
>
> It is important to notice that we are only interested in the number of birds on a specific tree; therefore the sample space will be independent on other birds features, such as their color, size, and so on, so forth.

Exercise. Let's make things more complicated, if we were to consider both the number of birds on tree A and on tree B, what would be the sample space?

Solution. Each event in this case, can be represented as a **tuple** of the form (x, y), where x is the number of birds on tree A and y is the number of birds on tree B. Therefore the possible events are: $\Omega = \{(0, 2), (1, 1), (0, 2)\}$.

Taking into account the assignment of the previous exercises, we can notice that there are two important aspects to consider:

- the sudden noise occurs at **random intervals**
- the happening of the noise can be considered an **event**. At **each event**, each bird has the same probability of moving

Clearly, the concept we have failed to represent so far is **time**. To be more precise, we can write the sample spaces prior to any noise event as respectively: $\Omega_0 = \{0, 1, 2\}$ and $\Omega_0 = \{(0, 2), (1, 1), (2, 0)\}$.

If we wanted to describe the **full experiment** we would need to descrive an **infinite sequence** of states. This is what's called a **random sequence**. We can briefly describe it as in Equation 1.

$$\Omega = \{(x_0, x_1, x_2, ...) \mid x_i \in \Omega_0 \ \forall i \in \mathbb{N}\} \tag{1}$$

> 💡 **Tip**
>
> Before starting to solve any kind of probability problem, also during the exam, always start by **defining** the **sample space**. This is a fundamental step that will help at better understanding the problem and avoid mistakes throughout the solution process.

Once we have defined the sample space, we can start defining **events**. For example, suppose the event A stands to represent the case when "initially both birds are on tree B": $A = \{(0, 2)\}$. Provided we have this information, Equation 2 shows the sample space for the presence on **tree A** conditioned on the fact that the starting position is the one described by A.

$$\Omega_A = \{0, 1, x_3, 1, x_4, ... \mid x_i \in \{0, 2\}\} \subset \Omega \tag{2}$$

1.1.2. Discrete vs Continuous Sample Spaces

One additional important aspect to consider is the **countability** of the sample space. Indeed we may find ourselves working with either **discrete** or **continuous** sample spaces. A discrete sample space is one that is either *finite* or *countably infinite*, meaning that its elements can be put into a one-to-one correspondence with the natural numbers. On the other hand, a continuous sample space contains an uncountably infinite number of outcomes, often represented by intervals of real numbers.

For instance, the time between consecutive noise events in the previous example could be modeled by means of a *continuous sample space*: $\Omega = \{t_0, t_1, t_2, \dots \mid t_j \in \mathbb{R}^+\}$.

Instead of recording the birds' positions only after each noise event, we could decide to record their positions **continuously over time**. In this case the sample space would become:

$$\Omega = \left\{ (X_t)_{t \geq 0} : x_t \in \{0, 1, 2\} \right\},$$

where at any given point in time t, X_t represents the number of birds on tree A.

1.2. Basic Set Operations for Events

This section is focused on introducing some basic set operations that can be performed on events. Since events are subsets of the sample space, we can apply standard set operations such as union, intersection, and complement to them.

Union of Events

The **union** of two events A and B, denoted by $A \cup B$, represents the event that either event A occurs, event B occurs, or both events occur. In other words, the union of events includes all outcomes that are in either event.

Intersection of Events

The **intersection** of two events A and B, denoted by $A \cap B$, represents the event that both event A and event B occur simultaneously. The intersection of events includes only those outcomes that are common to both events.

Complement of an Event

The **complement** of an event A, denoted by $\overline{A}$, represents the event that event A does not occur. The complement of an event includes all outcomes that are in the sample space but not in event A.

> **Definition 1.2 (Set Difference)**
>
> Given two sets A and B, the **set difference** of A and B, denoted by $A \setminus B$, is defined as the set of elements that are in A but not in B. Formally, it can be expressed as:
>
> $$A \setminus B = A \cap \overline{B}$$

De Morgan's Laws

De Morgan's Laws provide a relationship between the union and intersection of sets through complementation.

