> # Topic 2 — Multiple events, conditioning, and independence

2.1 Two or more events on the same sample space

Frequently we end up dealing with multiple events $A, B, \ldots$ defined on a single sample space Ω. For instance, Ω might denote the possible outcomes of picking a card from a deck, and the events of interest could include $A = \{\omega : \omega \text{ is a number}\}$, $B = \{\omega : \omega \text{ is a spade}\}$, and so on. Instead of just the single event probabilities $\Pr(A)$ and $\Pr(B)$, we are also interested in probabilities of combinations of events, such as $\Pr(A \cap B)$ and $\Pr(A \cup B)$.

2.1.1 Human intuition about combined event probabilities is fallible

Let's start with Exercise 1.2.25 from the textbook. The famous psychologists Kahneman and Tversky spent a lot of time trying to understand the nature of human rationality and decision making. For instance, do humans think probabilistically? In one experiment, they told subjects the following:

> Linda is 31, single, outspoken, and very bright. She majored in philosophy in college. As a student, she was deeply concerned with racial discrimination and other social issues, and participated in anti-nuclear demonstrations.

and then asked them to assess the likelihood of three events:

(A) Linda is active in the feminist movement.

(B) Linda is a bank teller.

(C) Linda is a bank teller who is active in the feminist movement.

A significant majority of respondents deemed A to be the most likely, followed by C, and lastly B. But $\Pr(C) = \Pr(A \cap B) \leq \Pr(B)$! The respondents exhibited what is called a *conjunction fallacy*.

2.1.2 Joint probability tables

One way to write down the probabilities of pairs of events is in a table of all possible outcomes. Let's take another exercise from the textbook, 1.2.15. Here two students, John and Mary, are taking a math course in which the possible grades are A, B, and C. The sample space of possible outcomes can be written as $\Omega = \{A, B, C\}^2$, where the first coordinate of $\omega \in \Omega$ is John's grade and the second is Mary's grade. We can write down the probabilities of the nine possible outcomes in a 3×3 table:

		Mary		
		A	B	C
	A	p_{AA}	p_{AB}	p_{AC}
John	B	p_{BA}	p_{BB}	p_{BC}
	C	p_{CA}	p_{CB}	p_{CC}

The particular problem (see book) can be solved by writing the evidence in terms of these probabilities and adding and subtracting appropriate events.

2.1.3   The union bound

As another example, suppose that in a city,

> 60% of people have a car,
>
> 20% of people have a bicycle, and
>
> 10% of people have a motorcycle.

Anybody who lacks all three ends up walking to work, and nobody walks unless they have to. Based on this data, what fraction of people walk to work?

We don't have enough information to give an exact answer. This is because we don't know about the overlap in ownership of different vehicles: what fraction of the people with cars also have bicycles, and so on. Nonetheless, we can give upper and lower bounds. To get started, consider the experiment of picking a random person in the city, and let Ω be the space of outcomes. We then define the four events of interest: $C = \{\text{owns a car}\}$, $B = \{\text{owns a bicycle}\}$, $M = \{\text{owns a motorcycle}\}$, and $W = \{\text{walks}\}$. Since $W = \Omega - (C \cup B \cup M)$, we need to bound $\Pr(C \cup B \cup M)$.

The least $\Pr(C \cup B \cup M)$ could possibly be is 0.6, which occurs when everyone who has a bicycle or motorcycle also owns a car. The most $\Pr(C \cup B \cup M)$ could be is 0.9, which happens when the sets C, B, M are disjoint. Therefore $0.1 \leq \Pr(W) \leq 0.4$.

Generalizing, we get the extremely useful **union bound**: for any events $A_1, A_2, \ldots, A_k$,

$$\Pr(A_1 \cup \cdots \cup A_k) \ \leq \ \Pr(A_1) + \cdots + \Pr(A_k).$$

2.2   Balls in bins, or urn problems

You have m indistinguishable balls and in front of you is a row of n bins. You place each ball into a bin chosen at random.

