CS30 (Discrete Math in CS), Summer 2021 : Lecture 13

Topic: Probability: Basics

Disclaimer: These notes have not gone through scrutiny and in all probability contain errors. Please discuss in Piazza/email errors to deeparnab@dartmouth.edu

---

- **Experiments and Outcomes : Sample Spaces.**

  Every time you hear a question beginning with, "What are the chances that ... ", an *experiment* with an (yet) unknown *outcome* has taken place. Let us consider a few examples:

  – What is the chance that a tossed coin lands heads?
  – What is the chance we get a score of 7 with two rolled dice?
  – What is the chance that two people in a party share a birthday?

  In each of these questions above, there are experiments involved. We toss a coin. We roll two dice. We note down people's birthdays. Each of these experiments have outcomes involved. In the first one, the outcomes are heads or tails. In the second one, the outcomes are tuples indicating the rolls of each die. In the third one, the outcome is a list of birthdays of the people in the party.

  The *first step* in answering in any question of the form above, we must be sure what the *experiment* is, and what the possible *outcomes* are. This forms the ***sample space*** of the question at hand.

  > **Exercise:** *What is the* **sample space** *for the following questions (google the definitions if you don't know them):*
  >
  > – *What is the chance of getting a royal flush in a five hand poker draw?*
  > – *What is the chance of getting four aces in a bridge hand?*
  > – *What is the chance of seeing a run of* 5 *heads when you toss a coin* 100 *times?*

- **Events.**

  Ok. So you have figured out the *sample space* of the question at hand, and let us call Ω the set of all possible outcomes. So in the first question $\Omega = \{\texttt{heads}, \texttt{tails}\}$, for the second question Ω is set $\{(1,1),(1,2),\ldots,(6,5),(6,6)\}$. For the third question, Ω is *all* possible lists of birthdays. How large is this list? (You can answer this now).

  We go back to the question and figure out the "set of outcomes we are interested in". Sometimes it is just one outcome; in the first question, we are only interested in the outcome `heads`. In the second question, we are interested in multiple outcomes such as $(3,4)$ or $(6,1)$, and so on. Similarly, in the third event, there are multiple outcomes which are of interest; any list which has *at least* two repeating birthdays is of interest to us. In general, a question of the form "what are the chances" asks the question of an *event*. Formally, an event $\mathcal{E} \subseteq \Omega$ is the subset of outcomes of interest.

  > **Exercise:** *What are the* **events** *for the following questions*
  >
  > – *What is the chance of getting a royal flush in a five hand poker draw?*
  > – *What is the chance of getting four aces in a bridge hand?*
  > – *What is the chance of seeing a run of* 5 *heads when you toss a coin* 100 *times?*

- **Randomness and Probabilities.**

  Right. Now we have the sample space Ω figured out. We have also figured out the relevant outcomes $\mathcal{E} \subseteq \Omega$ which we are interested in. Now comes the time to assign *likelihoods* to all the outcomes in Ω so that we can reasonably answer the question at hand. Formally, if we let Ω denote the set of possible outcomes, then the likelihood of any outcome $o \in \Omega$ is a "score" called the *probability* $\mathbf{Pr}[o]$ which is a *real number* between 0 and 1. Apart from this, the only constraint on $\mathbf{Pr}$ is that $\sum_{o\in\Omega} \mathbf{Pr}[o] = 1$.

  This function $\mathbf{Pr}$ which assigns a non-negative real value to each outcome is called the *probability distribution*. If $\mathbf{Pr}[o] = 1/|\Omega|$ for all $o \in \Omega$, then the distribution is called the *uniform distribution over the sample space*.

  How does one assign these scores? Ultimately, at some level these are determined by the *modeling assumptions* of the experiment at hand. However, what math allows us to do is to make as few and as *elementary* assumptions as possible. Let's discuss this by considering the questions again from the previous bullet point.