> **Theorem 1.1 (The Morgan 1st Law)**
>
> Given sets $E_1, ..., E_n$, the complement of their union is equal to the intersection of their complements:
>
> $$\overline{E_1 \cup E_2 \cup ... \cup E_n} = \overline{E_1} \cap \overline{E_2} \cap ... \cap \overline{E_n}$$
>
> ♡

> **Theorem 1.2 (The Morgan 2nd Law)**
>
> Given sets $E_1, ..., E_n$, the complement of their intersection is equal to the union of their complements:
>
> $$\overline{E_1 \cap E_2 \cap ... \cap E_n} = \overline{E_1} \cup \overline{E_2} \cup ... \cup \overline{E_n}$$
>
> ♡

Disjoint and Exhaustive Events

There are two important concepts related to events that are worth mentioning: **disjoint** and **exhaustive** events.

> **Definition 1.3 (Disjoint Events)**
>
> Events A and B are **disjoint** if their intersection is empty:
>
> $$A \cap B = \text{emptyset}$$
>
> Events $E_1, E_2, ...$ are **mutually exclusive** or **pairwise disjoint** if any two of these events are disjoint:
>
> $$E_i \cap E_j = \emptyset \ \forall i \neq j$$
>
> ♣

> **Definition 1.4 (Exhaustive Events)**
>
> A collection of events $E_1, E_2, ...$ is said to be **exhaustive** if their union covers the entire sample space:
>
> $$E_1 \cup E_2 \cup ... = \Omega$$
>
> ♣

Partitions of the Sample Space

Defining mutual exclusivity and exhaustivity allows us to introduce the concept of **partitions** of the sample space. This is a useful concept when dealing with events that cover the entire sample space without overlapping.

We say that a collection of **mutually exclusive** and **exhaustive** events $E_1, E_2, ...$ forms a **partition** of the sample space Ω. This concept is clearly shown in Figure 1
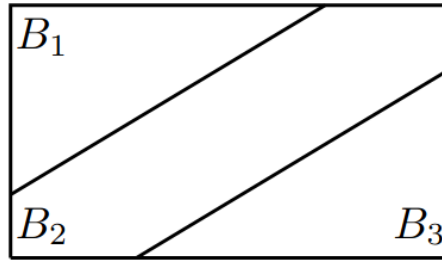
Figure 1: Example of a partition of the sample space

> **⬚ Remark**
>
> Any event $A \subset \Omega$ can be written in terms of the union of its intersections with the elements of the partition. This is illustrated in Figure 2:
>
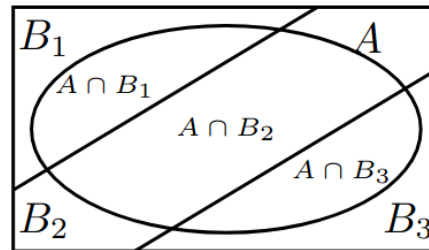> $$A = (A \cap B_1) \cup (A \cap B_2) \cup (A \cap B_3) \tag{3}$$

Figure 2: Event represented in terms of a partition of the sample space

1.3. Definition of Probability

In this section we are going to try to provide a formal definition for the notion of **probability**. First of all the reason why we introduced sample spaces and events in the first place, is that events are the entities for which we can compute *probabilities*.

In a very coarse way, we can think of probability as a **measure** that assigns to each event a number between 0 and 1, representing the **likelihood** of that event occurring. In a more technical language, a probability is a **function** which maps each event from the sample space to a real number in the interval $[0, 1]$.

A little Digression

In mathematics, a function che be described as the following:

$$f(x) = \sin(x)$$
$$f : \mathbb{R} \to [-1, 1]$$

We can see that we need to define both the **domain** and the **codomain** of the function along with the **rule** that describes how to map elements from the domain to elements in the codomain. Since probability is a function we'll need to do the same for it.

Sigma Algebra

Now that we have seen how a probability can be seen as a function, to formally define it, we need to specify its domain, codomain, and the mapping rule. This will give us a solid foundation to work with probabilities in a rigorous manner.

As per the **codomain**, we have already mentioned that probabilities are real numbers in the interval $[0, 1] \in \mathbb{R}$. The matter becomes a little bit more complicated when it comes to defining the **domain**. We can imagine the domain of the probability function as a *collection of events* with some specific properties. This collection is called a **sigma-algebra** (or **σ-algebra**). Typically, a sigma-algebra is denoted by the symbol $\mathcal{M}$.

