Lecture 5: Independence of more than two events, law of total probability, Bayes' formula

1 Mutual independence of n events

We start by recalling the following definition.

Definition 1.1. Let $A_1, A_2, \cdots, A_n$ be n events. We say that $A_1, \cdots, A_n$ are *mutually independent*, if for any sub-collection of these events, namely for any sub-collection $\{j_1, \cdots, j_m\} \subseteq \{1, \cdots, n\}$, the following identity holds true:

$$\mathbb{P}(A_{j_1} \cap \cdots \cap A_{j_m}) = \mathbb{P}(A_{j_1}) \times \cdots \times \mathbb{P}(A_{j_m}).$$

Let us use a few examples to see how many equations are required in the definition of mutual independence.

Example 1.1. (i) If there are only two events A, B involved, there is only one equation needed, which is exactly the definition of independence for two events.
(ii) If there are three events A, B, C involved, we need *all* of the following equations to hold in order to establish mutual independence among the three events:

$$\text{sub-collection of two events: } \begin{cases} \mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B), \\ \mathbb{P}(A \cap C) = \mathbb{P}(A) \cdot \mathbb{P}(C), \\ \mathbb{P}(B \cap C) = \mathbb{P}(B) \cdot \mathbb{P}(C), \end{cases}$$

$$\text{sub-collection of three events: } \mathbb{P}(A \cap B \cap C) = \mathbb{P}(A) \cdot \mathbb{P}(B) \cdot \mathbb{P}(C).$$

In other words, there are 4 equations needed in total.
(iii) If there are four events A, B, C, D, we need all of the following equations:

$$\begin{matrix}\text{sub-collection of} \\ \text{two events:}\end{matrix} \begin{cases} \mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B), & \mathbb{P}(A \cap C) = \mathbb{P}(A) \cdot \mathbb{P}(C), \\ \mathbb{P}(A \cap D) = \mathbb{P}(A) \cdot \mathbb{P}(D), & \mathbb{P}(B \cap C) = \mathbb{P}(B) \cdot \mathbb{P}(C), \\ \mathbb{P}(B \cap D) = \mathbb{P}(B) \cdot \mathbb{P}(D), & \mathbb{P}(C \cap D) = \mathbb{P}(C) \cdot \mathbb{P}(D), \end{cases}$$

$$\begin{array}{c}\text{sub-collection of}\\ \text{three events:}\end{array}\begin{cases}\mathbb{P}(A\cap B\cap C)=\mathbb{P}(A)\cdot\mathbb{P}(B)\cdot\mathbb{P}(C),\\ \mathbb{P}(A\cap B\cap D)=\mathbb{P}(A)\cdot\mathbb{P}(B)\cdot\mathbb{P}(D),\\ \mathbb{P}(A\cap C\cap D)=\mathbb{P}(A)\cdot\mathbb{P}(C)\cdot\mathbb{P}(D),\\ \mathbb{P}(B\cap C\cap D)=\mathbb{P}(B)\cdot\mathbb{P}(C)\cdot\mathbb{P}(D),\end{cases}$$

and

$$\text{sub-collection of four events: } \mathbb{P}(A\cap B\cap C\cap D)=\mathbb{P}(A)\cdot\mathbb{P}(B)\cdot\mathbb{P}(C)\cdot\mathbb{P}(D).$$

There are 11 equations in total.
(iv) More generally, if there are n events involved, one needs totally 2^n-1-n equations (why?) to test mutual independence!

Property. Let $A_1, A_2, \cdots, A_n$ be a given family of mutually independent events. We can form many new collections of mutually independent events out of them. The basic principle is to select sub-collections of events from $\{A_1, \cdots, A_n\}$ in a *non-overlapping* way and perform arbitrary operations within each sub-collection we select. For instance, each of the following families of events are mutually independent:

(i) $A_1\cap A_2$, A_3^c;
(ii) $A_1\cup A_2, A_3\cup A_4, A_5, A_7^c$;
(iii) $A_1\cup A_2, A_3\cap A_4, (A_7\backslash A_9)\cup A_{15}$.

Let us justify the independence for Case (i).

