Z. Abel, B. Chapman, E. Demaine
revised Sunday 28th April, 2024

Lecture 20: Independence

1 Basic definitions

Suppose I flip a fair coin ten times and get tails every time. Am I now "due for" a heads, if I flip it again? No! (This mistaken idea is called the "gambler's fallacy" and is surprisingly common in real life). Each flip of a fair coin has an equal chance of giving heads or tails, regardless of the previous outcomes. This concept is captured by the mathematical notion of *independence*.

Definition 1. *Events A, B are independent if*

$$\Pr[A \mid B] = \Pr[A] \quad \textit{or} \quad \Pr[B] = 0.$$

(In the case $\Pr[B] = 0$*, the conditional probability* $\Pr[A \mid B]$ *is not defined.) Alternatively and equivalently, A and B are independent if*

$$\Pr[A \cap B] = \Pr[A] \cdot \Pr[B].$$

A few examples:

1. If A and B are *disjoint* events with nonzero probabilities, then they cannot be independent! For $\Pr[A \cap B] = 0$, but $\Pr[A] \cdot \Pr[B] \neq 0$. Intuitively, knowing that A happened *yields information* about whether B happened.

2. Suppose we flip two fair coins. Let A be the event that the first comes out heads and B the event the second comes out heads. These are independent—in fact this was implicitly part of our definition of a *fair* coin that coin flips are independent. In reality, with physical coins, it need not be the case! See e.g. https://arxiv.org/abs/2310.04153.

3. Suppose again we flip two coins. Let A be the event that the first comes out heads, and B the event that both are heads. These events are *not* independent, for

   $$\Pr[A] = \frac{1}{2}, \Pr[B] = \frac{1}{4}, \Pr[A \cap B] = \frac{1}{4}.$$

   Indeed, the events $A \cap B$ and B are actually identical: knowing that B happened tells us that A happened with certainty.

4. Once again, we flip two fair coins. Let A be the event that the first comes out heads, and B the event that both have the same result (heads or tails). These events *are* independent:
$$\Pr[A] = \frac{1}{2}, \Pr[B] = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}, \Pr[A \cap B] = \frac{1}{4}.$$

5. We flip two coins which are independent but *biased*: each coin has a probability p of coming up heads and $1-p$ of tails. The events A and B are as in the previous example. Now they are *not* independent:
$$\Pr[A] = p, \Pr[B] = p^2 + (1-p)^2, \Pr[A \cap B] = p^2.$$
Or alternatively, using conditional probability:
$$\Pr[B \mid A] = p,$$
since if A occurs, then the only way the two coins can agree is for both to be heads. If p is very small, then $\Pr[B \mid A] \approx 0$ but $\Pr[B] \approx 1$. This is an example where the original definition in terms of conditional probabilities is nicer to work with than the product definition.

2 Multiple events: mutual and pairwise independence

What if we have 3 events A, B, C? We say that these events are *mutually independent* if

$$\begin{aligned} \Pr[A \cap B] &= \Pr[A]\Pr[B] \\ \Pr[A \cap C] &= \Pr[A]\Pr[C] \\ \Pr[B \cap C] &= \Pr[B]\Pr[C] \\ \Pr[A \cap B \cap C] &= \Pr[A]\Pr[B]\Pr[C] \end{aligned}$$

Is the last equation really necessary? Perhaps we can derive it from the first three conditions? The answer is no, as demonstrated by the following example. Suppose we flip three fair independent coins, and let A be the event that coins 1 and 2 agree, B the event that coins 2 and 3 agree, and C the event that coins 3 and 1 agree.

$$\begin{aligned} \Pr[A] = \Pr[B] &= \Pr[C] = 1/2 \\ \Pr[A \cap B] = \Pr[A \cap C] &= \Pr[B \cap C] = 1/4, \end{aligned}$$

since the event $A \cap B$ occurs iff all three coins have the same result, which occurs with chance $1/8 + 1/8 = 1/4$ (and likewise for $B \cap C$ and $A \cap C$). So the first three conditions are satisfied. However, the last condition is not:

$$\Pr[A \cap B \cap C] = \Pr[\text{all three coins agree}] = 1/4 \neq \Pr[A]\Pr[B]\Pr[C].$$

Thus, A, B, C are *not* mutually independent.

We say that three events that satisfy the first three conditions (on pairs being independent), but not the fourth condition are *pairwise independent*. Pairwise independence occurs frequently in computer science since it's often a "good enough" substitute for mutual independence, and much easier to achieve.

There are extensions of these definitions to more than 3 events which are straightforward: consult the textbook for the definitions.

3 Independence in reality

Knowing whether events are independent makes a big difference in understanding reality. For instance, take the 2016 election, where Trump's victory was a surprise to many analysts. Someone who was unaware of the importance of independence in probability might have reasoned as follows: for Trump to win the election, he needs to win all three of Pennsylvania, Michigan, and Wisconsin. Polls show that for each state, he has a low chance of winning (0.21 for PA, 0.23 for MI, 0.165 for WI). Thus, the chance he wins all three must be the product of these, which is very small indeed (0.008).