Definition 1.5 (Sigma Algebra)

A collection $\mathcal{M}$ of events is a **σ - algebra** on a sample space Ω if:

- it includes the sample space:

$$\Omega \in \mathcal{M}$$

- every event in $\mathcal{M}$ is contained along with its complement:

$$E \in \mathcal{M} \Rightarrow \overline{E} \in \mathcal{M}$$

- every finite or countable collection of events in $\mathcal{M}$ is contained along with its union:

$$E_1, E_2, ... \in \mathcal{M} \Rightarrow E_1 \cup E_2 \cup ... \in \mathcal{M}$$

♣

We can notice the following important aspects about sigma-algebras:

- $\mathcal{M} = \{\emptyset, \Omega\}$ is the smallest possible sigma-algebra on Ω, called **degenerate**
- $\mathcal{M} = 2^{\Omega} = \{E : E \subset \Omega\}$ is the largest possible sigma-algebra on Ω, called **power set**

> 🗨 **Remark**
>
> When $\Omega \subseteq \mathbb{N}$, is **countable** the most common choice for the associated sigma algebra is the **power set** $\mathcal{M} = 2^{\Omega}$.
>
> On the other hand, when dealing with **uncountable** $\Omega \subseteq \mathbb{R}$, the power set *too large* to be useful. In this case a common choice for the sigma-algebra is the **Borel Sigma Algebra**, denoted by $\mathcal{B}$, which contains all possible sets that one could practically think about except for everything that could get created by some strange recursive process that resembles the construction of *fractals*.

Axiomatic Definition of Probability

Now that we have defined the sigma-algebra, we can finally provide a formal definition of probability.

> **Definition 1.6 (Probability)**
>
> Assume a sample space Ω and a sigma-algebra of events $\mathcal{M}$ defined on it. **Probability**
>
> $$\mathbb{P} \to [0, 1] \tag{4}$$
>
> is a function of events with the domain $\mathcal{M}$ and the range $[0, 1]$ that satisfies the following conditions (which are called the **axioms of probability**):
>
> - **Unit Measure**: the sample space has unit probability: $\mathbb{P}[\Omega] = 1$
> - **Sigma-Additivity**: for any finite or countable collection of mutually exclusive events $E_1, E_2, ... \in \mathcal{M}$, the probability of their union is equal to the sum of their individual probabilities:
>
> $$\mathbb{P}[E_1 \cup E_2 \cup ...] = \mathbb{P}[E_1] + \mathbb{P}[E_2] + ...$$
>
> ♣

It is good to notice that, from the first properties, we can derive that the computing the probability of the sample space amounts to say: '*something happened*'. The second property becomes fundamental when dealing with events that can be broken down into simpler, mutually exclusive events, allowing us to compute their probabilities more easily.

All rules of probability are a direct consequence of Definition 1.6°. This will allow us to compute probabilities for all events in our interest. Following we outline some of the most important probability rules that will be useful in the next chapters.

- $\mathbb{P}[\emptyset] = 0$: this is easy to verify; indeed the we know from the axioms that $\mathbb{P}[\Omega] = 1$. From the second axiom we know that the union of any disjoint event has probability equal to their sum, that is $\mathbb{P}(\Omega) + \mathbb{P}[\emptyset] = 1 \Rightarrow \mathbb{P}[\emptyset] = 0$ ∎
- $\mathbb{P}[A \cup B] = \mathbb{P}[A] + \mathbb{P}[B] - \mathbb{P}[A \cap B]$; we can actually notice that the following relation holds: $\mathbb{P}[A \cup B] = \mathbb{P}[A \cap \overline{B}] + \mathbb{P}[B \cap \overline{A}] + \mathbb{P}[A \cap B]$.

We can see how the second formulation is supported by the second axiom in that the three members of the summation form a **partition** of the event $A \cup B$. The reason why the first formulation is commonly preferred, is that that, if A and B are **independent** the probability of their intersection is given by $\mathbb{P}[A]\mathbb{P}[B]$.

Intuitively, saying that two experiments are independent means that the outcome of one does not affect the outcome of the other one. To be more formal we should also consider that the occurrence of an event doesn't even influence the events' probability in the other experiment.

We would even like to be more formal about the definition of independency but it's not possible without first introducing the concept of **conditional probability**.

Inclusion - Exclusion Principle

There are many cases in which we may be interested in computing the probability (or the amount of elements) of the union of multiple events. We already know how to do that in case of two events. In case we have three or more events we can generalize through the **Inclusion-Exclusion Principle**. We know that in the two event case we have:

$$\text{count}(A_1 \cup A_2) = \text{count}(A_1) + \text{count}(A_2) - \text{count}(A_1 \cap A_2)$$

Suppose now that we have three non-disjoint events A_1, A_2, A_3 as represented in .