Proof. We check the definition directly:

$$\begin{aligned}\mathbb{P}((A_1\cap A_2)\cap A_3^c)&=\mathbb{P}(A_1\cap A_2)-\mathbb{P}(A_1\cap A_2\cap A_3)\\&=\mathbb{P}(A_1)\cdot\mathbb{P}(A_2)-\mathbb{P}(A_1)\cdot\mathbb{P}(A_2)\cdot\mathbb{P}(A_3)\\&=\mathbb{P}(A_1)\cdot\mathbb{P}(A_2)\cdot(1-\mathbb{P}(A_3))\\&=\mathbb{P}(A_1\cap A_2)\cdot\mathbb{P}(A_3^c),\end{aligned}$$

where the second equality follows from two of the equations in the definition of mutual independence. □

On the other hand, as we have emphasised, it is important to form these collections in a non-overlapping way, for otherwise we will destroy independence. For instance, the events $A_1\cap A_2$ and $A_2\cap A_3$ are not independent in general (these two collections have a common member A_2 as an overlap). Indeed, if they were independent, by definition we would have

$$\mathbb{P}((A_1 \cap A_2) \cap (A_2 \cap A_3)) = \mathbb{P}(A_1 \cap A_2) \cdot \mathbb{P}(A_2 \cap A_3). \tag{1.1}$$

Since A_1, A_2, A_3 are assumed to be mutually independently, we also know that the left hand side of (1.1) equals

$$\mathbb{P}(A_1 \cap A_2 \cap A_3) = \mathbb{P}(A_1) \cdot \mathbb{P}(A_2) \cdot \mathbb{P}(A_3)$$

and the right hand side of (1.1) equals

$$\mathbb{P}(A_1) \cdot \mathbb{P}(A_2) \cdot \mathbb{P}(A_2) \cdot \mathbb{P}(A_3) = \mathbb{P}(A_1) \cdot \mathbb{P}(A_2)^2 \cdot \mathbb{P}(A_3)$$

As a result, we would obtain the following identity:

$$\mathbb{P}(A_1) \cdot \mathbb{P}(A_2) \cdot \mathbb{P}(A_3) = \mathbb{P}(A_1) \cdot \mathbb{P}(A_2)^2 \cdot \mathbb{P}(A_3).$$

However, this equation cannot be true if

$$\mathbb{P}(A_1) > 0, \ \mathbb{P}(A_3) > 0, \ 0 < \mathbb{P}(A_2) < 1 \quad \text{(why?)},$$

which is almost always the case for most examples.

Remark 1.1. As one particular equation from the definition of mutual independence, we have

$$\mathbb{P}(A_1 \cap A_2 \cap \cdots \cap A_n) = \mathbb{P}(A_1) \cdot \cdots \cdot \mathbb{P}(A_n). \tag{1.2}$$

However, only knowing (1.2) is far from being sufficient to guarantee the mutual independence of $A_1, \cdots, A_n$. A simple example is the three events $A, A, \emptyset$. Apparently, we have

$$\mathbb{P}(A \cap A \cap \emptyset) = \mathbb{P}(\emptyset) = 0 = \mathbb{P}(A) \cdot \mathbb{P}(A) \cdot \mathbb{P}(\emptyset).$$

However, these three events are not mutually independent since A is not independent of itself (unless in the extreme case when $\mathbb{P}(A) = 0$ or 1).

An application: system reliability

The notion of independence has a nice application in the study of system reliability. In a simplified context, we can think of a system as a circuit consisting of several components, connected in a mixture of sequential and parallel ways. The following figure gives two examples of a system:

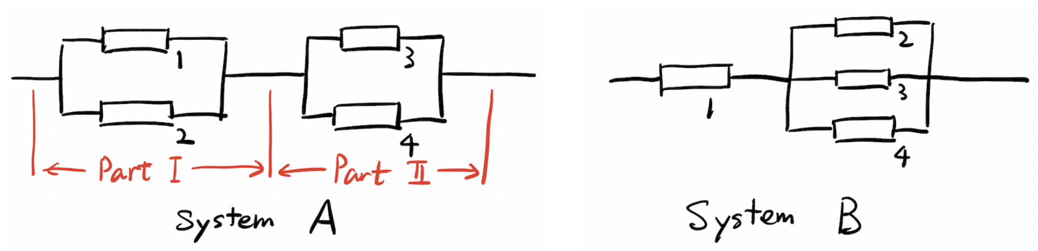

Figure 1: System reliability.

In a given system, it is often assumed that all the components function independently, but each component has a particular probability of failure. Current cannot flow through a component if it fails to function. The system functions normally if and only if the current can flow from the beginning (left) to the end (right). We interpret the *reliability* of the system as the probability that the system functions normally.

Let us try to understand which of the two systems given in Figure 1 is more reliable. We assume that every component has the same failure probability of 0.1.