This reasoning is incorrect because the events are not independent! There are many reasons they could be correlated, such as systematic errors in the polls, or election-day events and changes in "voter morale" of each party (for instance, early exit poll results from one state may convince voters from another state to show up to the polls).

Without any prior assumptions on correlations, what's the highest upper bound we can place on Trump winning all three states? The best we can say is 0.165 (the chance of the least likely individual event). It is consistent with the information at hand that $\Pr[PA \cap MI \cap WI] = \Pr[WI]$. Note that it cannot be greater, since the event $PA \cap MI \cap WI$ is a subset of the event WI.

For a more detailed calculation that tries to estimate the correlations and the probability of all three states being won by Trump, check out this discussion: https://chance.amstat.org/2018/11/epic-fail/.

4 Conditional independence

A mistaken intuition is that independence is related to causality: if A and B are not independent, then A must cause B or vice versa. This is not so! In fact, dependence can arise between two independent, causally unrelated events due to *conditioning* on a third event.

First, a mathematical definition: we say A and B are independent *given* C (or *conditioned on* C), if

$$\Pr[A \mid C] \Pr[B \mid C] = \Pr[A \cap B \mid C].$$

This is the "right" definition to choose since it reduces to our definition of independence from the start of lecture, applied to the probability space obtained by conditioning our original probability space on C.

As a specific example, consider a (drastically oversimplified) model of dating. Suppose that every person can be either attractive or not attractive with probability 1/2 each, and nice or not nice with probability 1/2 each. Suppose moreover that these are independent: defining events A for attractive and N for nice, we have that A and N are independent, and $\Pr[A] = \Pr[N] = 1/2$.

Now, suppose we condition on the event R of being "romantically interesting": let's suppose that anyone who is either attractive or nice (or both) is interesting. (Again, a drastically oversimplified assumption!) Are the events A and N independent given R? They are not!

$$\begin{aligned}\Pr[A \mid R] &= \frac{\Pr[A \cap R]}{\Pr[R]} = \frac{1/2}{3/4} = 2/3 \\ \Pr[N \mid R] &= \frac{\Pr[N \cap R]}{\Pr[R]} = \frac{1/2}{3/4} = 2/3 \\ \Pr[A \cap N \mid R] &= \frac{\Pr[A \cap N \cap R]}{\Pr[R]} = \frac{1/4}{3/4} = 1/3.\end{aligned}$$

In fact, it appears that attractiveness is *anticorrelated* with niceness:

$$\Pr[A \mid R \cap N] = 1/2 < \Pr[A \mid R] = 2/3.$$

But there's no causal relationship between attractiveness and niceness! The effect is purely caused by the "filter" applied by conditioning on R.

5 Odds and ends

The Birthday Principle (Paradox?) In last week's recitation, you found the probability of a group of d students, each of whose birthdays are independently and uniformly chosen from n possibilities, have no two students with the same birthday, is equal to

$$\Pr[\text{no equal birthdays}] = \frac{n-1}{n}\frac{n-2}{n}\cdots\frac{n-(d-1)}{n}.$$

Let us upper-bound this quantity to get an estimate of how many students we need before a birthday "collision" becomes likelier than not. Using the inequality $1 - x \le e^{-x}$ for $x > 0$ (provable using calculus), write

$$\begin{aligned}\Pr[\text{no equal birthdays}] &= \frac{n-1}{n}\frac{n-2}{n}\cdots\frac{n-(d-1)}{n}. \quad = \left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{d-1}{n}\right) \\ &\le e^{-1/n}e^{-2/n}\cdots e^{-(d-1)/n} \\ &= e^{-\frac{d(d-1)}{2n}}.\end{aligned}$$

This becomes small when $d(d-1)$ is about as big as n, or when $d \approx \sqrt{n}$. We can see that for $n = 365$, we need $d = 23$ to get this bound to be below 1/2. So for 23 students whose birthdays are distributed according to the assumptions, it's more likely than not that there's a pair with the same birthday!

This "square root" scaling is counterintuitive and important in computer science in several applications (hashing, cryptography, testing random data, etc.).

Revisiting the Gambler's Fallacy. If you actually flipped a coin 50 times and got heads every time, would you really guess 50-50 for the next flip? No! Because you'd start suspecting that the coin is biased! In fact, the common-sensically "rational" thing to do is the guess that after 50 heads, the next flip will *also* be heads—the precise opposite of the gambler's fallacy.

One way to formalize this is using Bayes' rule. At the start, you would assign a large prior probability that the coin is fair, and a small probability that it is perfectly biased. The more successive heads you see, the higher you should weight the probability that the coin is actually biased.

$$\begin{aligned}\Pr[H \mid H^{50}] &= \Pr[H \mid \text{biased} \cap H^{50}] \Pr[\text{biased} \mid H^{50}] + \Pr[H \mid \text{fair} \cap H^{50}] \Pr[\text{fair} \mid H^{50}] \\ &= 1 \cdot \Pr[\text{biased} \mid H^{50}] + \frac{1}{2} \cdot \Pr[\text{fair} \mid H^{50}].\end{aligned}$$

Indeed, much work has been done on trying to give a formal theory of "reasoning under uncertainty" using Bayesian methods. These perspectives have also been quite influential in AI and machine learning.