Let's write the sample space as $\Omega = \{1, 2, \ldots, m\}^n$; in each outcome $\omega = (\omega_1, \ldots, \omega_n)$, the value ω_i represents the number of balls in the ith bin.

Here are some interesting questions one can ask.

1. What is the chance that the ith bin is empty? (What is $\Pr(\omega_i = 0)$?)

   Well, there are m balls, and we want every single one of them to miss the ith bin. The probability that the first ball misses this bin is $(n-1)/n = 1 - 1/n$. The probability that the second ball misses is also $1 - 1/n$, as with the third, and fourth, and so on. Therefore,

   $$\Pr(i\text{th bin empty}) \ = \ \left(1 - \frac{1}{n}\right)^m \ \leq \ (e^{-1/n})^m \ = \ e^{-m/n},$$

   where we've used the formula $e^x \geq 1 + x$ that we discussed earlier.

2. What is the chance that there is an empty bin if $m = n$?

   Intuitively, one would expect this to be pretty close to 1 (that is, pretty much certain), because the complementary event – no empty bins – would occur only if every single bin received exactly one ball. Let's analyze this latter probability.

   $$\Pr(\text{every bin gets a ball}) \ = \ \left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right)\cdots\frac{1}{n} \ = \ \frac{n!}{n^n}.$$

   This is miniscule; for instance, it is less than $1/2^{n/2}$.

3. What is the chance that there is an empty bin if $m = 2n \ln n$?

   When m is increased from n to $2n \ln n$, the chance of an empty bin drops from ≈ 1 to ≈ 0. To see this, let A_i be the event that the ith bin is empty. We proved above that $\Pr(A_i) \leq \exp(-m/n) = \exp(-2 \ln n) = 1/n^2$. Therefore

$$\Pr(\text{some bin is empty}) \;=\; \Pr(A_1 \cup \cdots \cup A_n) \;\leq\; \Pr(A_1) + \cdots + \Pr(A_n) \;\leq\; n \cdot \frac{1}{n^2} \;=\; \frac{1}{n}$$

   (the first inequality is the union bound). Thus with high probability (at least $1 - 1/n$), every bin gets at least one ball.

4. *The coupon collector problem.* Many probability questions turn out to be thinly disguised balls-and-bins problems. Here's an example. Suppose that each cereal box contains one of k action figures (chosen uniformly at random). How many cereal boxes should you buy in order to collect all k figures?

   If you make the following associations:

$$
\begin{aligned}
\text{bin} &\equiv \text{action figure} \\
\text{throw a ball} &= \text{buy a box of cereal} \\
\text{every bin gets a ball} &= \text{you collect all the figures,}
\end{aligned}
$$

   then you see that what we are really asking is, how many balls do you need to throw into $n = k$ bins in order to hit all of them? And we've already seen that $m = 2k \ln k$ balls (cereal boxes) will suffice.

5. What is the probability that some bin gets two or more balls?

   This is another situation in which it is easier to study the complementary event, that every bin gets at most one ball.

$$\Pr(\text{every bin gets 0 or 1 ball}) \;=\; \left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right) \cdots \left(1 - \frac{m-1}{n}\right) \;=\; \frac{n!}{(n-m)! \, n^m}.$$

   Although this is exact, it is a little difficult to fathom, and so let's try an approximation instead, using $1 - i/n \leq e^{-i/n}$:

$$
\begin{aligned}
\Pr(\text{every bin gets 0 or 1 ball}) \;&=\; \left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right) \cdots \left(1 - \frac{m-1}{n}\right) \\
&\leq\; \exp(-1/n)\exp(-2/n) \cdots \exp(-(m-1)/n) \\
&=\; \exp\left(-\frac{1}{n}\left(1 + \cdots + (m-1)\right)\right) \\
&=\; \exp\left(-\frac{m(m-1)}{2n}\right).
\end{aligned}
$$

   This is an excellent approximation when $m \ll n$. It tells us that if $m \geq c\sqrt{n}$ (for some small constant c), then this probability is at most $1/2$. That is, if you throw (roughly) $\sqrt{n}$ (or more) balls into n bins, then chances are that some bin will get two or more balls.

