Conditional Probability

A pharmaceutical company is marketing a new test for a certain medical condition. According to clinical trials, the test has the following properties:

1. When applied to an affected person, the test comes up positive in 90% of cases, and negative in 10% (these are called "false negatives").
2. When applied to a healthy person, the test comes up negative in 80% of cases, and positive in 20% (these are called "false positives").

Suppose that the incidence of the condition in the US population is 5%. When a random person is tested and the test comes up positive, what is the probability that the person actually has the condition? (Note that this is presumably *not* the same as the simple probability that a random person has the condition, which is just $\frac{1}{20}$.)

This is an example of a conditional probability: we are interested in the probability that a person has the condition (event A) *given that* he/she tests positive (event B). Let's write this as $\Pr[A|B]$.

How should we compute $\Pr[A|B]$? Well, since event B is guaranteed to happen, we need to look not at the whole sample space Ω, but at the smaller sample space consisting only of the sample points in B. What should the probabilities of these sample points be? If they all simply inherit their probabilities from Ω, then the sum of these probabilities will be $\sum_{\omega\in B}\Pr[\omega] = \Pr[B]$, which in general is less than 1. So we need to *scale* the probability of each sample point by $\frac{1}{\Pr[B]}$. I.e., for each sample point $\omega \in B$, the new probability becomes

$$\Pr[\omega|B] = \frac{\Pr[\omega]}{\Pr[B]}.$$

Now it is clear how to compute $\Pr[A|B]$: namely, we just sum up these scaled probabilities over all sample points that lie in both A and B:

$$\Pr[A|B] = \sum_{\omega\in A\cap B}\Pr[\omega|B] = \sum_{\omega\in A\cap B}\frac{\Pr[\omega]}{\Pr[B]} = \frac{\Pr[A\cap B]}{\Pr[B]}.$$

Definition 16.1 (conditional probability): For events A,B in the same probability space, such that $\Pr[B] > 0$, the conditional probability of A given B is

$$\Pr[A|B] = \frac{\Pr[A\cap B]}{\Pr[B]}.$$

Let's go back to our medical testing example. The sample space here consists of all people in the US — denote their number by N (so $N \approx 250$ million). The population consists of four disjoint subsets:

TP: the true positives (90% of $\frac{N}{20} = \frac{9N}{200}$ of them);

FP: the false positives (20% of $\frac{19N}{20} = \frac{19N}{100}$ of them);

TN: the true negatives (80% of $\frac{19N}{20} = \frac{76N}{100}$ of them);

FN: the false negatives (10% of $\frac{N}{20} = \frac{N}{200}$ of them).

Now let A be the event that a person chosen at random is affected, and B the event that he/she tests positive. Note that B is the union of the disjoint sets TP and FP, so

$$|B| = |TP| + |FP| = \tfrac{9N}{200} + \tfrac{19N}{100} = \tfrac{47N}{200}.$$

Thus we have

$$\Pr[A] = \tfrac{1}{20} \qquad \text{and} \qquad \Pr[B] = \tfrac{47}{200}.$$

Now when we condition on the event B, we focus in on the smaller sample space consisting only of those $\frac{47N}{200}$ individuals who test positive. To compute $\Pr[A|B]$, we need to figure out $\Pr[A \cap B]$ (the part of A that lies in B). But $A \cap B$ is just the set of people who are both affected and test positive, i.e., $A \cap B = TP$. So we have

$$\Pr[A \cap B] = \frac{|TP|}{N} = \frac{9}{200}.$$

Finally, we conclude from Definition 17.1 that

$$\Pr[A|B] = \frac{\Pr[A \cap B]}{\Pr[B]} = \frac{9/200}{47/200} = \frac{9}{47} \approx 0.19.$$

This seems bad: if a person tests positive, there's only about a 19% chance that he/she actually has the condition! This sounds worse than the original claims made by the pharmaceutical company, but in fact it's just another view of the same data.

[Incidentally, note that $\Pr[B|A] = \frac{9/200}{1/20} = \frac{9}{10}$; so $\Pr[A|B]$ and $\Pr[B|A]$ can be very different. Of course, $\Pr[B|A]$ is just the probability that a person tests positive given that he/she has the condition, which we knew from the start was 90%.]

To complete the picture, what's the (unconditional) probability that the test gives a correct result (positive or negative) when applied to a random person? Call this event C. Then

$$\Pr[C] = \tfrac{|TP|+|TN|}{N} = \tfrac{9}{200} + \tfrac{76}{100} = \tfrac{161}{200} \approx 0.8.$$

So the test is about 80% effective overall, a more impressive statistic.