For System A, we can divide it into Part I and Part II. Since these two parts are connected sequentially, we know that the system functions normally (i.e. current flows from left to right) if and only if both parts function normally. In addition, we have assumed that all the components are working independently. Therefore, by independence we have

$$\mathbb{P}(\text{System A functions}) = \mathbb{P}(\text{Part I functions}) \times \mathbb{P}(\text{Part II functions}).$$

We now compute the probability that Part I functions. Since the two components in Part I are connected in parallel, we know that Part I functions (i.e. current can flow through Part I) if and only if at least one of the two components functions. In this scenario, it is simpler to consider the complement of the events. Namely, we have

$$\begin{aligned}
&\mathbb{P}(\text{Part I functions}) \\
&= 1 - \mathbb{P}(\text{Part I fails}) \\
&= 1 - \mathbb{P}(\text{Component 1 and 2 both fails}) \\
&= 1 - \mathbb{P}(\text{Component 1 fails}) \times \mathbb{P}(\text{Component 2 fails}) \quad \text{(by independence)} \\
&= 1 - 0.1 \times 0.1 \\
&= 0.99.
\end{aligned}$$

The same result is true for Part II. Therefore, the reliability of the first system is given by

$$\mathbb{P}(\text{System 1 functions}) = 0.99 \times 0.99 = 0.9801.$$

The analysis of System B is entirely similar. We write down the formula directly:

$$\mathbb{P}(\text{System B functions}) = (1 - 0.1) \times (1 - 0.1^3) = 0.8991.$$

In other words, System A is more reliable than System B.

Here is a more general mathematical question for entertainment. Suppose that we are given n components at hand, each having failure probability p. Using these n components, how do we design the most reliable system?

2 The law of total probability

There are two particular important tools for computing probabilities. The first one is known as the law of total probability, and the other one is known as Bayes' formula. As we will see, the philosophy behind these two principles are very much in reverse to each other.

We first discuss the law of total probability. The basic idea behind this law can be summarised as follows. We think of an event as the effect / result due to one of several distinct causes / reasons. In this way, we compute the probability of the event by *conditioning on one of the possible causes* and *adding over these possibilities.*

Before giving the precise mathematical formulation, we first need to introduce the notion of a partition. Let Ω be the given sample space.

Definition 2.1. A collection of events $\{A_1, A_2, A_3 \cdots\}$ (it could be a finite collection or an infinite sequence) is said to be a *partition* of Ω, if these events are mutually disjoint and exhaustive, namely if

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

and

$$\cup_n A_n = \Omega.$$

The following figure gives an example of a (finite) partition of Ω :

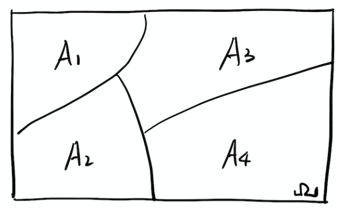

Figure 2: A partition of Ω.

To describe the law of total probability, let H be a given event and let $\{A_1, A_2, A_3, \cdots\}$ be a given partition (either finite or countably infinite) of Ω. Our goal is to compute $\mathbb{P}(H)$. For this purpose, we first write

$$\begin{aligned} H &= H \cap \Omega = H \cap (\cup_n A_n) \quad \text{(because the } A_n\text{'s are exhaustive)} \\ &= \cup_n (H \cap A_n) \end{aligned}$$

Since the events $H \cap A_n$ (with different n's) are also mutually disjoint, by the countable additivity axiom we have

$$\mathbb{P}(H) = \sum_n \mathbb{P}(H \cap A_n).$$

By further writing

$$\mathbb{P}(H \cap A_n) = \mathbb{P}(A_n) \cdot \mathbb{P}(H|A_n),$$

we obtain the law of total probability which is stated as follows.

The law of total probability. Let H be an event and let $\{A_1, A_2, \cdots\}$ be a given partition of Ω. Suppose that $\mathbb{P}(A_n) > 0$ for each n (in order to make sense of the conditional probability). Then we can compute $\mathbb{P}(H)$ by the formula:

$$\mathbb{P}(H) = \sum_n \mathbb{P}(A_n) \cdot \mathbb{P}(H|A_n). \tag{2.1}$$

Remark 2.1. The right hand side of (2.1) is a finite sum if the partition is finite, and it is an infinite sum (i.e. a series) if the partition is given by a sequence of events. In the above discussion, we only write the subscript n (without specifying its range) to allow both cases.

As mentioned earlier , the law of total probability can be interpreted as follows.

Interpretation. We interpret H as an effect and interpret a partition $\{A_1, A_2, \cdots\}$ as the possible distinct causes leading to the effect H. The effect of H is due to precisely one of the causes A_n. As a result, the law of total probability allows us to compute $\mathbb{P}(H)$ by first conditioning on one of these causes and then adding up all these (mutually disjoint and exhaustive) possibilities.

Example 2.1. An HIV test has 90% accuracy in the sense that if a person is HIV positive, there is a 90% chance that he/she will be tested positive. The test has 5% error in the sense that if a person is not HIV positive, there is still a 5% chance that he/she will be tested positive. Suppose that 0.01% of the population is HIV positive. What is the probability that a randomly selected person is tested positive?