  - *What is the chance that a tossed coin lands heads?* The sample space is $\Omega = \{\texttt{heads}, \texttt{tails}\}$. What is $\mathbf{Pr}[\texttt{heads}]$? Well, if we assume that the coin was fair, and the toss was fair, then it is *reasonable* to assume that $\mathbf{Pr}[\texttt{heads}] = \mathbf{Pr}[\texttt{tails}]$, and therefore since they sum to 1, each of these should be $1/2$. This is *one* model. Another model could be that perhaps the coin is not fair, but in that case the model must specify what the probability of getting a heads is.
  - *What is the chance we get a score of* 7 *with two rolled dice?* It seems reasonable to assume is that each of the two die is fair, that is, each die returns a answer in $\{1, 2, 3, 4, 5, 6\}$ equally likely. That is the *elementary* assumption. Indeed, this assumption will let us argue that if we took $\Omega = \{(1, 1), (1, 2), \ldots, (6, 5), (6, 6)\}$, then the probability would indeed be uniform over this sample space.
  - *What is the chance that two people in a party share a birthday?* This is slightly trickier. The sample space Ω, remember, is a list of dates. Are all lists equally likely? Is that a reasonable assumption? Depends. In truth, it doesn't seem to be so. Data shows that more people are born in September than in February. So the lists should have more September than Februaries. But for the *purpose of a first cut*, it doesn't seem to be too bad to assume all birthdates are equally likely. And we will make this assumption to answer this question.

  Ok, so where are we? When faced with a question we figure out the sample space Ω. We make our modeling assumptions, make them as simple and as reasonable as possible. Sometimes these assumptions immediately lead to the $\mathbf{Pr}[o]$ for each $o \in \Omega$ (like the coin question above, like the birthday question above). Sometimes not (for the dice question above).

  If we figure out the $\mathbf{Pr}[o]$ for every $o \in \Omega$, then we can figure out the probability of an event

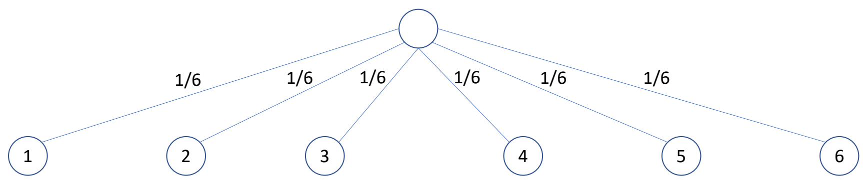
  $$\mathbf{Pr}[\mathcal{E}] = \sum_{o\in\mathcal{E}} \mathbf{Pr}[o]$$

  And indeed, if the distribution is uniform, then $\mathbf{Pr}[o] = \frac{1}{|\Omega|}$ then, $\mathbf{Pr}[\mathcal{E}] = \frac{|\mathcal{E}|}{|\Omega|}$.

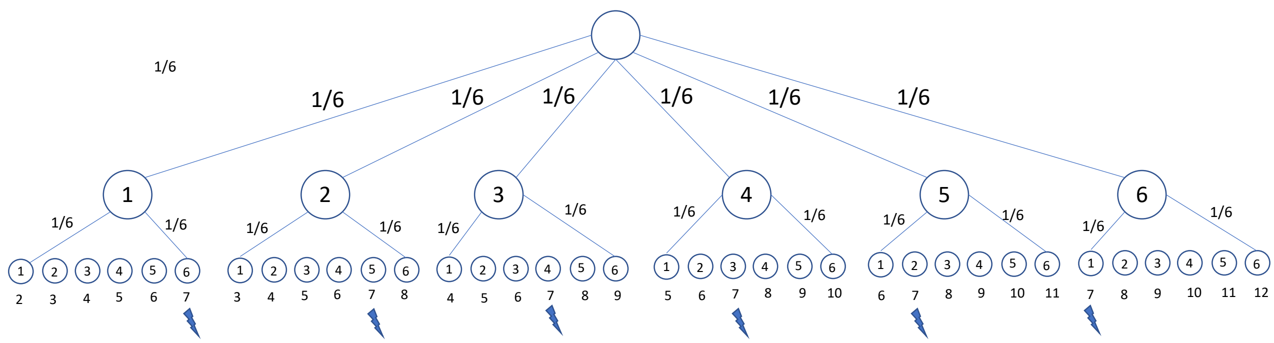
- **Figuring out Outcome Probabilities and the Probability of the Event.**

Let's figure out the outcome probabilities for the dice question. We do so by drawing the "tree diagram". This is very useful technique when there are "multiple" elementary assumptions which lead to the outcomes.

Back to the dice question. We first roll the first die, and draw the situations as a tree below. The outcome is written in the circle. The probabilities are written on the edge. These are equally likely due to our modeling assumption.