But how impressive is it? Suppose we ignore the test and just pronounce everybody to be healthy. Then we would be correct on 95% of the population (the healthy ones), and wrong on the affected 5%. I.e., this trivial test is 95% effective! So we might ask if it is worth running the test at all. What do you think?

Here are a couple more examples of conditional probabilities, based on some of our sample spaces from the previous lecture.

1. **Balls and bins.** Suppose we toss $m = 3$ balls into $n = 3$ bins; this is a uniform sample space with $3^3 = 27$ points. We already know that the probability the first bin is empty is $(1 - \frac{1}{3})^3 = (\frac{2}{3})^3 = \frac{8}{27}$. What is the probability of this event *given that* the second bin is empty? Call these events A, B

respectively. To compute $\Pr[A|B]$ we need to figure out $\Pr[A \cap B]$. But $A \cap B$ is the event that both the first two bins are empty, i.e., all three balls fall in the third bin. So $\Pr[A \cap B] = \frac{1}{27}$ (why?). Therefore,

$$\Pr[A|B] = \frac{\Pr[A \cap B]}{\Pr[B]} = \frac{1/27}{8/27} = \frac{1}{8}.$$

Not surprisingly, $\frac{1}{8}$ is quite a bit less than $\frac{8}{27}$: knowing that bin 2 is empty makes it significantly less likely that bin 1 will be empty.

2. **Dice.** Roll two fair dice. Let A be the event that their sum is even, and B the event that the first die is even. By symmetry it's easy to see that $\Pr[A] = \frac{1}{2}$ and $\Pr[B] = \frac{1}{2}$. Moreover, a little counting gives us that $\Pr[A \cap B] = \frac{1}{4}$. What is $\Pr[A|B]$? Well,

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

In this case, $\Pr[A|B] = \Pr[A]$, i.e., conditioning on B does not change the probability of A.

Independent events

Definition 16.2 (independence): Two events A, B in the same probability space are independent if $\Pr[A|B] = \Pr[A]$.

Note that independence is symmetric: i.e., if $\Pr[A|B] = \Pr[A]$ then it must also be the case that $\Pr[B|A] = \Pr[B]$. To see this, use the definition of conditional probabilities:

$$\Pr[B|A] = \frac{\Pr[A \cap B]}{\Pr[A]} = \frac{\Pr[A \cap B]}{\Pr[B]} \times \frac{\Pr[B]}{\Pr[A]} = \frac{\Pr[A|B]}{\Pr[A]} \times \Pr[B] = \Pr[B].$$

In the last step here, we used our assumption that $\Pr[A|B] = \Pr[A]$. [We are assuming here that $\Pr[A]$ and $\Pr[B]$ are both > 0. Otherwise the conditional probabilities are not defined.]

Examples: In the balls and bins example above, events A, B are *not* independent. In the dice example, events A, B *are* independent.

Knowing that events are independent is very useful, because of the following simple observation:

Theorem 16.1: *If A, B are independent, then* $\Pr[A \cap B] = \Pr[A]\Pr[B]$.

Proof: From the definition of conditional probability we have

$$\Pr[A \cap B] = \Pr[A|B]\Pr[B] = \Pr[A]\Pr[B],$$

where in the second step we have used independence. □

Note that the condition in Theorem 17.1 actually holds if *and only if* A and B are independent. In fact, this condition is often given as the *definition* of independence, rather than the definition we are using.

All the above generalizes to any finite set of events:

Definition 16.3 (mutual independence): Events $A_1, \ldots, A_n$ are mutually independent if for every $1 \le i \le n$ and every subset $I \subseteq \{1, \ldots, n\} - \{i\}$,

$$\Pr[A_i | \textstyle\bigcap_{j \in I} A_j] = \Pr[A_i].$$

I.e., the probability of A_i does not depend on *any combination of* the other events.

Theorem 16.2: *If events $A_1, \ldots, A_n$ are mutually independent, then*

$$\Pr[A_1 \cap \ldots \cap A_n] = \Pr[A_1] \times \Pr[A_2] \times \cdots \times \Pr[A_n].$$

We won't prove this theorem here because it is a special case of the more general Theorem 17.3, which we will prove below (check this!). Note that it is possible to construct three events A, B, C such that each *pair* is independent but the triple A, B, C is *not* mutually independent.