The number elements in the union in such case would be given by:

$$\text{count}\left(\bigcup_{i=1}^{3} A_i\right) = \sum_{i=1}^{3} \text{count}(A_i) - \sum_{i<j} \text{count}(A_i \cap A_j) + \text{count}\left(\bigcap_{i=1}^{3} A_i\right)$$
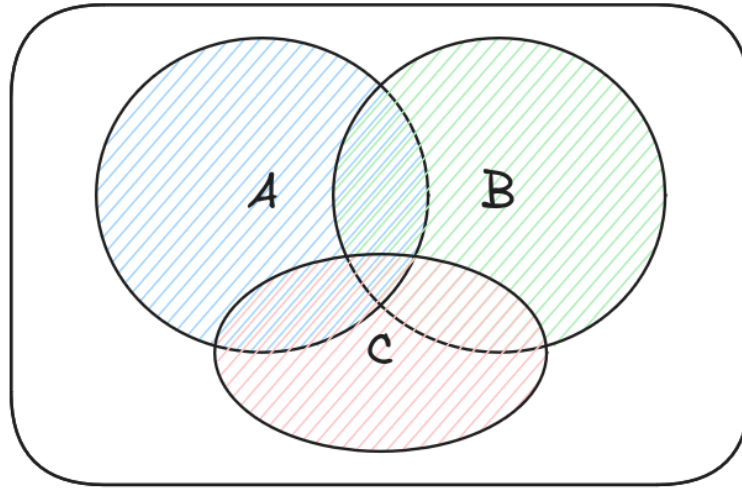

Figure 3: Three non-disjoint events

We are not going to see the full generalized version of the formula, mainly because it is so ugly. Anyway as a general principle we can say that to compute the count of the union multiple events:

- we first sum the counts of each individual event
- we subtract the counts ef each pairwise union (even number of events)
- we add back the counts of each triple-wise union (odd number of events)
- we keep on alternating between subtracting the count of "even-unions" and adding the count of "odd-unions" until we reach the union of all events

1.4. Conditional Probability

After defining the basic notion of probability, it's time to move on to a slightly more powerful concept, which is the one of **conditional probability**. The following definition gives us a way to compute conditional probabilities of two events.

> **Definition 1.7 (Conditional Probability)**
>
> Given two events A and B we can define the conditional probability of one event *given* that the other one has occurred as follows:
>
> $$\mathbb{P}\{A|B\} = \frac{\mathbb{P}\{A \cap B\}}{\mathbb{P}\{B\}} \tag{5}$$
>
> ♣

To understand the meaning of Equation 5 we can consider the following example.

> **Example: Computation of conditional probability**
>
> Suppose we throw a die 2 times and observe that the sum is 7. We can assert the following:
>
> - the probability that the event $(6, 6)$ happened is 0, since it is impossible to obtain a 7 with $(6, 6)$
> - the probability the rolling $(3, 4)$ can be computed even without looking at Equation 5. Indeed if we know that the sum of the tosses is 7, we can manually compute the sample space: $\{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\}$. By simply noticing that the event $(3, 4)$ is present only once in the sample space we conclude that its probability is given by $\frac{1}{6}$.

Formula Derivation

The reason behind the formulation of Equation 5 is the following: before an experiment the sample space Ω is the set of all possible experiments outcomes. Due to the fact that we are considering a probability we are dealing with a sigma algebra $\mathcal{M}$, we can make safely state the following:

$$A, B \in \mathcal{M} \Rightarrow A \cap B \in \mathcal{M} \Rightarrow \mathbb{P}\{A\}, \mathbb{P}\{B\}, \mathbb{P}\{A \cap B\} \text{ are known}$$

or at least can be computed in some way. We can say that $\mathbb{P} : \mathcal{M} \to [0, 1]$ is a **prior probability**. If now we perform the experiment and know that B happens we can update the probability to incorporate the new knowledge by computing a new $\mathbb{P} \mid B : \mathcal{M} \to [0, 1]$, a **posterior probability**. In practical terms, if B happened, $\overline{B}$ becomes impossible and we can **restrict our sample space** to only those outcomes in which B happens. Therefore the new sample space becomes $\Omega_B = \{\omega \in \Omega \mid \omega \in B\}$.