6. *Birthday paradox.* We saw earlier than in a room with 23 people, chances are two of them will share the same birthday.

   This is yet another balls-and-bins problem:

$$
\begin{aligned}
\text{bin} &\equiv \text{day of the year} \\
\text{throw a ball} &\equiv \text{select birthday of a person in the room} \\
\text{bin with 2 balls} &\equiv \text{two people with the same birthday.}
\end{aligned}
$$

The number of bins is $n = 365$ and the number of balls is $m = 23$. Using the formula we just obtained,

$$\Pr(\text{two people share the same birthday}) \;=\; 1 - \Pr(\text{every bin has at most one ball})$$

$$\geq \;\; 1 - \exp\left(-\frac{m(m-1)}{2n}\right) \;\; \geq \;\; \frac{1}{2}.$$

7. *Coin tossing.*

   Even the tossing of a fair coin, m times, can be modeled using balls and bins. Imagine there are just $n = 2$ bins (call one bin H and the other bin T).

2.3   Conditional probability

You meet a stranger, a random citizen of the United States. What's the chance that he (or, equally likely, she) votes for the Democratic party? Who knows, but it's probably close to 50%.

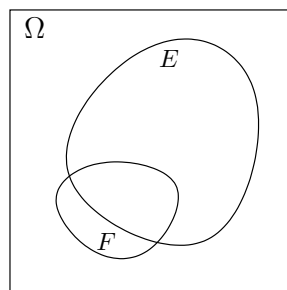$$\Omega = \{\text{citizens of the US}\}, \quad E = \{\omega \in \Omega : \text{votes Dem}\}.$$

Now what if you find out a little bit more information: this stranger likes spicy food.

$$F = \{\omega \in \Omega : \text{likes spicy food}\}.$$

How does change the odds? What is $\Pr(E|F)$ ("probability of E given F")? Again, it's hard to say, but for illustrative purposes here is a (definitely incorrect) set of probabilities:

$$\Pr(E) = 0.5, \quad \Pr(F) = 0.2, \quad \Pr(E \cap F) = 0.15.$$

The Venn diagram for this scenario is

If you stare at it a little bit, you will probably conclude, correctly, that $\Pr(E|F)$ is the chance that a point drawn from the small blob lies inside the large blob, namely $0.15/0.2 = 0.75$. By reasoning this way, you are implicitly using **the formula for conditional probability**

$$\Pr(E|F) \;=\; \frac{\Pr(E \cap F)}{\Pr(F)}.$$

We will make heavy use of this and of the equivalent

$$\Pr(E \cap F) = \Pr(E|F)\Pr(F).$$

2.3.1   A few examples

1. *Pregnancy test.* How accurate is a pregnancy test? Here we make up some numbers and compute the relevant conditional probabilities.

   The sample space in this case consists of women who use the test; call this Ω. There are two events on this space that we care about: $P = \{\text{actually pregnant}\}$ and $T = \{\text{test says pregnant}\}$. Suppose that the following are determined (warning: these are fabricated!):

   $$T \subset P, \quad \Pr(P) = 0.4, \quad \Pr(T) = 0.3.$$

   There are two conditional probabilities we'd like to compute.