Combinations of events

In most applications of probability in Computer Science, we are interested in things like $\Pr[\bigcup_{i=1}^n A_i]$ and $\Pr[\bigcap_{i=1}^n A_i]$, where the A_i are simple events (i.e., we know, or can easily compute, the $\Pr[A_i]$). The intersection $\bigcap_i A_i$ corresponds to the logical AND of the events A_i, while the union $\bigcup_i A_i$ corresponds to their logical OR. As an example, if A_i denotes the event that a failure of type i happens in a certain system, then $\bigcup_i A_i$ is the event that the system fails.

In general, computing the probabilities of such combinations can be very difficult. In this section, we discuss some situations where it can be done.

Intersections of events

From the definition of conditional probability, we immediately have the following product rule (sometimes also called the chain rule) for computing the probability of an intersection of events.

Theorem 16.3: **[Product Rule]** *For events A, B, we have*

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

More generally, for events $A_1, \ldots, A_n$,

$$\Pr[\bigcap_{i=1}^n A_i] = \Pr[A_1] \times \Pr[A_2|A_1] \times \Pr[A_3|A_1 \cap A_2] \times \cdots \times \Pr[A_n|\bigcap_{i=1}^{n-1} A_i].$$

Proof: The first assertion follows directly from the definition of $\Pr[B|A]$ (and is in fact a special case of the second assertion with $n = 2$).

To prove the second assertion, we will use simple induction on n (the number of events). The base case is $n = 1$, and corresponds to the statement that $\Pr[A] = \Pr[A]$, which is trivially true. For the inductive step, let $n > 1$ and assume (the inductive hypothesis) that

$$\Pr[\bigcap_{i=1}^{n-1} A_i] = \Pr[A_1] \times \Pr[A_2|A_1] \times \cdots \times \Pr[A_{n-1}|\bigcap_{i=1}^{n-2} A_i].$$

Now we can apply the definition of conditional probability to the two events A_n and $\bigcap_{i=1}^{n-1} A_i$ to deduce that

$$\begin{aligned}\Pr[\bigcap_{i=1}^n A_i] = \Pr[A_n \cap (\bigcap_{i=1}^{n-1} A_i)] &= \Pr[A_n|\bigcap_{i=1}^{n-1} A_i] \times \Pr[\bigcap_{i=1}^{n-1} A_i] \\ &= \Pr[A_n|\bigcap_{i=1}^{n-1} A_i] \times \Pr[A_1] \times \Pr[A_2|A_1] \times \cdots \times \Pr[A_{n-1}|\bigcap_{i=1}^{n-2} A_i],\end{aligned}$$

where in the last line we have used the inductive hypothesis. This completes the proof by induction. □

Note that Theorems 17.1 and 17.2 are special cases of the product rule for *independent* events.

Sequences of trials

Many experiments can be viewed as *sequences* of simpler experiments, or *trials*. In these cases, it is more natural to define the probability space in terms of *conditional probabilities*, using the product rule. As an illustration, consider the sample space Ω of n tosses of a biased coin discussed in the previous lecture. We can write Ω as the *product* $\Omega = \Omega_1 \times \Omega_2 \times \ldots \times \Omega_n$, where $\Omega_i = \{H, T\}$ is the sample space of the ith coin toss.[1]

How should we define probabilities in Ω? Each sample point in Ω is an n-tuple $\omega = (\omega_1, \omega_2, \ldots, \omega_n)$, where $\omega_i \in \Omega_i$ is the outcome of the ith toss. Using the product rule, we must have

$$\Pr[\omega] = \Pr[\omega_1] \times \Pr[\omega_2|\omega_1] \times \cdots \times \Pr[\omega_n|\omega_1, \omega_2, \ldots, \omega_{n-1}]. \tag{1}$$

So if we define all the conditional probabilities $\Pr[\omega_i|\omega_1, \omega_2, \ldots, \omega_{i-1}]$, we will in fact have defined the entire probability space!

Now the key fact in this example is that the coin tosses are supposed to be *independent*, so each conditional probability is just the same as the corresponding unconditional probability for a coin. Thus equation (1) becomes

$$\Pr[\omega] = \Pr[\omega_1] \times \Pr[\omega_2] \times \cdots \times \Pr[\omega_n].$$

Thus we see that the probability of any sample point ω is $\Pr[\omega] = p^r(1-p)^{n-r}$, where r is the number of Heads in ω. Of course, this is exactly how we defined this sample space in the previous lecture. However, the point is that now we have a rational basis for our definition: it is an inevitable consequence of the fact that the coin tosses are independent.