Let's understand the reason why we need to divide by $\mathbb{P}\{B\}$ in Equation 5:

- suppose we are interested in the whole sample space Ω

- prior to knowing that B happened, the probability of Ω is obviously 1: $\mathbb{P}\{\Omega\} = 1$

- after knowing that B happened, the new sample space becomes $\Omega_B = \Omega \cap B = B$

- since the new sample space is $\Omega \cap B$, its probability must still be 1: $\mathbb{P}\{\Omega \cap B\} = 1$, but since $\Omega \cap B = B$ we have that $\mathbb{P}\{B \mid B\} = 1$

- since we want to make sure that $\mathbb{P}\{B \mid B\} = 1$ we can compute $\frac{\mathbb{P}\{\Omega \cap B\}}{\mathbb{P}\{B\}} = 1$, in order to normalize the probability to 1

We can now focus on the single posterior probability of the event A given that B happened following a similar reasoning:

- for any event A prior to knowing that B happened, its probability is given by $\mathbb{P}\{A\}$

- after knowing that B happened, the new sample space becomes $\Omega_B = \Omega \cap B = B$

- since we need to compute the probability of A in the new sample space, we need to restrict A to only those outcomes in which B happens, therefore the new event becomes $A \cap B$. Thus the probability of A in the new sample space becomes $\mathbb{P}\{A \cap B\}$

- to scale everything back to the previous sample space we need to divide by $\mathbb{P}\{B\}$.

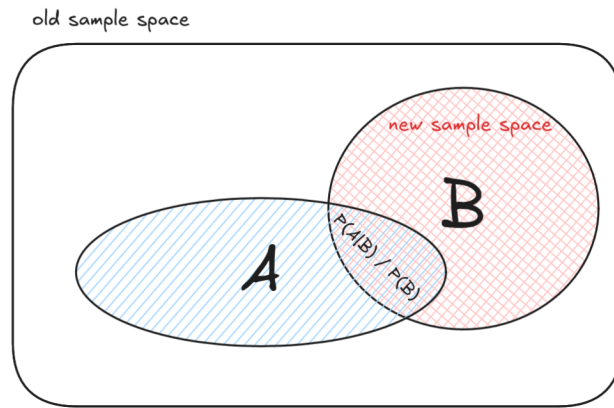The previous reasoning is clearly represented in the Venn diagram shown in Figure 4.

Figure 4: Venn diagram representation of conditional probability

We are now ready to provide a formal definition of independency between two events.

> **Definition 1.8 (Independent Events)**
>
> Given two events A and B, we say that they are independent if by knowing that one event has changed the probability of the other event remains the same. Formally, this can be expressed as:
>
> $$\mathbb{P}\{A \mid B\} = \frac{\mathbb{P}\{A \cap B\}}{\mathbb{P}\{B\}} = \mathbb{P}\{A\} \tag{6}$$
>
> ♣

> ⚠ **Warning**
>
> Although the notions of independency and disjointedness may seem similar at a first glance, they are actually the **opposite**. Indeed if two events are disjoint, the occurrence of one

event implies that the other event cannot occur, which means that knowing one event has occurred changes the probability of the other event to zero. Therefore, disjoint events are not independent.

From the previous definition we can derive the following important relation for independent events.

> **Theorem 1.3**
>
> Given two independent events A, B we can look at their definition in order to derive a different formulation for their intersection:
>
> $$\mathbb{P}\{A \cap B\} = \mathbb{P}\{A\}\mathbb{P}\{B\} \tag{7}$$

1.5. Conditional Probability and Bayes' theorem

This section will introduce one of the most important results in all probability theory. Before doing so, we show a couple of examples.

1.5.1. Exercise - Crashes in a Computer Program (Baron 2.35)

Problem Statement

A new computer program consists of two modules. The first module contains an error with probability 0.2. The second module is more complex and has a probability of 0.4 of containing an error, *independently of the first one.*

An error in the first module alone causes the program to crash with probability 0.5. For the second module, an error causes a crash with probability 0.8. If there are errors in both modules the probability of a crash rises to 0.9. Suppose that the program has crashed. What is the probability of error in both modules? ∎

Definition of the Sample Space and Sigma Algebra

Let's first try to map all the information provided into terms we are already familiar with. The **experiment** basically consists in running the program. The **sample space Ω** should contain all results of this experiment, that is, the program itself and the result of executing it. We can notice how this. Since the experiment is very complicated it makes sense to try and focus only on relevant outcomes. Let's list all the relevant events that must be included in our **sigma algebra**:

- E_1: module 1 contains an error$\rightarrow \overline{E}_1 \in \mathcal{M}$ (by σ-algebra properties)