   - What is the chance of pregnancy if the test comes out positive?
     Well, since $T \subset P$, we have $\Pr(P|T) = 1.0$. Algebraically, $\Pr(P \cap T) = \Pr(T)$, so $\Pr(P|T) = \Pr(P \cap T)/\Pr(T) = 1.0$.
   - What is the chance of pregnancy if the test comes out negative?
     Let $T^c = \Omega - T$ be the event that the test is negative. Then $\Pr(T^c) = 1 - \Pr(T) = 0.7$ and $\Pr(P \cap T^c) = \Pr(P) - \Pr(P \cap T) = 0.4 - 0.3 = 0.1$. These are the two probabilities we need.

     $$\Pr(P|T^c) \;=\; \frac{\Pr(P \cap T^c)}{\Pr(T^c)} \;=\; \frac{0.1}{0.7} \;\approx\; 0.14.$$

2. *Rolls of a die.* You roll a die twice. What is the probability that the sum of the two rolls is ≥ 10 if:

   - The first roll is 6?
     We could use the conditional probability formula, but that seems like overkill in so straightforward a situation.
     $$\Pr(\text{sum} \geq 10 \mid \text{first} = 6) \;=\; \Pr(\text{second} \geq 4) \;=\; \frac{1}{2}.$$

   - The first roll is ≥ 3?
     Okay, this is not so trivial anymore. The sample space is $\Omega = \{1,2,3,4,5,6\}^2$, with each of the 36 outcomes equally likely.

     $$\begin{aligned}
     \Pr(\text{sum} \geq 10 \mid \text{first} \geq 3) &= \frac{\Pr(\text{sum} \geq 10 \text{ AND first} \geq 3)}{\Pr(\text{first} \geq 3)} \\
     &= \frac{\Pr(\{(4,6),(5,5),(5,6),(6,4),(6,5),(6,6)\})}{4/6} = \frac{6/36}{4/6} = \frac{1}{4}.
     \end{aligned}$$

   - The first roll is < 6?

     $$\begin{aligned}
     \Pr(\text{sum} \geq 10 \mid \text{first} < 6) &= \frac{\Pr(\text{sum} \geq 10 \text{ AND first} < 6)}{\Pr(\text{first} < 6)} \\
     &= \frac{\Pr(\{(5,5),(5,6),(4,6)\})}{5/6} = \frac{3/36}{5/6} = \frac{1}{10}.
     \end{aligned}$$

3. *To be invisible or to fly?* A few years ago, the question was going round: "which super power would you rather possess: the ability to make yourself invisible or to fly?" It was perhaps an urban legend that men usually wanted to fly while women usually wanted to be invisible; and there was much speculation

about the psychological reasons for this. But suppose the following statistics were obtained (warning: these numbers are fabricated!):

$$\Pr(\text{fly}|\text{male}) = 0.8, \quad \Pr(\text{invisible}|\text{female}) = 0.6.$$

Then what is the overall fraction of people who would choose the ability to fly?

This cannot be answered unless it is known what fraction of the population is male and what fraction is female. Suppose it is 60-40. Then

$$
\begin{aligned}
\Pr(\text{fly}) \;&=\; \Pr(\text{fly AND male}) + \Pr(\text{fly AND female}) \\
&=\; \Pr(\text{fly}|\text{male})\Pr(\text{male}) + \Pr(\text{fly}|\text{female})\Pr(\text{female}) \\
&=\; (0.8)(0.6) + (0.4)(0.4) \;=\; 0.64.
\end{aligned}
$$

That is, 64% would like to fly.

2.3.2   The summation rule

In the last example, we implicitly used a **summation formula**. Suppose $A_1, \ldots, A_k \subset \Omega$ are disjoint events whose union is Ω (that is, exactly one of them will occur). Then for any event $E \subset \Omega$,

$$\Pr(E) \;=\; \sum_{i=1}^{k} \Pr(E \cap A_i) \;=\; \sum_{i=1}^{k} \Pr(E|A_i)\Pr(A_i).$$

(If the A_i are not disjoint, simply replace the first equality with a $\leq$.) We will often write $\Pr(E \cap A_i)$ as $\Pr(E, A_i)$.

2.3.3   The Monty Hall problem

This probability puzzle is weakly related to a game on an old TV show called *Let's Make a Deal*, and has been renamed after the host of that show. The host brings the game player to a room with three closed doors. One of the doors leads to a treasure chest while the other two doors each lead to a goat. The player picks a door (at random, presumably), hoping for the best. Now something interesting happens. Instead of opening that door, the host opens one of the other two doors to reveal a goat. The player is then allowed to either stick to his original guess, or to switch to the other unopened door. Which should he do?