Now, we roll the second die. We write the outcome of the second die in the circles in the second level. The sums are written below. The "lightnings" indicate the circles of interest.



What is the probability of landing up in a smaller circle (leaf of the tree)? We can figure out by "walking down" the tree. For instance take the left-most small circle. To get there, we must first take the left most edge from the root. The chances of doing this is $\frac{1}{6}$. Having done that, we have to take the left most edge again. So the total probability is $\frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36}$. Indeed, this is the probability of *each outcome* in the second level.

To figure out the probability of the event, since there are 6 of these outcomes which lead to the sum of 7, the probability of the event that we see a 7 is indeed $6/36 = 1/6$.

Exercise: *What is the probability that the sum of two rolled dice leads to a prime number?*

- **Notation.**

  With events, we often mix-and-match notation from Boolean Logic and Sets.

- Given an event $\mathcal{E}$, the negation event $\neg\mathcal{E}$ is used to denote the event that $\mathcal{E}$ doesn't take place. That is, it is simply the subset $\neg\mathcal{E} = \Omega \setminus \mathcal{E}$. Sometimes, $\neg\mathcal{E}$ is denoted as $\overline{\mathcal{E}}$.

$$\mathbf{Pr}[\mathcal{E}] + \mathbf{Pr}[\neg\mathcal{E}] = 1$$

- Given two events $\mathcal{E}$ and $\mathcal{F}$, the notation $\mathcal{E} \cup \mathcal{F}$ is precisely the union of the subsets in the sample space. $\mathbf{Pr}[\mathcal{E} \cup \mathcal{F}]$ captures the likelihood that at least one of the events takes place.
- Given two events $\mathcal{E}$ and $\mathcal{F}$, the notation $\mathcal{E} \cap \mathcal{F}$ is precisely the intersection of the subsets in the sample space. $\mathbf{Pr}[\mathcal{E} \cap \mathcal{F}]$ captures the likelihood that both the events takes place.
- Two events $\mathcal{E}$ and $\mathcal{F}$ are *disjoint* or *exclusive* if $\mathcal{E} \cap \mathcal{F} = \varnothing$. That is, they both can't occur simultaneously. A collection of events $\mathcal{E}_1, \mathcal{E}_2, \ldots, \mathcal{E}_k$ are *mutually exclusive* if $\mathcal{E}_i \cap \mathcal{E}_j = \varnothing$ for $i \neq j$.
- For mutually exclusive events,

$$\mathbf{Pr}[\mathcal{E}_1 \cup \mathcal{E}_2 \cup \cdots \mathcal{E}_k] = \sum_{i=1}^{k} \mathbf{Pr}[\mathcal{E}_i]$$

- The Inclusion-Exclusion formula (for two events, aka Baby version) tells us

$$\mathbf{Pr}[\mathcal{E} \cup \mathcal{F}] = \mathbf{Pr}[\mathcal{E}] + \mathbf{Pr}[\mathcal{F}] - \mathbf{Pr}[\mathcal{E} \cap \mathcal{F}]$$

  Do you see why? It is *exactly* the baby-version of inclusion-exclusion if $\mathbf{Pr}$ is a *uniform distribution*. Indeed, if this is the case then $\mathbf{Pr}[\mathcal{E} \cup \mathcal{F}] = \frac{|\mathcal{E} \cup \mathcal{F}|}{|\Omega|}$, and the proof follows by applying baby IE. What if it is not uniform?

  Exercise: *Prove the above even when* $\mathbf{Pr}$ *is* not *a uniform distribution.*

- **The Birthday "Paradox".**

  Let's tackle the party question: in a party of N people, what is the chance that at least two persons share a birthday. We will assume just for simplicity that no one is born on the leap day. As we discussed above, the sample space is

  Ω = Set of all length N sequences where each entry is a date among the 365 dates.

  We know that

$$|\Omega| = 365^N \tag{1}$$

  What is the event we are interested in? We are interested in the following subset $\mathcal{E}$ of Ω

  $\mathcal{E}$ = Set of all length N sequences where each entry is a date among the 365 dates, **and** at least two entries are same.

  What is the probability assumption? We will assume (again for simplicity) that every person is *equally likely* to be born on any of the days. That is, the probability of a random person's birthday being January 31 is the same as it being February 14, etc.