- E_2: module 2 contains an error $\rightarrow \overline{E}_2 \in \mathcal{M}$ (by σ-algebra properties)

- Ω: any event happens $\rightarrow \emptyset \in \mathcal{M}$ (by σ-algebra properties)

- C: program crashes $\rightarrow \overline{C} \in \mathcal{M}$ (by σ-algebra properties)

By the properties of sigma algebras we can also state the following events must be included in $\mathcal{M}$:

- $E_1 \cup E_2 \in \mathcal{M}$, since sigma algebras are closed under union

- $\overline{E_1 \cup E_2} \in \mathcal{M}$, since sigma algebras are closed under complement

- $\overline{E_1} \cap \overline{E_2} \in \mathcal{M}$, by applying De Morgan laws

The same rational can be applied starting from the complement events. Therefore the sigma algebra will also contain $E_1 \cap E_2$. Notice that by simply considering the events E_1, E_2, C we can already keep track of all the possible events we may be interested in ($C \cap E_1, C \cap E_2, ...$).

Information Extraction

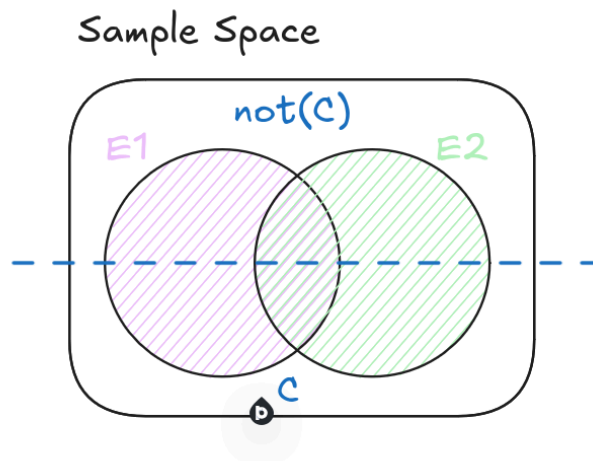Let's now try to extract valuable information from the problem statement:

- $\mathbb{P}\{E_1\} = 0.2$

- $\mathbb{P}\{E_2\} = 0.4$

Since $E_1 \perp E_2$ we can use Equation 7 to compute the probability of their intersection: $\mathbb{P}\{E_1 \cap E_2\} = \mathbb{P}\{E_1\}\mathbb{P}\{E_2\} = 0.08$. The problem also provides us with the following information:

- $\mathbb{P}\{C \mid E_1 \cap \overline{E_2}\} = 0.5$: crash given an error on first module alone

- $\mathbb{P}\{C \mid \overline{E_1} \cap E_2\} = 0.8$: crash given an error on second module alone

- $\mathbb{P}\{C \mid E_1 \cap E_2\} = 0.9$: crash given an error on both modules

Information Organization

Now that we have extracted all the relevant information from the problem statement, we can try to organize it in a more structured way. The main approach in order to compute probability is *divide and conquer*. So what we do is trying to divide our events in the smallest possible pieces, that are mutually exclusive and that cover the whole space. This amounts to finding **relevant partitions**. Consider the following picture:

One partition we can consider is the one given by the events E_1, E_2 combined and their complements:

- $B_1 = \overline{E_1} \cap \overline{E_2} \leftrightarrow \overline{E_1 \cup E_2}$: this represents the case in which no module has an error

- $B_2 = E_1 \cap \overline{E_2}$: this represents the case in which only the first module has an error

- $B_3 = \overline{E}_1 \cap E_2$: this represents the case in which only the second module has an error

- $B_4 = E_1 \cap E_2$: this represents the case in which both modules have an error

We can easily notice that these events (B_1, B_2, B_3, B_4) are mutually exclusive and span the whole sample space. Therefore they form a partition of the sample space. Another partition we can consider is the one given by the crash event and its complement: $(C, \overline{C})$.

In order to divide and conquer, we can **identify** the **partition** for which we have **prior probabilities** (probabilities which are not conditioned on other events) and the ones for which we need to compute **conditional probabilities**. This step allows us to *find a logical temporal order of events*. In our case we can notice that we have prior probabilities for the events in the partition (B_1, B_2, B_3, B_4) and conditional probabilities for the events in the partition $(C, \overline{C})$.