Solution. To answer this question, we first define the following events:

$$\begin{gathered} H : \text{a randomly selected person is tested positive}, \\ A_1 : \text{the person is HIV positive}, \\ A_2 : \text{the person is not HIV positive}. \end{gathered}$$

We can think of the event H (tested positive) as the effect of two possible causes: either the person is indeed HIV positive or the person is healthy (but the test makes a mistake). These two causes are apparently disjoint and exhaustive. We can thus use the law of total probability:

$$\begin{aligned} \mathbb{P}(H) &= \mathbb{P}(A_1) \cdot \mathbb{P}(H|A_1) + \mathbb{P}(A_2) \cdot \mathbb{P}(H|A_2) \\ &= 0.0001 \cdot 0.9 + 0.9999 \cdot 0.05 \\ &\approx 0.05. \end{aligned}$$

A more interesting question for the previous example is the following. Given that the person is tested positive, what is the conditional probability that he/she is actually HIV positive? This question is important as it gives a clear indication on the effectivity of the test. The nature of this question is about finding a particular cause given the effect. This leads us to the other important tool called Bayes' formula.

3 Bayes' formula

We continue with the same assumption as for the law of total probability. Namely, H is a given event and $\{A_1, A_2, \cdots\}$ is a given partition of Ω. Bayes' formula is concerned with computing the conditional probability $\mathbb{P}(A_i|H)$ for some i. The formula is algebraically very simple to establish:

$$\begin{aligned}\mathbb{P}(A_i|H) &= \frac{\mathbb{P}(A_i \cap H)}{\mathbb{P}(H)} \quad \text{(definition of conditional probability)} \\ &= \frac{\mathbb{P}(A_i) \cdot \mathbb{P}(H|A_i)}{\sum_j \mathbb{P}(A_j) \cdot \mathbb{P}(H|A_j)} \quad \text{(applying L.O.T.P. to the denominator).}\end{aligned}$$

Bayes' formula. Let H be an event and let $\{A_1, A_2, \cdots\}$ be a given partition of Ω. Suppose that $\mathbb{P}(H) > 0$ and $\mathbb{P}(A_n) > 0$ for each n. Then for each i, we can compute $\mathbb{P}(A_i|H)$ by the formula:

$$\mathbb{P}(A_i|H) = \frac{\mathbb{P}(A_i) \cdot \mathbb{P}(H|A_i)}{\sum_j \mathbb{P}(A_j) \cdot \mathbb{P}(H|A_j)}.$$

Interpretation. The quantity $\mathbb{P}(A_i|H)$ is the conditional probability that "given that the effect / result H occurs, it is due to Cause / Reason i".

Example 3.1. In Example 2.1, let us compute the conditional probability that the person is HIV positive given that he/she is tested positive? This is given by

$$\mathbb{P}(A_1|H) = \frac{\mathbb{P}(A_1) \cdot \mathbb{P}(H|A_1)}{\mathbb{P}(H)}.$$

Since we have already computed $\mathbb{P}(H)$ via the law of total probability, we directly substitute the numbers here:

$$\mathbb{P}(A_1|H) = \frac{0.0001 \times 0.9}{0.05} = 0.0018.$$

In other words, this is suggesting that even the test result is positive, it is rather unlikely that the the person has HIV. Therefore, the test is of no value. This is a surprising result, as by a first glance on the assumptions, there is no clue suggesting that the test is useless. It is after performing the probabilistic analysis (i.e. computing $\mathbb{P}(A_1|H)$) that we reach such a conclusion (a good example of critical thinking with mathematics). To understand deeper about what is going on, let us recall the equation

$$
\begin{aligned}
\mathbb{P}(A_1|H) &= \frac{\mathbb{P}(A_1)\cdot\mathbb{P}(H|A_1)}{\mathbb{P}(A_1)\cdot\mathbb{P}(H|A_1)+\mathbb{P}(A_1)\cdot\mathbb{P}(H|A_1)} \\
&= \frac{0.0001\times 0.9}{0.0001\times 0.9+0.9999\times 0.05} \\
&= \frac{1}{1+\frac{0.9999\times 0.05}{0.0001\times 0.9}}. \qquad (3.1)
\end{aligned}
$$

The main problem here is that, the number 0.0001 (the HIV rate) is extremely small, which is of course the nature of any rare disease. This results in the fact that the term 0.0001×0.9 is much small than the term 0.9999×0.5 in the denominator, leading to the smallness of $\mathbb{P}(A_1|H)$ in view of the expression (3.1). The expression (3.1) also suggests that, in order to make the test useful, we need the number 0.05 (the error rate) to be much smaller or the number 0.9 (the accuracy) to be much larger. In concise words, the test needs to be much more accurate than it is in order to be effective.