  This assumption is immensely helpful in figuring out $\mathbf{Pr}[o]$ for any $o \in \Omega$. Note, o is just a list of N dates. They lie on the leaves of the (extremely bushy) tree of height N. But, the numbers of *every* edge of the tree, *because of our uniformity assumption*, is $\frac{1}{365}$. Therefore, $\mathbf{Pr}[o]$ for any $o \in \Omega$ is $\frac{1}{365^N}$. Indeed, this means that the probability distribution on Ω is *uniform*.

And so,

$$\mathbf{Pr}[\mathcal{E}] = \frac{|E|}{|\Omega|}$$

Thus, we have boiled down our question of chance to a question in combinatorics (Ah, it pays off!). What is the size of $|\mathcal{E}|$? By now we know this is an "at least" kind of statement. Let's look at the *negation event*.

$\neg\mathcal{E}$ = Set of all length N sequences where each entry is a date among the 365 dates, **and** no entries are same.

But this is just asking "how many sequences of length N are there where each entry is from 1 to 365 and repetition now allowed?" The answer is (make sure you know this before reading)

$$|\neg\mathcal{E}| = \frac{365!}{(365-N)!} = 365 \cdot 364 \cdot 363 \cdots \cdot (365 - N + 1)$$

And so,

$$\mathbf{Pr}[\neg\mathcal{E}] = \frac{365 \cdot 364 \cdot 363 \cdots \cdot (365 - N + 1)}{365^N} = \left(\frac{365}{365}\right) \cdot \left(\frac{364}{365}\right) \cdot \left(\frac{363}{365}\right) \cdot \cdots \cdot \left(\frac{365 - N + 1}{365}\right)$$

This is the probability that *no two people* share a birthday. And therefore, the probability that at least two people share a birthday is

$$\mathbf{Pr}[\mathcal{E}] 1 - \frac{365 \cdot 364 \cdot 363 \cdots \cdot (365 - N + 1)}{365^N} = \left(\frac{365}{365}\right) \cdot \left(\frac{364}{365}\right) \cdot \left(\frac{363}{365}\right) \cdot \cdots \cdot \left(\frac{365 - N + 1}{365}\right)$$

This function gets closer and closer to 1 as N gets larger. And indeed, if $N > 365$ it becomes one. Do you see this? Do you also see why in a set of 366 people atleast two people must share a birthday (not counting the leap day)? Pigeon-hole-principle (UGP 2, Problem 1).

Anyway, how *fast* does this function grow with N. This is something which is trivial to do with a computer. (You should do it, like right now!). This is plotted in Figure (am plotting $N = 1$ to 100). At around $N = 23$, one breaks the 50% "barrier" (it is more likely for a collision of birthdays than not), and then rises pretty steadily. At around $N = 50$, it is almost a certaintly. At the size of our class, I would take a wager of $10000 : 1$ odds betting that there is a collision.

Remark:

- *One could do the above calculation for anything – why birthdays? One could ask, suppose one had a list of size N taken uniformly at random over a sample space of lists where each entry came uniformly at random from a set of sized M (for birthdays, $M = 365$). Roughly how large N should be (compared to M) to see a greater chance of collision than no collision. You could repeat the math as above. And then with some "mathy jugglery" (actually the binomial theorem, nothing more) one could show the break even point is roughly $\sqrt{M}$. Especially true if M is large. So, if we have $10,000$ choices, then at rougly 100-odd people we would get a collision, which could be a near certainly at around 200.*

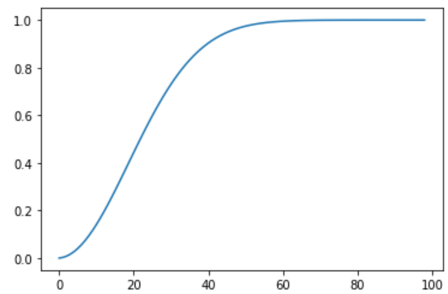


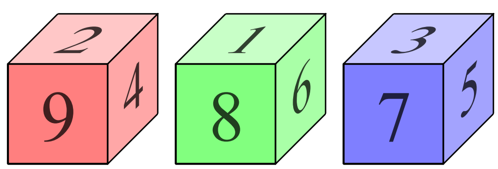
Figure 1: The probability of at least two people sharing a birthday with N, the number of people in the party.