Now we can **compute** the **available probabilities** from the basic information we have extracted. First we can compute the probability of the *intersection $B_4 = E_1 \cap E_2$*:

$$\mathbb{P}\{B_4\} = \mathbb{P}\{E_1 \cap E_2\} = \mathbb{P}\{E_1\}\mathbb{P}\{E_2\} = 0.08$$

Now we would like to compute the probability of B_1, for doing so we need the probability of the union, that we can compute as follows:

$$\mathbb{P}\{E_1 \cup E_2\} = \mathbb{P}\{E_1\} + \mathbb{P}\{E_2\} - \mathbb{P}\{E_1 \cap E_2\} = 0.2 + 0.4 - 0.08 = 0.52$$

Thus, the probability of the event B_1 becomes: $\mathbb{P}\{B_1\} = 1 - \mathbb{P}\{E_1 \cup E_2\} = 1 - 0.52 = 0.48$.

To compute the probability of B_2 we need to compute the following:

- $\mathbb{P}\{B_2\} = \mathbb{P}\{E_1 \cap \overline{E}_2\}$, since $E_1 \perp E_2$, then also $E_1 \perp \overline{E}_2$, therefore we can apply Equation 7 and compute: $\mathbb{P}\{E_1 \cap \overline{E}_2\} = \mathbb{P}\{E_1\}\mathbb{P}\{\overline{E}_2\} = 0.2(1 - 0.4) = 0.12$

- the same reasoning can be applied to compute B_3, for completeness though, we can also compute both B_2, B_3 as follows (B_3 case):

$$\mathbb{P}\{B_3\} = \mathbb{P}\{E_2\} - \mathbb{P}\{E_1 \cap E_2\} = 0.4 - 0.08 = 0.32$$

> 🗨 **Remark**
>
> Since $(B_1...B_4)$ form a partition we can verify that the sum of their probabilities is equal to 1: $\mathbb{P}\{B_1\} + \mathbb{P}\{B_2\} + \mathbb{P}\{B_3\} + \mathbb{P}\{B_4\} = 0.48 + 0.12 + 0.32 + 0.08 = 1$.

Assuming that if we have no crashes the program works fine, we can produce the following diagram that summarizes all the information we have gathered so far. Tha's illustrate Figure 5.

Now that we have all the information we have gathered we are finally ready to give a solution. Since we are asked to compute the probability of having errors in both modules given that the program has crashed. We can rewrite it in terms of our events as follows:

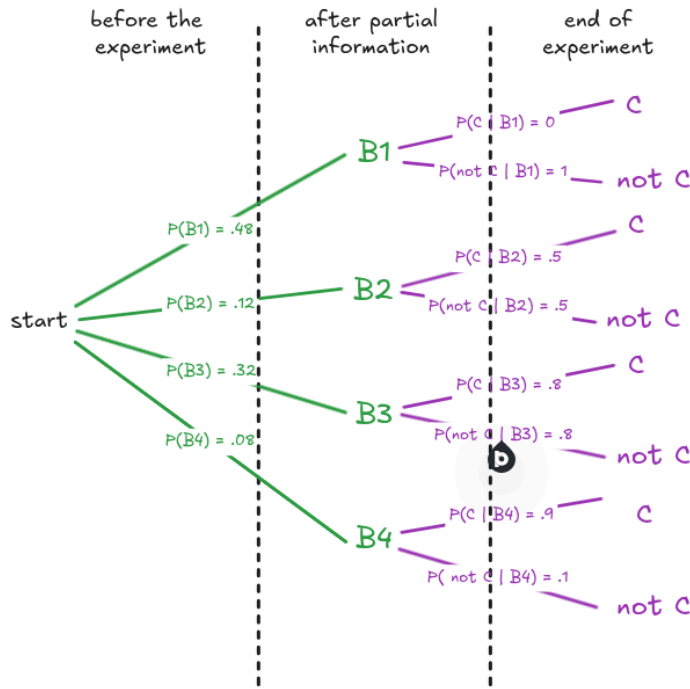$$\mathbb{P}\{E_1 \cap E_2 \mid C\} = \mathbb{P}\{B_4 \mid C\}$$

Figure 5: Summary of the information gathered from the problem statement and from the intermediate stages

We can retrieve the definition of conditional probability in Equation 5 and apply it to our case: $\mathbb{P}\{B_4 \mid C\} = \frac{\mathbb{P}\{B_4 \cap C\}}{\mathbb{P}\{C\}}$. To do so, we need different components.