In surveys, the majority of people feel intuitively that it doesn't make a difference. They reason that there are two unopened doors, and the treasure could be behind either of them, so each has a 50-50 chance of being the lucky door. But this is incorrect. The player should switch to the other door: by doing so, he will double his chances of getting the treasure! A conditional probability calculation shows why:

$$
\begin{aligned}
\Pr(\text{treasure in other door}) \;=\;\; & \Pr(\text{treasure in other door}|\text{initial guess correct})\Pr(\text{initial guess correct}) + \\
& \Pr(\text{treasure in other door}|\text{initial guess wrong})\Pr(\text{initial guess wrong}) \\
=\;\; & 0 \cdot \frac{1}{3} + 1 \cdot \frac{2}{3} \;=\; \frac{2}{3}.
\end{aligned}
$$

2.3.4   Another summation rule

There is also a summation rule for conditional probabilities. Suppose $A_1, \ldots, A_k$ are disjoint events whose union is Ω; and let $E, F \subset \Omega$ be any two events. Then

$$\Pr(E|F) \;=\; \sum_{i=1}^{k} \Pr(E \cap A_i | F) \;=\; \sum_{i=1}^{k} \Pr(E|A_i, F)\Pr(A_i|F).$$

(It's just like the regular summation rule, but operating within F rather than Ω.)

2.3.5   Sex bias in graduate admissions

In 1973, the dean of graduate admissions at Berkeley noticed something alarming: a male applicant had a higher chance of being admitted than a female applicant! Specifically,

$$\text{Pr}(\text{admitted}|\text{male}) = 0.44, \quad \text{Pr}(\text{admitted}|\text{female}) = 0.35.$$

This is a significant difference given that the sample space (applicants to graduate programs at Berkeley) was very large, of size 12,763. The dean wondered, which departments were responsible for this bias?

To determine this, data was collected for each of the departments on campus. And it was found that no department (with a few insignificant exceptions) had any bias against females! Specifically, for each department d,

$$\text{Pr}(\text{admitted}|\text{male, department} = d) \leq \text{Pr}(\text{admitted}|\text{female, department} = d).$$

How could this be?

Let's go back to the summation formulas for conditional probabilities.

$$
\begin{aligned}
\text{Pr}(\text{admit}|\text{male}) &= \sum_d \text{Pr}(\text{admit}|\text{male, dept} = d)\text{Pr}(\text{dept} = d|\text{male}) \\
\text{Pr}(\text{admit}|\text{female}) &= \sum_d \text{Pr}(\text{admit}|\text{female, dept} = d)\text{Pr}(\text{dept} = d|\text{female})
\end{aligned}
$$

The answer lies in the second term. Some departments (engineering) are easier to get into than others (humanities). Males tended to apply to the easier departments, while females on average applied to the harder departments. So while no individual department had bias, on the whole it looked like women had a lower probability of acceptance.

2.4   Bayes' rule

The following example is taken from *Probabilistic Reasoning in Intelligent Systems* by Judea Pearl:

> You wake up in the middle of the night to the shrill sound of your burglar alarm. What is the chance that a burglary attempt has taken place?

The relevant facts are:

- There is a 95% chance that an attempted burglary attempt will trigger the alarm. That is,

$$\text{Pr}(\text{alarm}|\text{burglary}) = 0.95.$$

- There is a 1% chance of a false alarm.

$$\text{Pr}(\text{alarm}|\text{no burglary}) = 0.01.$$

- Based on local crime statistics, there is a one-in-10,000 chance that a house will be burglarized on a given night.

$$\text{Pr}(\text{burglary}) = 10^{-4}.$$

We are interested in the chance of a burglary given that the alarm has sounded. We can use the conditional probability formula for this:

$$\Pr(\text{burglary}|\text{alarm}) \ = \ \frac{\Pr(\text{burglary, alarm})}{\Pr(\text{alarm})} \ = \ \frac{\Pr(\text{alarm}|\text{burglary})\Pr(\text{burglary})}{\Pr(\text{alarm})}.$$