Here are some more examples.

1. **Coin tosses.** Toss a fair coin three times. Let A be the event that all three tosses are heads. Then $A = A_1 \cap A_2 \cap A_3$, where A_i is the event that the ith toss comes up heads. We have

   $$\begin{aligned}\Pr[A] &= \Pr[A_1] \times \Pr[A_2|A_1] \times \Pr[A_3|A_1 \cap A_2] \\ &= \Pr[A_1] \times \Pr[A_2] \times \Pr[A_3] \\ &= \tfrac{1}{2} \times \tfrac{1}{2} \times \tfrac{1}{2} = \tfrac{1}{8}.\end{aligned}$$

   The second line here follows from the fact that the tosses are mutually independent. Of course, we already know that $\Pr[A] = \frac{1}{8}$ from our definition of the probability space in the previous lecture. The above is really a check that the space behaves as we expect.[2]

   If the coin is biased with heads probability p, we get, again using independence,

   $$\Pr[A] = \Pr[A_1]\Pr[A_2]\Pr[A_3] = p^3.$$

   And more generally, the probability of any sequence of n tosses containing r heads and $n-r$ tails is $p^r(1-p)^{n-r}$. This is in fact the reason we defined the probability space this way in the previous lecture: we defined the sample point probabilities so that the coin tosses would behave independently.

2. **Balls and bins.** The sample space here is the product $\Omega = \Omega_1 \times \Omega_2 \times \cdots \times \Omega_m$, where $\Omega_i = \{1, 2, \ldots, n\}$ is the set of n bins that could be chosen by the ith ball. (Recall that there are m balls and n bins.) Since

[1] Recall that, for sets A, B, the *Cartesian product* $A \times B$ is the set consisting of all (ordered) pairs (a, b) with $a \in A$ and $b \in B$.

[2] Strictly speaking, we should really also have checked from our original definition of the probability space that $\Pr[A_1]$, $\Pr[A_2|A_1]$ and $\Pr[A_3|A_1 \cap A_2]$ are all equal to $\frac{1}{2}$.

we specify that the balls are thrown independently, by similar reasoning to the above we must have $\Pr[\omega] = \Pr[\omega_1] \times \cdots \times \Pr[\omega_n] = (\frac{1}{n})^m$ for each sample point ω. This is again in agreement with our definition of this probability space in the previous lecture.

The product rule also gives us a new way to compute probabilities of events. Let A be the event that bin 1 is empty. Then we can write $A = \bigcap_{i=1}^m A_i$, where A_i is the event that ball i misses bin 1. Clearly $\Pr[A_i] = 1 - \frac{1}{n}$ for each i. Also, by construction of our sample space the events A_i are mutually independent. So by the product rule

$$\Pr[A] = \Pr[A_1] \times \Pr[A_2|A_1] \times \cdots \times \Pr[A_n|\textstyle\bigcap_{i=1}^{n-1} A_i] = \Pr[A_1] \times \Pr[A_2] \times \cdots \times \Pr[A_n] = \left(1 - \dfrac{1}{n}\right)^m.$$

This agrees with the answer we obtained in the previous lecture by counting sample points.

So now we have two methods of computing probabilities in many of our sample spaces. It is useful to keep these different methods around, both as a check on your answers and because in some cases one of the methods is easier to use than the other.

3. **Card shuffling.** Each sample point ω can be viewed as a sequence of 52 choices, $\omega = (\omega_1, \omega_2, \ldots, \omega_{52})$, where ω_i is the ith card of the deck. First we pick the top card, ω_1, from all 52 cards; then we pick the second card, ω_2, from the remaining 51 cards; then ω_3 from the remaining 50 cards; and so on.[3] The probability of choosing each ω_1 is clearly $\frac{1}{52}$; given ω_1, the probability of choosing each possible ω_2 is uniform, namely $\frac{1}{51}$; given ω_1 and ω_2, the probability of each possible ω_3 is $\frac{1}{50}$; and so on. These conditional probabilities imply, using the product rule:

   $$\Pr[\omega] = \Pr[\omega_1] \times \Pr[\omega_2|\omega_1] \times \ldots \times \Pr[\omega_{52}|\omega_1, \ldots, \omega_{51}] = \frac{1}{52} \times \frac{1}{51} \times \frac{1}{50} \times \ldots \times \frac{1}{2} \times \frac{1}{1} = \frac{1}{52!}.$$

   Reassuringly, this is in agreement with our definition of the probability space in the previous lecture, based on counting permutations.