First of all we need to compute the probability of $B_4 \cap C$, this can be done by backtracking on the tree in Figure 5:

$$\mathbb{P}\{B_4 \cap C\} = \mathbb{P}\{B_4\}\mathbb{P}\{C \mid B_4\} = 0.08 \cdot 0.9 = 0.072$$

Now wee need to compute the probability of a crash, that is $\mathbb{P}\{C\}$, this is a slightly more complicated matter. To do so, we need to take into consideration all the possible paths that lead from the root of the tree up to node C. Therefore we can write:

$$\mathbb{P}\{C\} = \mathbb{P}\{B_1\}\mathbb{P}\{C \mid B_1\} + \mathbb{P}\{B_2\}\mathbb{P}\{C \mid B_2\} + \mathbb{P}\{B_3\}\mathbb{P}\{C \mid B_3\} + \mathbb{P}\{B_4\}\mathbb{P}\{C \mid B_4\}$$

$$= \sum_{i=1}^{4} \mathbb{P}\{B_i\}\mathbb{P}\{C \mid B_i\}$$

$$= 0.48 \cdot 0 + 0.12 \cdot 0.5 + 0.32 \cdot 0.8 + 0.08 * 0.9 = 0.388$$

Now that we have all the elements we need, we can finally compute the probability we were looking for:

$$\mathbb{P}\{B_4 \mid C\} = \frac{\mathbb{P}\{B_4 \cap C\}}{\mathbb{P}\{C\}} = \frac{0.072}{0.388} \approx 0.1856 \ \blacksquare$$

1.5.2. Law of Total Probability

Even though we didn't explicitly mentioned, while resolving the last exercise we have actually applied probably what is one of the most important results in probability theory: the **law of total probability**.

Precisely, we did it when computing the probability of a crash $\mathbb{P}\{C\}$. We can now formally state the law as follows.

> **Axiom 1.1 (Law of Total Probability)**
>
> Given a partition of the sample space $(B_1, B_2, ...)$ and an event A, the probability of A can be computed as follows:
>
> $$\mathbb{P}\{A\} = \sum_i \mathbb{P}\{B_i\}\mathbb{P}\{A \mid B_i\} \tag{8}$$

This is the reason why in the previous example we spent so much time trying to compute the partition (B_1, B_2, B_3, B_4). In case we have only two events we can rewrite it as follows:

$$\mathbb{P}\{A\} = \mathbb{P}\{A \mid B\}\mathbb{P}\{B\} + \mathbb{P}\{A \mid \overline{B}\}\mathbb{P}\{\overline{B}\}$$

1.5.3. Bayes' Rule

Another important result that we can derive from the definition of conditional probability is **Bayes' Rule**. This rule allows us to *reverse* the conditioning of a probability. Formally we can state it as follows.

> **Theorem 1.4 (Bayes' Rule)**
>
> Given an event A and a partition of the sample space $B = (B_1, B_2, ..., B_k)$, the conditional probability of $B_i \in B$ given A can be computed as follows:
>
> $$\mathbb{P}\{B_i \mid A\} = \frac{\mathbb{P}\{A \mid B_i\}\mathbb{P}\{B_i\}}{\sum_{j=1}^{k} \mathbb{P}\{A \mid B_j\}\mathbb{P}\{B_j\}} \tag{9}$$
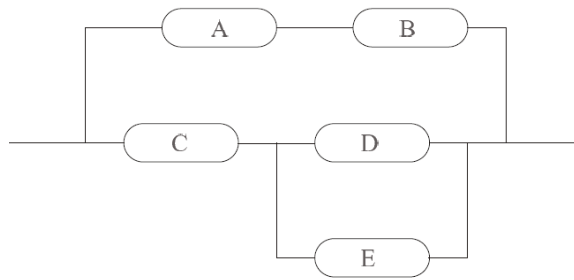
In particular, since B and $\overline{B}$ always form a partition of the sample space, we can rewrite Equation 9 for the case of two events as follows:

$$\mathbb{P}\{B \mid A\} = \frac{\mathbb{P}\{A \mid B\}\mathbb{P}\{B\}}{\mathbb{P}\{A \mid B\}\mathbb{P}\{B\} + \mathbb{P}\{A \mid \overline{B}\}\mathbb{P}\{\overline{B}\}} \tag{10}$$

1.5.4. Exercise - Reliability of a System (Baron 2.20)

Problem Statement

Consider the following system of connected components.

A B

C D

E

Calculate the reliability of the following system if each component is operable with probability 0.92 independently of the other components ∎

2. Ciao

> **Definition 2.1**
>
> Ciao
>
> ♣

Figures Index