The one term we don't immediately know is $\Pr(\text{alarm})$. By the summation rule,

$$\Pr(\text{alarm}) \ = \ \Pr(\text{alarm}|\text{burglary})\Pr(\text{burglary}) + \Pr(\text{alarm}|\text{no burglary})\Pr(\text{no burglary}).$$

Putting it all together,

$$\Pr(\text{burglary}|\text{alarm}) \ = \ \frac{0.95 \times 10^{-4}}{0.95 \times 10^{-4} + 0.01 \times (1 - 10^{-4})} \ = \ 0.00941,$$

about 0.94%. Thus our belief in a burglary has risen approximately a hundredfold from its default value of 10^{-4}, on account of the alarm.

It is frequently the case, as in this example, that we wish to update the chances of an event H based on new evidence E. In other words, we wish to know $\Pr(H|E)$. The derivation above implicitly uses the following formula, called **Bayes' rule**:

$$\Pr(H|E) \ = \ \frac{\Pr(E|H)\Pr(H)}{\Pr(E)}.$$

As another example, let's look at the Three Prisoner's Paradox, which is actually just a reformulation of the Monty Hall problem. The story goes that there are three prisoners A, B, C in a jail, and one of them is to be declared guilty and executed the following morning. As the night progresses, prisoner A is racked with worry, and calls the prison guard over. He wants to know whether he is the unlucky one. The guard replies, "I am not allowed to tell you whether you will be declared guilty. But I can say that prisoner B will be declared innocent." Prisoner A thinks about this for a little while and then starts worrying even more. Before he asked the question, it seemed like his chances of dying were one-in-three. But after his innocuous question, the chance seems to have risen to one-in-two!

Actually, Prisoner A's chances are still one in three. The two events of interest are

$$G_A \ = \ A \text{ will be declared guilty}$$
$$I_B \ = \ \text{the guard, when prompted, will declare } B \text{ to be innocent.}$$

Using the summation rule,

$$\Pr(I_B) \ = \ \Pr(I_B|G_A)\Pr(G_A) + \Pr(I_B|G_A^c)\Pr(G_A^c) \ = \ \frac{1}{2} \cdot \frac{1}{3} + \frac{1}{2} \cdot \frac{2}{3} \ = \ \frac{1}{2}.$$

Therefore, by Bayes' rule,

$$\Pr(G_A|I_B) \ = \ \frac{\Pr(I_B|G_A)\Pr(G_A)}{\Pr(I_B)} \ = \ \frac{\frac{1}{2} \cdot \frac{1}{3}}{\frac{1}{2}} \ = \ \frac{1}{3}.$$

2.5   Independence

Two events are called *independent* if the outcome of one (that is, whether or not it occurs) does not affect the probability that the other will occur. For instance, suppose you flip two fair coins. The outcome of either coin does not influence the other; therefore the two outcomes are independent.

Formally, we say events A and B (defined on some sample space Ω) are independent if

$$\Pr(A \cap B) \;=\; \Pr(A)\Pr(B).$$

Can you show that this definition implies the following?

- $\Pr(A|B) = \Pr(A)$.

- $\Pr(B|A) = \Pr(B)$.

- $\Pr(A|B^c) = \Pr(A)$.

In the following examples, are events A and B independent or not?

1. You have two children. $A = \{$first child is a boy$\}$, $B = \{$second child is a girl$\}$.

2. You throw two dice. $A = \{$first is a 6$\}$, $B = \{$sum $> 10\}$.

3. You get dealt two cards at random from a deck of 52. $A = \{$first is a heart$\}$, $B = \{$second is a club$\}$.

4. Same sample space. $A = \{$first is a heart$\}$, $B = \{$second is a 10$\}$.

5. The three scenarios depicted in these Venn diagrams.