   Let A be the event that the top card is an ace. Then we see immediately, just by considering the first choice ω_1, that $\Pr[A] = \frac{4}{52} = \frac{1}{13}$. And if B is the event that the top two cards have the same suit, then we can view B as the event that the second choice produces the same suit as the first. This can be calculated as follows:

   $$\Pr[\mathrm{suit}(\omega_2) = \mathrm{suit}(\omega_1)] = \sum_{\omega_1} \Pr[\omega_1] \times \Pr[\mathrm{suit}(\omega_2) = \mathrm{suit}(\omega_1)|\omega_1] = \sum_{\omega_1} \Pr[\omega_1] \times \frac{12}{51} = \frac{12}{51}.$$

4. **Poker hands.** Again we can view the sample space as a sequence of choices. First we choose one of the cards (note that it is not the "first" card, since the cards in our hand have no ordering) uniformly from all 52 cards. Then we choose another card from the remaining 51, and so on. For any given poker hand, the probability of choosing it is (by the product rule):

   $$\frac{5}{52} \times \frac{4}{51} \times \frac{3}{50} \times \frac{2}{49} \times \frac{1}{48} = \frac{1}{\binom{52}{5}},$$

   just as before. Where do the numerators 5, 4, 3, 2, 1 come from? Well, for the given hand the first card we choose can be any of the five in the hand: i.e., five choices out of 52. The second can be any of the remaining four in the hand: four choices out of 51. And so on. This arises because the order of the cards in the hand is irrelevant.

[3] Note that we can't write $\Omega = \Omega_1 \times \Omega_2 \times \ldots \times \Omega_{52}$ here since the sample space Ω_i depends on $\omega_1, \omega_2, \omega_{i-1}$. (E.g., Ω_2, the sample space for the second card, consists of all cards *except* ω_1.)

Let's use this view to compute the probability of a flush in a different way. Clearly this is $4 \times \Pr[A]$, where A is the probability of a Hearts flush. And we can write $A = \bigcap_{i=1}^{5} A_i$, where A_i is the event that the ith card we pick is a Heart. So we have

$$\Pr[A] = \Pr[A_1] \times \Pr[A_2|A_1] \times \cdots \times \Pr[A_5|\bigcap_{i=1}^{4} A_i].$$

Clearly $\Pr[A_1] = \frac{13}{52} = \frac{1}{4}$. What about $\Pr[A_2|A_1]$? Well, since we are conditioning on A_1 (the first card is a Heart), there are only 51 remaining possibilities for the second card, 12 of which are Hearts. So $\Pr[A_2|A_1] = \frac{12}{51}$. Similarly, $\Pr[A_3|A_1 \cap A_2] = \frac{11}{50}$, and so on. So we get

$$4 \times \Pr[A] = 4 \times \frac{13}{52} \times \frac{12}{51} \times \frac{11}{50} \times \frac{10}{49} \times \frac{9}{48},$$

which is exactly the same fraction we computed in the previous lecture.

So now we have two methods of computing probabilities in many of our sample spaces. It is useful to keep these different methods around, both as a check on your answers and because in some cases one of the methods is easier to use than the other.

5. **Monty Hall.** Recall that we defined the probability of a sample point by multiplying the probabilities of the sequence of choices it corresponds to; thus, e.g.,

   $$\Pr[(1,1,2)] = \tfrac{1}{3} \times \tfrac{1}{3} \times \tfrac{1}{2} = \tfrac{1}{18}.$$

   The reason we defined it this way is that we knew (from our model of the problem) the probabilities for each choice *conditional on* the previous one. Thus, e.g., the $\frac{1}{2}$ in the above product is the probability that Monty opens door 2 conditional on the prize door being door 1 and the contestant initially choosing door 1. Once again, we used these conditional probabilities to *define* the probabilities of our sample points.

Unions of events

You are in Las Vegas, and you spy a new game with the following rules. You pick a number between 1 and 6. Then three dice are thrown. You win if and only if your number comes up on at least one of the dice.