- *One could also ask, "Hey, your assumption about birthdays — that seems wrong. Maybe in real life your assumption doesn't hold, and then maybe in a set of* 50 *people we see all distinct birthdays. Turns out (and this is more difficult to show) that the* uniform assumption *is* worst *for collisions. That is, if more people were born in September than January, then in fact the chance of collision would* increase*. At some level it makes sense — we would get more January-born people in our party.*
- *Lastly, a puzzle. How would we ever* really *find out if there is some pair of people in our class who share the same birthday without revealing the "private info"? Hmmm...*

We didn't do this in class, but read on for more "paradoxes".

- **Intransitive Dice.**

Consider the following whacky dice. Each die has the same number on opposite sides.

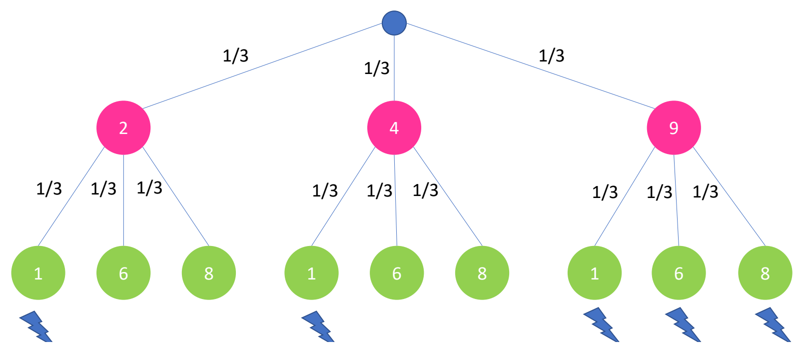


Let's call the red die R, the green die G, and the blue die B. Fix two of these: say R and G. We roll both of them. We are interested in the event that the red die gets a bigger number than the green die.

What is the sample space? Without loss of generality let us imagine the red die is rolled first and the green die is rolled. The sample space Ω is a collection of tuples indicating the possibilities of both rolls. Concretely,

$$\Omega = \{(2,1),(2,6),(2,8),(4,1),(4,6),(4,8),(9,1),(9,6),(9,8)\}$$

To figure out the probabilities of the outcomes, we could draw a tree diagram. The "lightnings" denote when the red die gets a bigger number than the green die.



Each of the circles in the second layer occur with probability $\frac{1}{3} \cdot \frac{1}{3} = \frac{1}{9}$. Therefore, the probability that the red die gets a bigger number than the green die is $5/9$. Therefore, we get that the Red die is a better die than Green die. We denote this as $R > G$ to show R "beats" G.

Exercise: *Drawing the tree diagrams, what is the probability that the green die beats the blue die with one roll? What is the probability that the blue die beats the red die? Do you see something counter-intuitive?*

- **The Monty Hall Problem.**

  Here is a question you may or may not have heard before:

  > *Suppose you're on a game show, and you're given the choice of three doors: Behind one door is a car; behind the others, goats. You pick a door, say No. 1, and the host, who knows what's behind the doors, opens another door, say No. 3, which has a goat. He then says to you, "Do you want to pick door No. 2?" Is it to your advantage to switch your choice?*

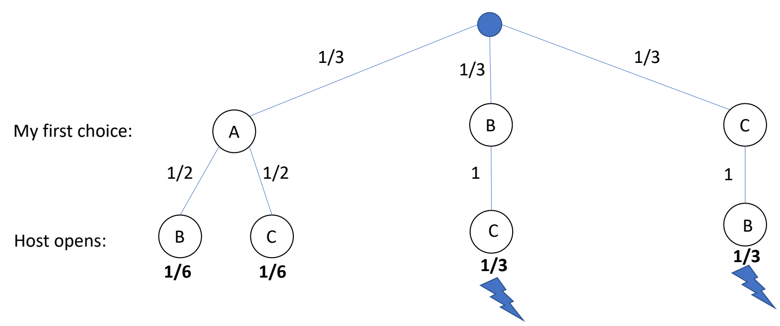
  At first glance, it may not seem like a question in probability at all. And indeed, it is ill-formulated. To formulate this well, we need to define the experiments, define the outcomes, clearly state the modeling assumptions, calculate the probabilities based on this, and then answer.

  So here are the assumptions which I would make depending on my reading:

  – The first door I pick is uniformly at random between the three doors. I don't know better.
  – The host who does know the answer chooses a door to open. He is constrained to open a door which has a goat. However, my belief is that if there are two door which has goats, then he would open one of the two also uniformly at random. Once again, my rationale is that the question doesn't specify it, and so I assume he doesn't know better.

  Once we have fixed this, then we can ask the question "What is the probability that the 'third door' has the car?" To answer this, again, we can use tree diagrams.

  Before we begin, let us *rename* the doors such that the car is behind Door 1, and Door 2 and 3 contains goats. This is without loss of generality – my choice and the host's choice, by my assumption, doesn't depend on the numbers. The tree diagram is below. The first layer shows the random choices for the first door. The second layer shows the random choices of the door opened by the host. Note, that it is random only in the first branch; if the door I choose in the first try has a goat, then the host has to open the third door (the one not containing the car, and the one I chose). The "lightning" indicates the outcomes in which switching leads me to the car. The numbers in bold indicates the probability of that outcome.



As you can see from the tree-diagram, the probability switching helps is $2/3$.

Answers to some exercises.

- **Exercise:** *What is the* **sample space** *for the following questions (google the definitions if you don't know them):*

  - *What is the chance of getting a royal flush in a five hand poker draw?*
  - *What is the chance of getting four aces in a bridge hand?*
  - *What is the chance of seeing a run of* 5 *heads when you toss a coin* 100 *times?*

  - The sample space is the set of 5-cards drawn from a 52-deck card with no repetitions allowed.
  - The sample space is the set of 4-cards drawn from a 52-deck card with no repetitions allowed.
  - The sample space is the set of all **seqeunces** of length 100 whose entries are heads or tails.

- **Exercise:** *What are the* **events** *for the following questions (google the definitions if you don't know them):*

  - *What is the chance of getting a royal flush in a five hand poker draw?*
  - *What is the chance of getting four aces in a bridge hand?*
  - *What is the chance of seeing a run of* 5 *heads when you toss a coin* 100 *times?*

  - The event is the set of 5-cards whose ranks are $A, K, Q, J, 10$ and their suits are the same. There are only 4 elements in this event set.
  - The event is the set of 4-cards whose ranks are A. There is only 1 elements in this event set.
  - The event is the set of all **seqeunces** of length 100 whose entries are heads or tails, and which contains 5 consecutive heads somewhere.

- **Exercise:** *What is the probability that the sum of two rolled dice leads to a prime number?*

I am not drawing the "tree-diagram" here, but perhaps you should till you are comfortable. We know that the sample space is $\Omega = \{(a,b) : 1 \le a \le 6, 1 \le b \le 6\}$ and $|\Omega| = 36$. We know that $\mathcal{E} = \{(a,b) : (a,b) \in \Omega,\ a+b = \text{ prime }\}$. In this case,

$$\mathcal{E} = \{(1,1),(1,2),(1,4),(1,6),(2,1),(2,3),(2,5),(3,2),(3,4),(4,1),(4,3),(5,2),(5,6),(6,1),(6,5)\}$$

And thus $|\mathcal{E}| = 15$. Thus, the answer is $\frac{15}{36} = \frac{5}{12}$.

- **Exercise:**

$$\mathbf{Pr}[\mathcal{E} \cup \mathcal{F}] = \mathbf{Pr}[\mathcal{E}] + \mathbf{Pr}[\mathcal{F}] - \mathbf{Pr}[\mathcal{E} \cap \mathcal{F}]$$

  Prove the above even when $\mathbf{Pr}$ *is* not *a uniform distribution.*

The way to do is to mimic the proof of the baby inclusion exclusion. We write $\mathcal{E} = (\mathcal{E} \cap \mathcal{F}) \cup (\mathcal{E} \setminus \mathcal{F})$ as a disjoint union. And thus,

$$\mathbf{Pr}[\mathcal{E}] = \mathbf{Pr}[\mathcal{E} \cap \mathcal{F}] + \mathbf{Pr}[\mathcal{E} \setminus \mathcal{F}]$$

And then we write $\mathcal{E} \cup \mathcal{F} = \mathcal{F} \cup (\mathcal{E} \setminus \mathcal{F})$ as a disjoint union, giving us

$$\mathbf{Pr}[\mathcal{E} \cup \mathcal{F}] = \mathbf{Pr}[\mathcal{F}] + \mathbf{Pr}[\mathcal{E} \setminus \mathcal{F}]$$

and the identity follows by rearrangement.