The casino claims that your odds of winning are 50%, using the following argument. Let A be the event that you win. We can write $A = A_1 \cup A_2 \cup A_3$, where A_i is the event that your number comes up on die i. Clearly $\Pr[A_i] = \frac{1}{6}$ for each i. Therefore,

$$\Pr[A] = \Pr[A_1 \cup A_2 \cup A_3] = \Pr[A_1] + \Pr[A_2] + \Pr[A_3] = 3 \times \frac{1}{6} = \frac{1}{2}.$$

Is this calculation correct? Well, suppose instead that the casino rolled six dice, and again you win iff your number comes up at least once. Then the analogous calculation would say that you win with probability $6 \times \frac{1}{6} = 1$, i.e., certainly! The situation becomes even more ridiculous when the number of dice gets bigger than 6.

The problem is that the events A_i are *not disjoint*: i.e., there are some sample points that lie in more than one of the A_i. (We could get really lucky and our number could come up on two of the dice, or all three.) So if we add up the $\Pr[A_i]$ we are counting some sample points more than once.

Fortunately, there is a formula for this, known as the Principle of Inclusion/Exclusion:

Theorem 16.4: **[Inclusion/Exclusion]** *For events $A_1, \ldots, A_n$ in some probability space, we have*

$$\Pr[\bigcup_{i=1}^{n} A_i] = \sum_{i=1}^{n} \Pr[A_i] - \sum_{\{i,j\}} \Pr[A_i \cap A_j] + \sum_{\{i,j,k\}} \Pr[A_i \cap A_j \cap A_k] - \cdots \pm \Pr[\bigcap_{i=1}^{n} A_i].$$

[In the above summations, $\{i, j\}$ denotes all unordered pairs with $i \neq j$, $\{i, j, k\}$ denotes all unordered triples of distinct elements, and so on.]

I.e., to compute $\Pr[\bigcup_i A_i]$, we start by summing the event probabilities $\Pr[A_i]$, then we *subtract* the probabilities of all pairwise intersections, then we *add* back in the probabilities of all three-way intersections, and so on.

We won't prove this formula here; but you might like to verify it for the special case $n = 3$ by drawing a Venn diagram and checking that every sample point in $A_1 \cup A_2 \cup A_3$ is counted exactly once by the formula. You might also like to prove the formula for general n by induction (in similar fashion to the proof of Theorem 17.3).

Taking the formula on faith, what is the probability we get lucky in the new game in Vegas?

$$\Pr[A_1 \cup A_2 \cup A_3] = \Pr[A_1] + \Pr[A_2] + \Pr[A_3] - \Pr[A_1 \cap A_2] - \Pr[A_1 \cap A_3] - \Pr[A_2 \cap A_3] + \Pr[A_1 \cap A_2 \cap A_3].$$

Now the nice thing here is that the events A_i are mutually independent (the outcome of any die does not depend on that of the others), so $\Pr[A_i \cap A_j] = \Pr[A_i]\Pr[A_j] = (\frac{1}{6})^2 = \frac{1}{36}$, and similarly $\Pr[A_1 \cap A_2 \cap A_3] = (\frac{1}{6})^3 = \frac{1}{216}$. So we get

$$\Pr[A_1 \cup A_2 \cup A_3] = \left(3 \times \tfrac{1}{6}\right) - \left(3 \times \tfrac{1}{36}\right) + \tfrac{1}{216} = \tfrac{91}{216} \approx 0.42.$$

So your odds are quite a bit worse than the casino is claiming!

When n is large (i.e., we are interested in the union of many events), the Inclusion/Exclusion formula is essentially useless because it involves computing the probability of the intersection of every non-empty subset of the events: and there are $2^n - 1$ of these! Sometimes we can just look at the first few terms of it and forget the rest: note that successive terms actually give us an overestimate and then an underestimate of the answer, and these estimates both get better as we go along.

However, in many situations we can get a long way by just looking at the first term:

1. **Disjoint events.** If the events A_i are all *disjoint* (i.e., no pair of them contain a common sample point — such events are also called *mutually exclusive*), then

   $$\Pr[\bigcup_{i=1}^{n} A_i] = \sum_{i=1}^{n} \Pr[A_i].$$

   [Note that we have already used this fact several times in our examples, e.g., in claiming that the probability of a flush is four times the probability of a Hearts flush — clearly flushes in different suits are disjoint events.]

2. **Union bound.** Always, it is the case that

   $$\Pr[\bigcup_{i=1}^{n} A_i] \leq \sum_{i=1}^{n} \Pr[A_i].$$

   This merely says that adding up the $\Pr[A_i]$ can only *over*estimate the probability of the union. Crude as it may seem, in the next lecture we'll see how to use the union bound effectively in a Computer Science example.